



(86) Date de dépôt PCT/PCT Filing Date: 2011/07/14
(87) Date publication PCT/PCT Publication Date: 2012/01/19
(85) Entrée phase nationale/National Entry: 2013/01/14
(86) N° demande PCT/PCT Application No.: US 2011/044063
(87) N° publication PCT/PCT Publication No.: 2012/009568
(30) Priorité/Priority: 2010/07/16 (US61/365,194)

(51) Cl.Int./Int.Cl. *C40B 40/08* (2006.01),
C07K 16/00 (2006.01), *C12N 1/19* (2006.01),
C12N 15/13 (2006.01), *C40B 30/02* (2006.01),
C40B 30/04 (2006.01), *C40B 40/02* (2006.01),
C40B 40/10 (2006.01), *C40B 50/00* (2006.01)

(71) Demandeur/Applicant:
ADIMAB, LLC, US

(72) Inventeurs/Inventors:
VASQUEZ, MAXIMILIANO, US;
SIVASUBRAMANIAN, ARVIND, US;
FELDHAUS, MICHAEL, US

(74) Agent: TORYS LLP

(54) Titre : BANQUES D'ANTICORPS
(54) Title: ANTIBODY LIBRARIES

(57) **Abrégé/Abstract:**

Provided is a method that overcomes the inadequacies inherent in the known methods for generating libraries of antibody-encoding polynucleotides by specifically designing the libraries with directed sequence and length diversity.



ANTIBODY LIBRARIES**ABSTRACT**

The present invention overcomes the inadequacies inherent in the known methods for generating libraries of antibody-encoding polynucleotides by specifically designing the libraries with directed sequence and length diversity.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 239

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 239

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

ANTIBODY LIBRARIES

RELATED APPLICATION

This application claims priority to U.S. provisional application Ser. No. 61/365,194, filed on July 16, 2010, incorporated herein in its entirety by this reference.

BACKGROUND

Antibodies have profound relevance as research tools and in diagnostic and therapeutic applications. However, the identification of useful antibodies is difficult and once identified, antibodies often require considerable redesign or “humanization” before they are suitable for therapeutic applications in humans.

Many methods for identifying antibodies involve display of libraries of antibodies derived by amplification of nucleic acids from B cells or tissues. Some of these methods have utilized synthetic libraries. However, many of these approaches have limitations. For example, most human antibody libraries known in the art contain only the antibody sequence diversity that can be experimentally captured or cloned from a biological source (*e.g.*, B cells). Accordingly, such libraries may over-represent some sequences, while completely lacking or under-representing other sequences, particularly those binding human antigens. Most synthetic libraries known in the art have other limitations, such as the occurrence of unnatural (*i.e.*, non-human) amino acid sequence motifs that have the potential to be immunogenic.

Accordingly, a need exists for diverse antibody libraries that contain candidate antibodies that are non-immunogenic (*i.e.*, are human) and have desired properties (*e.g.*, the ability to recognize a broad variety of antigens). However, obtaining such libraries requires balancing the competing objectives of generating diverse libraries while still maintaining the human character of the sequences within the library. The current invention provides antibody libraries that have these and other desirable features, and methods of making and using such libraries.

SUMMARY

The invention provides, among other things, improvements in the design and production of synthetic libraries that mimic the diversity of the natural human repertoire of CDRH3, CDRL3, heavy chain, light chain, and/or full-length (intact) antibody sequences. In some embodiments the invention defines and provides methods of generating theoretical segment pools of TN1, DH, N2, and H3-JH segments to consider for inclusion in a physical manifestation of a library (*e.g.*, polynucleotide or polypeptide) comprising or encoding CDRH3 sequences (*e.g.*, an antibody library). In certain embodiments the invention defines and provides methods of matching the individual members of these theoretical segment pools to a reference set of CDRH3 sequences, to determine the frequency of occurrence (or segment usage weight) of each of the segments in the theoretical segment pool in the reference set. While any set of CDRH3 sequences may be used as a reference set, the invention also defines and provides methods of generating particular reference sets or subsets of interest. For example, among other things, the present invention provides methods for filtering an original reference set to obtain a provided reference set with a preimmune character. Also provided are methods to define and/or identify segments that occur within the CDRH3 sequences in the reference set but not in the theoretical segment pool. Such segments can be added to a theoretical segment pool, for example in order to be considered for inclusion in a physical library. Although the frequency of occurrence of a particular segment in a reference set is useful to select segments for inclusion in a physical library, the invention also provides a number of physicochemical and biological properties that can be used (alone or together with any other criterion or criteria) to select segments for inclusion in a physical library.

In some embodiments the invention provides libraries that differ from certain other libraries known in the art in that they are not sitewise-stochastic in composition or sequence, and are therefore intrinsically less random than these certain other libraries of the art (*see e.g.*, Example 14 of US Pub. No. 2009/0181855, incorporated by reference in its entirety, for a discussion of information content and randomness). In some embodiments, degenerate oligonucleotides may be used to further increase the diversity of the members of a library while further improving matching with a reference set of

sequences (e.g., CDRH3, CDRL3, heavy chain, light chain, and/or full-length (intact) antibody sequences).

The invention also provides libraries whose members have sequences that are related to one another in that they would be selected for inclusion in a physical library by performing the analyses described herein, for example by generating a CDRH3 reference set as in Example 3; generating theoretical segment pools as in Examples 5-7; matching the members of a theoretical segment pool to the reference set as in Examples 4 and 8; and selecting members of the theoretical segment pool for inclusion in a physical library as in Examples 8-9. Also provided are methods of further increasing diversity in certain sequences by utilizing degenerate oligonucleotides as in Examples 12-16.

In some embodiments, the present invention provides polynucleotide and polypeptide libraries comprising CDRH3, CDRL3, heavy chain, light chain, and/or full-length (intact) antibody sequences, and methods of making and using such libraries.

In some embodiments, the invention provides libraries comprising, consisting essentially of, or consisting of any of the libraries or theoretical segment pools described herein.

In some embodiments, the present invention recognizes that by mimicking the *in vivo* activity of the enzyme TdT computationally, theoretical segment pools can be generated and subsequently matched to large reference datasets of CDR sequences to choose, for inclusion in a library, those theoretical segments that best recapitulate the CDR sequences in the reference data sets.

In certain embodiments, the invention provides libraries of polynucleotides comprising at least about 10^4 polynucleotides encoding CDRH3 polypeptides with the structure: [TN1]-[DH]-[N2]-[H3-JH], wherein: TN1 is a polypeptide corresponding to any of the TN1 polypeptides of Tables 9-10 and 18-26, or a polypeptide produced by translation of any of the TN1 polynucleotides of Tables 25-26; DH is a polypeptide corresponding to any of the DH polypeptides of Tables 9, 11, 17-25 and 28, or a polypeptide produced by translation of any of the DH-encoding polynucleotides of Tables 16, 25 and 27; N2 is a polypeptide corresponding to any of the N2 polypeptides of Tables 9, 12, 18-25 and 30, or a polypeptide produced by translation of any of the N2-encoding polynucleotides of Tables 25 and 29; and H3-JH is a polypeptide

corresponding to any of the H3-JH polypeptides of Tables 9, 13, 15, 18-25 and 32, or a polypeptide produced by translation of any of the H3-JH-encoding polynucleotides of Tables 14, 25 and 31.

In some embodiments, the invention provides libraries wherein at least about 1%, 5%, or 10% of the sequences in the library have the structure provided above, or that of any of the libraries provided herein.

In certain embodiments, the invention provides libraries comprising polynucleotides encoding CDRH3 polypeptides produced by the sets of TN1, DH, N2, and H3-JH polypeptides provided in any one of Tables 23-25.

In some embodiments, the invention provides libraries comprising polynucleotides encoding CDRH3 polypeptides produced by the set of TN1 polypeptides provided in Table 26, the set of DH polypeptides provided in Table 28, the set of N2 polypeptides provided in Table 30 and the set of H3-JH polypeptides provided in Table 32.

In certain embodiments, the invention provides libraries whose members show (or encode polypeptides that show) at least a certain percent identity with the polypeptides described above, for example, a library comprising at least about 10^4 polynucleotides encoding CDRH3 polypeptides with the structure: [TN1]-[DH]-[N2]-[H3-JH], wherein: TN1 is a polypeptide at least about 80%, 90%, or 95% identical to any of the TN1 polypeptides of Tables 9-10 and 18-26, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the TN1 polynucleotides of Tables 25-26; DH is a polypeptide at least about 80%, 90%, or 95% identical to any of the DH polypeptides of Tables 9, 11, 17-25 and 28, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the DH-encoding polynucleotides of Tables 16, 25 and 27; N2 is a polypeptide at least about 80%, 90%, or 95% identical to any of the N2 polypeptides of Tables 9, 12, 18-25 and 30, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the N2-encoding polynucleotides of Tables 25 and 29; and H3-JH is a polypeptide at least about 80%, 90%, or 95% identical to any of the H3-JH polypeptides of Tables 9, 13, 15, 18-25 and 32, or a polypeptide at least about 80%, 90%,

or 95% identical to a polypeptide produced by translation of any of the H3-JH-encoding polynucleotides of Tables 14, 25 and 31.

In some embodiments, the invention provides libraries comprising polynucleotides encoding light chain variable regions, wherein the light chain variable regions are selected from the group consisting of: (a) a VK1-05 sequence varied at one or more of positions 4, 49, and 46; (b) a VK1-12 sequence varied at one or more of positions 4, 49, 46, and 66; (c) a VK1-33 sequence varied at one or more of positions 4, 49, and 66; (d) a VK1-39 sequence varied at one or more of positions 4, 49, and 46; (e) a VK2-28 sequence varied at one or more of positions 2, 4, 46, and 49; (f) a VK3-11 sequence varied at one or more of positions 2, 4, 36, and 49; (g) a VK3-15 sequence varied at one or more of positions 2, 4, 48, and 49; (h) a VK3-20 sequence varied at one or more of positions 2, 4, 48, and 49; and/or (i) a VK4-1 sequence varied at one or more of positions 4, 46, 49, and 66.

In certain embodiments, the invention provides libraries comprising polynucleotides encoding light chain variable regions that comprise polypeptide sequences at least about 80%, 90%, or 95% identical to two or more of the light chain polypeptide sequences provided in Table 3.

In some embodiments, the invention provides libraries wherein the light chain variable regions comprise the polypeptide sequences provided in Table 3.

In certain embodiments, the invention provides libraries comprising polynucleotides encoding light chain variable regions, wherein the L3-VL polypeptide sequences of the light chain variable regions are varied at two or three residues between positions 89 to 94, inclusive, in comparison to an L3-VL germline sequence. In some embodiments, libraries containing a single light chain germline sequence and its variants are provided. In certain embodiments, variants produced from different light chain germline sequences can be combined to produce libraries encoding multiple light chain germline sequences and their variants. Any of the light chain L3-VL germline sequences provided herein may be varied at two or three residues between positions 89 to 94, inclusive, and one of ordinary skill in the art will recognize that any other L3-VL sequence can also be varied according to the principles described herein to produce

libraries provided by the invention. In some embodiments, the present invention comprises libraries containing polynucleotides that encode antibody light chain variable regions, wherein the antibody light chain variable regions comprise one or more of the following L3-VL sequences: (i) an amino acid sequence that is identical to an L3-VL germline sequence (e.g., see Table 1); (ii) an amino acid sequence that contains two substitutions between residues 89-94, inclusive, in comparison to an L3-VL germline sequence; and (iii) an amino acid sequence that contains three substitutions between residues 89-94, inclusive, in comparison to an L3-VL germline sequence. In some embodiments, each antibody light chain variable region on a library includes one or more of the above L3-VL sequences. In some embodiments, such a library is combined with one or more sets of other nucleic acids that may or may not encode antibody light chain variable regions, and may or may not contain such L3-VL sequences. In some embodiments, the present invention comprises libraries containing polynucleotides that encode an antibody light chain variable region having an amino acid sequence as set forth in Table 4, or a polynucleotide sequence as set forth in one or more of Tables 5-7, wherein two or three residues at positions 89-94, inclusive, are varied.

In some embodiments, the present invention comprises libraries containing polynucleotides that encode an antibody light chain variable region, wherein, across the library, all encoded antibody light chain variable regions are identical to one another except for substitutions of residues at positions between residue 89 and residue 94, inclusive and further wherein, across the library, sequences of any two encoded antibody light chain variable regions differ from one another at not more than 3 positions.

In some embodiments, the invention provides libraries comprising polynucleotides encoding light chain variable regions comprising polypeptide sequences at least about 80%, 90%, or 95% identical to polypeptides produced by translation of two or more of the polynucleotide sequences provided in Tables 5-7. In certain embodiments all members of the library are at least about 80%, 90%, or 95% identical to polypeptides produced by translation of two or more of the polynucleotide sequences provided in Tables 5-7.

In certain embodiments, the invention provides a library comprising light chain variable regions that comprise the polypeptides produced by translation of the polynucleotide sequences provided in Tables 5-7. In certain embodiments, all members of the library comprise the polypeptides produced by translation of the polynucleotide sequences provided in Tables 5-7.

In some embodiments, any of the libraries described herein as containing or encoding CDRL3 and/or light chain variable regions, contains or encodes such CDRL3 and/or light chain variable regions in the context of complete light chains. Furthermore, in some embodiments, such libraries (and/or complete light chain libraries) further contain or encode one or more heavy chain CDRH3, variable domains, or intact heavy chains. In some embodiments, provided libraries include or encode intact antibodies such as, for example, intact intact IgGs.

In some embodiments, provided libraries include or encode human antibodies or antibody fragments; in some such embodiments, provided libraries include or encode intact human antibodies.

In certain embodiments, the invention provides libraries that comprise nucleic acid vectors containing library nucleic acids described herein. In many embodiments, each such library member comprises the same vector.

In some embodiments, the invention provides host cells containing one or more provided libraries, for example including a vector. In some embodiments the host cell is a yeast, and in certain embodiments the yeast is *Saccharomyces cerevisiae*.

In some embodiments, the invention provides antibodies isolated from the libraries described herein.

In certain embodiments, the invention provides kits containing any of the libraries described herein.

In some embodiments, the invention provides representations of libraries and/or theoretical segment pools in a computer readable format, for example, the TN1 polypeptides of Tables 10, 23-25 and 26; the DH polypeptides of Tables 11, 23-25 and 28; the N2 polypeptides of Tables 12, 23-25 and 30; the H3-JH polypeptides of Tables

13, 15, 17, 23-25 and 32; the TN1 polynucleotides of Tables 25-26; the DH polynucleotides of Tables 25 and 27; the N2 polynucleotides of Tables 25 and 29; and/or the H3-JH polynucleotides of Tables 25 and 31.

In certain embodiments, the invention provides a representation of the polynucleotide sequences of the Human Preimmune Set (Appendix A), or the polypeptide expression products thereof, in a computer readable format.

In some embodiments, the invention provides a method of making synthetic polynucleotides encoding a CDRH3 library, comprising: (a) providing a theoretical segment pool containing TN1, DH, N2, and H3-JH segments; (b) providing a reference set of CDRH3 sequences; (c) utilizing the theoretical segment pool of (a) to identify the closest match(es) to each CDRH3 sequence in the reference set of (b); (d) selecting segments from the theoretical segment pool for inclusion in a synthetic library; and (e) synthesizing the synthetic CDRH3 library. In certain embodiments, the invention provides libraries made by this method. In some embodiments, the segments selected for inclusion in the synthetic library are selected according to their segment usage weight in the reference set of CDRH3 sequences.

In certain embodiments, the invention provides a method of making synthetic polynucleotides encoding a CDRL3 library, comprising: (i) obtaining a reference set of light chain sequences, wherein the reference set contains light chain sequences with VL segments originating from the same IGVL germline gene and/or its allelic variants; (ii) determining which amino acids occur at each of the CDRL3 positions in the reference set that are encoded by the IGVL gene; (iii) synthesizing light chain variable domain encoding sequences wherein two positions between positions 89 and 94, inclusive, contain degenerate codons encoding two or more of the five most frequently occurring amino acid residues at the corresponding positions in the reference set; and (iv) synthesizing the polynucleotides encoding the CDRL3 library. In certain embodiments, the invention provides libraries made by this method.

In some embodiments, the invention provides a method of using any of the libraries of the invention to isolate an antibody binding an antigen, comprising contacting

the polypeptide expression products of said libraries with an antigen and isolating polypeptide expression products that bind to the antigen.

In certain embodiments, the number of N-linked glycosylation sites, deamidation motifs, and/or Cys residues in the libraries of the invention are reduced or eliminated in comparison to libraries produced by amplification of a repertoire from a biological source.

The invention provides a number of polynucleotide and polypeptide sequences and segments that can be used to build larger polynucleotide and polypeptide sequences (e.g., TN1, DH, N2, and H3-JH segments that can be used to build CDRH3). One of ordinary skill in the art will readily recognize that in some instances these sequences can be more succinctly represented by providing consensus sequences after alignment of the sequences provided by the invention, and that these consensus sequences fall within the scope of the invention and may be used to more succinctly represent any of the sequences provided herein.

BRIEF DESCRIPTION OF THE DRAWING

Figure 1 shows that Vernier residues 4 and 49 (starred) in VK1-39 have a diversity index comparable to or greater than the diversity indices of the CDR positions (*i.e.*, at or above 0.07, in this example).

Figure 2 shows that clinically validated CDRL3 sequences deviate little from germline-like sequences (n=35).

Figure 3 shows the percent of sequences in the jumping dimer CDRL3 libraries of the invention and a previous CDRL3 library, VK-v1.0, with *X* or fewer mutations from germline. Here, *FX* is the percentage of sequences in a library with *X* or fewer mutations from germline.

Figure 4 shows the application of a provided method used to generate nucleotide sequences encoding the parent H3-JH segments.

Figure 5 shows the general schematic of an approach used to select segments from a theoretical segment pool for inclusion in a theoretical and/or synthetic library.

Figure 6 shows the frequency of “Good” and “Poor” expressing CDRH3 sequences isolated from yeast-based libraries described in US 2009/0181855, and their comparison to the sequences contained in the library design described therein (“Design”), as a function of the DH segment hydrophobicity (increasing to the right).

Figure 7 shows the percentage of CDRH3 sequences in the LUA-141 library and Exemplary Library Design 3 (ELD-3) that match CDRH3 sequences from Lee-666 and Boyd-3000 with zero, one, two, three, or more than three amino acid mismatches.

Figure 8 shows that Exemplary Library Design 3 (ELD-3) and the Extended Diversity Library Design both return better matches to clinically relevant CDRH3 sequences than the LUA-141 library.

Figure 9 shows that the combinatorial efficiency of Exemplary Library Design 3 (ELD-3) is greater than that of the LUA-141 library. Specifically, the ELD-3 segments are more likely to yield a unique CDRH3 than the LUA-141 library segments.

Figure 10 shows the amino acid compositions of the Kabat-CDRH3s of LUA-141, Exemplary Library Design 3 (ELD-3), and Human CDRH3 sequences from the HPS (Human H3).

Figure 11 shows the Kabat-CDRH3 length distribution of LUA-141, Exemplary Library Design 3 (ELD-3), and Human CDRH3 sequences from the HPS (Human H3).

Figure 12 shows the percentage of CDRH3 sequences in the Extended Diversity library that match CDRH3 sequences from Boyd *et al.* with zero to thirty-two amino acid mismatches

Figure 13 shows the Kabat-CDRH3 length distribution of Exemplary Library Design 3 (“ELD-3”), the Extended Diversity Library Design (“Extended Diversity”) and human CDRH3 sequences from the Boyd *et al.* data set (“Boyd 2009”).

Figure 14 shows the amino acid compositions of the Kabat-CDRH3s of the Extended Diversity Library Design (“Extended Diversity”) and human CDRH3 sequences from the Boyd *et al.* dataset (“Boyd 2009”).

Figure 15 shows the combinatorial efficiency of the Extended Diversity Library Design by matching 20,000 randomly selected sequences from the (same) design. About 65% of the sequences appear only once in the design and about 17% appear twice.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

The present invention provides, among other things, polynucleotide and polypeptide libraries, methods of producing and using the libraries, kits containing the libraries, and computer readable forms of representations of libraries and/or theoretical segment pools disclosed herein. Libraries taught in this application can be described, at least in part, in terms of components (*e.g.*, polynucleotide or polypeptide “segments”) from which they are assembled. Among other things, the present invention specifically provides and contemplates these polynucleotide or polypeptide segments, methods of producing and using such segments, and kits and computer readable forms of representations that include such library segments.

In certain embodiments, the invention provides antibody libraries specifically designed based on sequences and CDR length distribution in a naturally occurring human antibody repertoire. It is estimated that, even in the absence of antigenic stimulation, an individual human makes at least about 10^7 different antibody molecules (Boyd *et al.*, Science Translational Medicine, 2009, 1: 1). The antigen-binding sites of many antibodies can cross-react with a variety of related but different epitopes. In addition, the human antibody repertoire is large enough to ensure that there is an antigen-binding site to fit almost any potential epitope, albeit potentially with low affinity.

The mammalian immune system has evolved unique genetic mechanisms that enable it to generate an almost unlimited number of different light and heavy chains in a remarkably economical way, by combinatorially joining chromosomally separated gene segments prior to transcription. Each type of immunoglobulin (Ig) chain (*i.e.*, kappa light, lambda light, and heavy) is synthesized by combinatorial assembly of DNA sequences, selected from two or more families of gene segments, to produce a single polypeptide chain. Specifically, the heavy chains and light chains each consist of a variable region and a constant (C) region. The variable regions of the heavy chains are encoded by DNA sequences assembled from three families of gene sequences: variable (IGHV), diversity (IGHD), and joining (IGHJ). The variable regions of light chains are encoded by DNA sequences assembled from two families of gene sequences for each of the kappa and lambda light chains: variable (IGLV) and joining (IGLJ). Each variable

region (heavy and light) is also recombined with a constant region, to produce a full-length immunoglobulin chain.

While combinatorial assembly of the V, D and J gene segments make a substantial contribution to antibody variable region diversity, further diversity is introduced *in vivo*, at the pre-B cell stage, via imprecise joining of these gene segments and the introduction of non-templated nucleotides at the junctions between the gene segments (see *e.g.*, U.S. Pub. No. 2009/0181855, which is incorporated by reference in its entirety, for more information).

After a B cell recognizes an antigen, it is induced to proliferate. During proliferation, the B cell receptor locus undergoes an extremely high rate of somatic mutation that is far greater than the normal rate of genomic mutation. The mutations that occur are primarily localized to the Ig variable regions and comprise substitutions, insertions and deletions. This somatic hypermutation enables the production of B cells that express antibodies possessing enhanced affinity toward an antigen. Such antigen-driven somatic hypermutation fine-tunes antibody responses to a given antigen.

Synthetic antibody libraries of the instant invention have the potential to recognize any antigen, including antigens of human origin. The ability to recognize antigens of human origin may not be present in other antibody libraries, such as antibody libraries prepared from human biological sources (*e.g.*, from human cDNA), because self-reactive antibodies are removed by the donor's immune system via negative selection.

Still further, the present invention provides strategies that streamline and/or simplify certain aspects of library development and/or screening. For example, in some embodiments, the present invention permits use of cell sorting technologies (*e.g.*, fluorescence activated cell sorting, FACS) to identify positive clones, and therefore bypasses or obviates the need for the standard and tedious methodology of generating a hybridoma library and supernatant screening.

Yet further, in some embodiments, the present invention provides libraries and/or sublibraries that accommodate multiple screening passes. For example, in some embodiments, provided libraries and/or sublibraries can be screened multiple times. In

some such embodiments, individual provided libraries and/or sublibraries can be used to discover additional antibodies against many targets.

Before further description of the invention, certain terms are defined.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by one of ordinary skill in the art relevant to the invention. Unless otherwise specified, the Kabat numbering system is used throughout the application. The definitions below supplement those in the art and are directed to the embodiments described in the current application.

The term “amino acid” or “amino acid residue,” as would be understood by one of ordinary skill in the art, typically refers to an amino acid having its art recognized definition such as an amino acid selected from the group consisting of: alanine (Ala or A); arginine (Arg or R); asparagine (Asn or N); aspartic acid (Asp or D); cysteine (Cys or C); glutamine (Gln or Q); glutamic acid (Glu or E); glycine (Gly or G); histidine (His or H); isoleucine (Ile or I); leucine (Leu or L); lysine (Lys or K); methionine (Met or M); phenylalanine (Phe or F); proline (Pro or P); serine (Ser or S); threonine (Thr or T); tryptophan (Trp or W); tyrosine (Tyr or Y); and valine (Val or V), although modified, synthetic, or rare amino acids may be used as desired. Generally, amino acids can be grouped as having a non-polar side chain (*e.g.*, Ala, Cys, Ile, Leu, Met, Phe, Pro, Val); a negatively charged side chain (*e.g.*, Asp, Glu); a positively charged side chain (*e.g.*, Arg, His, Lys); or an uncharged polar side chain (*e.g.*, Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, and Tyr).

As would be understood by those of ordinary skill in the art, the term “antibody” is used herein in the broadest sense and specifically encompasses at least monoclonal antibodies, polyclonal antibodies, multi-specific antibodies (*e.g.*, bispecific antibodies), chimeric antibodies, humanized antibodies, human antibodies, and antibody fragments. An antibody is a protein comprising one or more polypeptides substantially or partially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta,

epsilon and mu constant region genes, as well as myriad immunoglobulin variable region genes.

The term “antibody binding region” refers to one or more portions of an immunoglobulin or antibody variable region capable of binding an antigen(s). Typically, the antibody binding region is, for example, an antibody light chain (or variable region or one or more CDRs thereof), an antibody heavy chain (or variable region or one or more CDRs thereof), a heavy chain Fd region, a combined antibody light and heavy chain (or variable regions thereof) such as a Fab, F(ab')₂, single domain, or single chain antibodies (scFv), or any region of a full length antibody that recognizes an antigen, for example, an IgG (*e.g.*, an IgG1, IgG2, IgG3, or IgG4 subtype), IgA1, IgA2, IgD, IgE, or IgM antibody.

“Antibody fragments” comprise a portion of an intact antibody, for example, one or more portions of the antigen-binding region thereof. Examples of antibody fragments include Fab, Fab', F(ab')₂, and Fv fragments, diabodies, linear antibodies, single-chain antibodies, and multi-specific antibodies formed from intact antibodies and antibody fragments.

The term “antibody of interest” refers to an antibody that has a property of interest that is identified and/or isolated from a library of the invention. Exemplary properties of interest include, for example, but are not limited to, binding to a particular antigen or epitope, binding with a certain affinity, cross-reactivity, blocking a binding interaction between two molecules, and/or eliciting a certain biological effect.

The term “canonical structure,” as understood by those of ordinary skill in the art, refers to the main chain conformation that is adopted by the antigen binding (CDR) loops. From comparative structural studies, it has been found that five of the six antigen binding loops have only a limited repertoire of available conformations. Each canonical structure can be characterized by the torsion angles of the polypeptide backbone. Correspondent loops between antibodies may, therefore, have very similar three dimensional structures, despite high amino acid sequence variability in most parts of the loops (Chothia and Lesk, J. Mol. Biol., 1987, 196: 901; Chothia *et al.*, Nature, 1989, 342: 877; Martin and Thornton, J. Mol. Biol., 1996, 263: 800, each of which is incorporated by reference in its entirety). Furthermore, there is a relationship between the adopted loop

structure and the amino acid sequences surrounding it. As is known in the art, the conformation of a particular canonical class is determined by the length of the loop and the amino acid residues residing at key positions within the loop, as well as within the conserved framework (*i.e.*, outside of the loop). Assignment to a particular canonical class can therefore be made based on the presence of these key amino acid residues. The term “canonical structure” may also include considerations as to the linear sequence of the antibody, for example, as catalogued by Kabat (Kabat *et al.*, in “Sequences of Proteins of Immunological Interest,” 5th Edition, U.S. Department of Health and Human Services, 1992). The Kabat numbering scheme is a widely adopted standard for numbering the amino acid residues of an antibody variable domain in a consistent manner, and is used herein unless indicated otherwise. Additional structural considerations can also be used to determine the canonical structure of an antibody. For example, those differences not fully reflected by Kabat numbering can be described by the numbering system of Chothia *et al.* and/or revealed by other techniques, for example, crystallography and two or three-dimensional computational modeling. Accordingly, a given antibody sequence may be placed into a canonical class which allows for, among other things, identifying appropriate chassis sequences (*e.g.*, based on a desire to include a variety of canonical structures in a library). Kabat numbering of antibody amino acid sequences and structural considerations as described by Chothia *et al.*, and their implications for construing canonical aspects of antibody structure, are described in the literature.

The terms “CDR”, and its plural “CDRs”, refer to a complementarity determining region (CDR) of which three make up the binding character of a light chain variable region (CDRL1, CDRL2 and CDRL3) and three make up the binding character of a heavy chain variable region (CDRH1, CDRH2 and CDRH3). CDRs contribute to the functional activity of an antibody molecule and are separated by amino acid sequences that comprise framework regions. The exact definitional CDR boundaries and lengths are subject to different classification and numbering systems. CDRs may therefore be referred to by Kabat, Chothia, contact or other boundary definitions, including for example the CDRH3 numbering system described below. Despite differing boundaries, each of these systems has some degree of overlap in what constitutes the so called

"hypervariable regions" within the variable region. CDR definitions according to these systems may therefore differ in length and boundary areas with respect to the adjacent framework region. See, for example Kabat *et al.*, in "Sequences of Proteins of Immunological Interest," 5th Edition, U.S. Department of Health and Human Services, 1992; Chothia *et al.*, J. Mol. Biol., 1987, 196: 901; and MacCallum *et al.*, J. Mol. Biol., 1996, 262: 732, each of which is incorporated by reference in its entirety.

The "CDRH3 numbering system" used herein defines the first amino acid of CDRH3 as being at position 95 and the last amino acid of CDRH3 as position 102. Note that this is a custom numbering system that is not according to Kabat. The amino acid segment, beginning at position 95 is called "TN1" and, when present, is assigned numbers 95, 96, 96A, 96B, etc. Note that the nomenclature used in the current application is slightly different from that used in U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379. In those applications, position 95 was designated a "Tail" residue, while here, the Tail (T) has been combined with the N1 segment, to produce one segment, designated "TN1." The TN1 segment is followed by the "DH" segment, which is assigned numbers 97, 97A, 97B, 97C, etc. The DH segment is followed by the "N2" segment, which, when present, is numbered 98, 98A, 98B, etc. Finally, the most C-terminal amino acid residue of the "H3-JH" segment is designated as number 102. The residue directly before (N-terminal) it, when present, is 101, and the one before (if present) is 100. The rest of the H3-JH amino acids are numbered in reverse order, beginning with 99 for the amino acid just N-terminal to 100, 99A for the residue N-terminal to 99, and so forth for 99B, 99C, etc. Examples of CDRH3 sequence residue numbers may therefore include the following:

13 Amino Acid CDR-H3 with TN1 and N2

(95)	(96)	(96A)	(97)	(97A)	(97B)	(97C)	(97D)	(98)	(99)	(100)	(101)	(102)
-----				-----				-----		-----		
TN1				DH				N2		H3-JH		

10 Amino Acid CDR-H3 without TN1 and N2

(97) (97A) (97B) (97C) (97D) (97E) (97F) (97G) (101) (102)
 |-----|-----|
 DH H3-JH

“Chassis” of the invention are portions of the antibody heavy chain variable (IGHV) or light chain variable (IGLV) domains that are not part of CDRH3 or CDRL3, respectively. A chassis of the invention is defined as the portion of the variable region of an antibody beginning with the first amino acid of FRM1 and ending with the last amino acid of FRM3. In the case of the heavy chain, the chassis includes the amino acids including from position 1 to position 94. In the case of the light chains (kappa and lambda), the chassis are defined as including from position 1 to position 88. The chassis of the invention may contain certain modifications relative to the corresponding germline variable domain sequences. These modifications may be engineered (*e.g.*, to remove N-linked glycosylation sites) or naturally occurring (*e.g.*, to account for naturally occurring allelic variation). For example, it is known in the art that the immunoglobulin gene repertoire is polymorphic (Wang *et al.*, Immunol. Cell. Biol., 2008, 86: 111; Collins *et al.*, Immunogenetics, 2008, 60: 669, each incorporated by reference in its entirety); chassis, CDRs and constant regions representative of these allelic variants are also encompassed by the invention. In some embodiments, the allelic variant(s) used in a particular embodiment of the invention may be selected based on the allelic variation present in different patient populations, for example, to identify antibodies that are non-immunogenic in these patient populations. In certain embodiments, the immunogenicity of an antibody of the invention may depend on allelic variation in the major histocompatibility complex (MHC) genes of a patient population. Such allelic variation may also be considered in the design of libraries of the invention. In certain embodiments of the invention, the chassis and constant regions are contained in a vector, and a CDR3 region is introduced between them via homologous recombination.

As used herein, a sequence designed with “directed diversity” has been specifically designed to contain both sequence diversity and length diversity. Directed diversity is not stochastic.

As used herein, the term “diversity” refers to a variety or a noticeable heterogeneity. The term “sequence diversity” refers to a variety of sequences which are collectively representative of several possibilities of sequences, for example, those found in natural human antibodies. For example, CDRH3 sequence diversity may refer to a variety of possibilities of combining the known human TN1, DH, N2, and H3-JH segments to form CDRH3 sequences. The CDRL3 sequence diversity (kappa or lambda) may refer to a variety of possibilities of combining the naturally occurring light chain variable region contributing to CDRL3 (*i.e.*, “L3-VL”) and joining (*i.e.*, “L3-JL”) segments, to form CDRL3 sequences. As used herein, “H3-JH” refers to the portion of the IGHJ gene contributing to CDRH3. As used herein, “L3-VL” and “L3-JL” refer to the portions of the IGLV and IGLJ genes (kappa or lambda) contributing to CDRL3, respectively.

As used herein, the term “expression” refers to steps involved in the production of a polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

The term “framework region” refers to the art-recognized portions of an antibody variable region that exist between the more divergent (*i.e.*, hypervariable) CDRs. Such framework regions are typically referred to as frameworks 1 through 4 (FRM1, FRM2, FRM3, and FRM4) and provide a scaffold for the presentation of the six CDRs (three from the heavy chain and three from the light chain) in three dimensional space, to form an antigen-binding surface.

The term “full-length heavy chain” refers to an immunoglobulin heavy chain that contains each of the canonical structural domains of an immunoglobulin heavy chain, including the four framework regions, the three CDRs, and the constant region.

The term “full-length light chain” refers to an immunoglobulin light chain that contains each of the canonical structural domains of an immunoglobulin light chain, including the four framework regions, the three CDRs, and the constant region.

The term “germline-like,” when used with respect to the CDRL3 sequences of the light chains of the invention, means those sequences consisting of combinations of: (i) the first six wild-type residues contributed to CDRL3 by the IGVJ germline gene (*i.e.*, positions 89 to 94 in the Kabat numbering system; “L” is kappa or lambda); and (ii) one

of several amino acid sequences, two one to four amino acids in length, largely, but not exclusively, derived from the JL segment ("L," again is kappa or lambda). For kappa CDRL3 sequences of the most common lengths (*i.e.*, 8, 9, and 10 residues), the sequences of (ii) number twenty and are: FT, LT, IT, RT, WT, YT, [X]T, [X]PT, [X]FT, [X]LT, [X]IT, [X]RT, [X]WT, [X]YT, [X]PFT, [X]PLT, [X]PIT, [X]PRT, [X]PWT and [X]PYT, where [X] corresponds to the amino acid residue found at position 95 (Kabat) in the respective VK germline sequence. X is most commonly P, but may also be S or any other amino acid residue found at position 95 of a VK germline sequence. For eight exemplified VK chassis exemplified herein, the corresponding 160 germline-like sequences, (*i.e.*, 20 sequences of two to four amino acids in length combined with positions 89 to 94 of each of eight VK germline sequences) are provided in Table 1. A similar approach is applied to define germline-like CDRL3 sequences for lambda light chains. As for the kappa sequences described above, the intact, un-mutated portions of CDRL3 encoded by the IGVL genes (in this case, IGV λ) would be combined with the sequences largely, but not exclusively, derived from the J λ segment. Here, the latter sequences (corresponding to (ii), above), number five and are: YV, VV, WV, AV or V. In addition, and as described in Example 7 of US 2009/0818155, one could further allow for variation at the last position of the V λ -gene-encoded portion of CDRL3 by considering partial codons, while still considering the resulting sequences "germline-like." More specifically, the entire "minimalist library" of Example 7 in US 2009/0818155 would be defined as "germline-like." One of ordinary skill in the art will readily recognize that these methods can be extended to other VK and V λ sequences.

The term "genotype-phenotype linkage," as understood by those of ordinary skill in the art, refers to the fact that the nucleic acid (genotype) encoding a protein with a particular phenotype (*e.g.*, binding an antigen) can be isolated from a library. For the purposes of illustration, an antibody fragment expressed on the surface of a phage can be isolated based on its binding to an antigen (*e.g.*, U.S. Patent No. 5,837,500). The binding of the antibody to the antigen simultaneously enables the isolation of the phage containing the nucleic acid encoding the antibody fragment. Thus, the phenotype (antigen-binding characteristics of the antibody fragment) has been "linked" to the genotype (nucleic acid encoding the antibody fragment). Other methods of maintaining a

genotype-phenotype linkage include those of Wittrup *et al.* (U.S. Patent Nos. 6,300,065, 6,331,391, 6,423,538, 6,696,251, 6,699,658, and U.S. Pub. No. 20040146976, each of which is incorporated by reference in its entirety), Miltenyi (US Patent No. 7,166,423, incorporated by reference in its entirety), Fandl (U.S. Patent No. 6,919,183, US Pub No. 20060234311, each incorporated by reference in its entirety), Clausell-Tormos *et al.* (Chem. Biol., 2008, 15: 427, incorporated by reference in its entirety), Love *et al.* (Nat. Biotechnol., 2006, 24: 703, incorporated by reference in its entirety), and Kelly *et al.* (Chem. Commun., 2007, 14: 1773, incorporated by reference in its entirety). The term can be used to refer to any method which localizes an antibody protein together with the gene encoding the antibody protein, in a way in which they can both be recovered while the linkage between them is maintained.

The term “heterologous moiety” is used herein to indicate the addition of a moiety to an antibody wherein the moiety is not part of a naturally-occurring antibody. Exemplary heterologous moieties include drugs, toxins, imaging agents, and any other compositions which might provide an activity that is not inherent in the antibody itself.

As used herein, the term “host cell” is intended to refer to a cell comprising a polynucleotide of the invention. It should be understood that such terms refer not only to the particular subject cell but to the progeny or potential 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 as used herein.

As used herein, the term “human antibody CDR library” includes at least a polynucleotide or polypeptide library which has been designed to represent the sequence diversity and length diversity of naturally occurring CDRs in human antibodies (*e.g.*, the term “CDR” in “human antibody CDR library” may be substituted with “CDRL1,” “CDRL2,” “CDRL3,” “CDRH1,” “CDRH2,” and/or “CDRH3”). Known human CDR sequences are represented in various data sets, including Jackson *et al.*, J. Immunol Methods, 2007, 324: 26; Martin, Proteins, 1996, 25: 130; Lee *et al.*, Immunogenetics, 2006, 57: 917, Boyd *et al.*, Science Translational Medicine, 2009, 1: 1, and WO/2009/036379, each of which is incorporated by reference in its entirety, and the HPS, which is provided in Appendix A.

The term “Human Preimmune Set,” or “HPS,” refers to a reference set of 3,571 curated human preimmune heavy chain sequences corresponding to the GI Nos. provided in Appendix A.

An "intact antibody" is one comprising full-length heavy- and light- chains (*i.e.*, four frameworks, three CDRs, and a constant region for each of the heavy and light chains). An intact antibody is also referred to as a “full-length” antibody.

The term “length diversity” refers to a variety in the length of a family of nucleotide or amino acid sequence. For example, in naturally occurring human antibodies, the heavy chain CDR3 sequence varies in length, for example, from about 2 amino acids to over about 35 amino acids, and the light chain CDR3 sequence varies in length, for example, from about 5 to about 16 amino acids.

The term “library” refers to a set of entities comprising two or more entities having diversity as described herein, and/or designed according to the methods of the invention. For example, a “library of polynucleotides” refers to a set of polynucleotides comprising two or more polynucleotides having diversity as described herein, and/or designed according to the methods of the invention. A “library of polypeptides” refers to a set of polypeptides comprising two or more polypeptides having diversity as described herein, and/or designed according to the methods of the invention. A “library of synthetic polynucleotides” refers to a set of polynucleotides comprising two or more synthetic polynucleotides having diversity as described herein, and/or designed according to the methods of the invention. Libraries where all members are synthetic are also encompassed by the invention. A “human antibody library” refers to a set of polypeptides comprising two or more polypeptides having diversity as described herein, and/or designed according to the methods of the invention, for example a library designed to represent the sequence diversity and length diversity of naturally occurring human antibodies. In some embodiments, the term “library” may refer to a set of entities sharing similar structural or sequence characteristics, for example, a “heavy chain library,” “light chain library,” “antibody library,” and/or “CDRH3 library.”

The term “physical realization” refers to a portion of a theoretical (*e.g.*, computer-based) or synthetic (*e.g.*, oligonucleotide-based) diversity that can actually be physically sampled, for example, by any display methodology. Exemplary display methodology

include: phage display, ribosomal display, and yeast display. For synthetic sequences, the size of the physical realization of a library depends on (1) the fraction of the theoretical diversity that can actually be synthesized, and (2) the limitations of the particular screening method. Exemplary limitations of screening methods include the number of variants that can be screened in a particular assay (*e.g.*, ribosome display, phage display, yeast display) and the transformation efficiency of the host cells (*e.g.*, yeast, mammalian cells, bacteria) which are used in a screening assay. For the purposes of illustration, given a library with a theoretical diversity of 10^{12} members, an exemplary physical realization of the library (*e.g.*, in yeast, bacterial cells, or ribosome display) that can maximally include 10^{11} members will, therefore, sample about 10% of the theoretical diversity of the library. However, if fewer than 10^{11} members of the library with a theoretical diversity of 10^{12} are synthesized, and the physical realization of the library can maximally include 10^{11} members, less than 10% of the theoretical diversity of the library is sampled in the physical realization of the library. Similarly, a physical realization of the library that can maximally include more than 10^{12} members would "oversample" the theoretical diversity, meaning that each member may be present more than once (assuming that the entire 10^{12} theoretical diversity is synthesized).

The term "polynucleotide(s)" refers to nucleic acids such as DNA molecules and RNA molecules and analogs thereof (*e.g.*, DNA or RNA generated using nucleotide analogs or using nucleic acid chemistry). As desired, the polynucleotides may be made synthetically, *e.g.*, using art-recognized nucleic acid chemistry or enzymatically using, *e.g.*, a polymerase, and, if desired, be modified. Typical modifications include methylation, biotinylation, and other art-known modifications. In addition, the nucleic acid molecule can be single-stranded or double-stranded and, where desired, linked to a detectable moiety. The representation of nucleotide bases herein follows International Union of Pure and Applied Chemistry (IUPAC) nomenclature (see U.S. Pub. No. 2009/0181855, incorporated by reference in its entirety).

"Preimmune" antibody libraries have sequence diversities and length diversities similar to naturally occurring human antibody sequences before these sequences have undergone negative selection and/or somatic hypermutation. For example, the set of sequences described in Lee *et al.* (Immunogenetics, 2006, 57: 917, incorporated by

reference in its entirety) and the Human Preimmune Set (HPS) described herein (see Appendix A) are believed to represent sequences from the preimmune repertoire. In certain embodiments of the invention, the sequences of the invention will be similar to these sequences (*e.g.*, in terms of composition and length).

As used herein, the term “sitewise stochastic” describes a process of generating a sequence of amino acids, where only the amino acid occurrences at the individual positions are considered, and higher order motifs (*e.g.*, pair-wise correlations) are not accounted for (*e.g.*, see Knappik, *et al.*, J Mol Biol, 2000, 296: 57, and analysis provided in U.S. Publication No. 2009/0181855, each incorporated by reference in its entirety).

The term “split-pool synthesis” refers to a procedure in which the products of a plurality of individual first reactions are combined (pooled) and then separated (split) before participating in a plurality of second reactions. For example, U.S. Publication No. 2009/0181855 (incorporated by reference in its entirety) describes the synthesis of 278 DH segments (products), each in a separate reaction. After synthesis, these 278 segments are combined (pooled) and then distributed (split) amongst 141 columns for the synthesis of the N2 segments. This enables the pairing of each of the 278 DH segments with each of 141 N2 segments.

As used herein, “stochastic” describes a process of generating a random sequence of nucleotides or amino acids, which is considered as a sample of one element from a probability distribution (*e.g.*, see U.S. Patent No. 5,723,323).

As used herein, the term “synthetic polynucleotide” refers to a molecule formed through a chemical process, as opposed to molecules of natural origin, or molecules derived via template-based amplification of molecules of natural origin (*e.g.*, immunoglobulin chains cloned from populations of B cells via PCR amplification are not “synthetic” as used herein). In some instances, for example, when referring to libraries of the invention that comprise multiple segments (*e.g.*, TN1, DH, N2, and/or H3-JH), the invention encompasses libraries in which at least one, two, three, or four of the aforementioned components is synthetic. By way of illustration, a library in which certain components are synthetic, while other components are of natural origin or derived via template-based amplification of molecules of natural origin, would be encompassed

by the invention. Libraries that are fully synthetic would, of course, also be encompassed by the invention.

The term “theoretical diversity” refers to the maximum number of variants in a library design. For example, given an amino acid sequence of three residues, where residues one and three may each be any one of five amino acid types and residue two may be any one of 20 amino acid types, the theoretical diversity is $5 \times 20 \times 5 = 500$ possible sequences. Similarly if sequence X is constructed by combination of 4 amino acid segments, where segment 1 has 100 possible sequences, segment 2 has 75 possible sequences, segment 3 has 250 possible sequences, and segment 4 has 30 possible sequences, the theoretical diversity of fragment X would be $100 \times 75 \times 200 \times 30$, or 5.6×10^5 possible sequences.

The term “theoretical segment pool” refers to a set of polynucleotide or polypeptide segments that can be used as building blocks to assemble a larger polynucleotide or polypeptide. For example, a theoretical segment pool containing TN1, DH, N2, and H3-JH segments can be used to assemble a library of CDRH3 sequences by concatenating them combinatorially to form a sequence represented by [TN1]-[DH]-[N2]-[H3-JH], and synthesizing the corresponding oligonucleotide(s). The term “theoretical segment pool” can apply to any set of polynucleotide or polypeptide segments. Thus, while a set of TN1, DH, N2, and H3-JH segments are collectively considered a theoretical segment pool, each of the individual sets of segments also comprise a theoretical segment pool, specifically a TN1 theoretical segment pool, a DH theoretical segment pool, an N2 theoretical segment pool, and an H3-JH theoretical segment pool. Any subsets of these theoretical segment pools containing two or more sequences can also be considered theoretical segment pools.

The term “unique,” as used herein, refers to a sequence that is different (*e.g.*, has a different chemical structure) from every other sequence within the designed set (*e.g.*, the theoretical diversity). It should be understood that there are likely to be more than one copy of many unique sequences from the theoretical diversity in a particular physical realization. For example, a library comprising three unique sequences at the theoretical level may comprise nine total members if each sequence occurs three times in the

physical realization of the library. However, in certain embodiments, each unique sequence may occur only once, less than once, or more than once.

The term "variable" refers to the portions of the immunoglobulin domains that exhibit variability in their sequence and that are involved in determining the specificity and binding affinity of a particular antibody (*i.e.*, the "variable domain(s)"). Variability is not evenly distributed throughout the variable domains of antibodies; it is concentrated in sub-domains of each of the heavy and light chain variable regions. These sub-domains are called "hypervariable" regions or "complementarity determining regions" (CDRs). The more conserved (*i.e.*, non-hypervariable) portions of the variable domains are called the "framework" regions (FRM). The variable domains of naturally occurring heavy and light chains each comprise four FRM regions, largely adopting a β -sheet configuration, connected by three hypervariable regions, which form loops connecting, and in some cases forming part of, the β -sheet structure. The hypervariable regions in each chain are held together in close proximity by the FRM and, with the hypervariable regions from the other chain, contribute to the formation of the antigen-binding site (see Kabat *et al.* Sequences of Proteins of Immunological Interest, 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md., 1991, incorporated by reference in its entirety). The constant domains are not directly involved in antigen binding, but exhibit various effector functions, such as, for example, antibody-dependent, cell-mediated cytotoxicity and complement activation.

Libraries of the invention containing "VKCDR3" sequences and "V λ CDR3" sequences refer to the kappa and lambda sub-sets of the light chain CDR3 (CDRL3) sequences, respectively. Such libraries may be designed with directed diversity, to collectively represent the length and sequence diversity of the human antibody CDRL3 repertoire. "Preimmune" versions of these libraries have similar sequence diversities and length diversities as naturally occurring human antibody CDRL3 sequences before these sequences undergo negative selection and/or somatic hypermutation. Known human CDRL3 sequences are represented in various data sets, including the NCBI database, WO/2009/036379, and Martin, Proteins, 1996, 25: 130 each incorporated by reference in its entirety.

General Design of Libraries

Antibody libraries provided by the present invention may be designed to reflect certain aspects of the preimmune repertoire as created by the human immune system. Certain libraries of the invention are based on rational design informed by collections of human V, D, and J genes, and large databases of human heavy and light chain sequences (*e.g.*, publicly known germline sequences and sequences from Jackson *et al.*, J. Immunol Methods, 2007, 324: 26; Lee *et al.*, Immunogenetics, 2006, 57: 917; Boyd *et al.*, Science Translational Medicine, 2009, 1: 1-8, each incorporated by reference in its entirety; and sequences compiled from rearranged VK and V λ sequences (see WO/2009/036379, also incorporated by reference in its entirety). Additional information may be found, for example, in Scaviner *et al.*, Exp. Clin. Immunogenet., 1999, 16: 234; Tomlinson *et al.*, J. Mol. Biol., 1992, 227: 799; and Matsuda *et al.*, J. Exp. Med., 1998, 188: 2151, each incorporated by reference in its entirety.

In certain embodiments of the invention, segments representing the possible V, D, and J diversity found in the human repertoire, as well as junctional diversity (*i.e.*, TN1 and N2), are synthesized *de novo* as single or double-stranded DNA oligonucleotides. In certain embodiments of the invention, oligonucleotides encoding CDR sequences are introduced into yeast along with one or more acceptor vectors containing heavy or light chain chassis sequences and constant domains. No primer-based PCR amplification or template-directed cloning steps from mammalian cDNA or mRNA are employed. Through standard homologous recombination, the recipient yeast recombines the CDR segments with the acceptor vectors containing the chassis sequences and constant regions, to create a properly ordered synthetic, full-length human heavy chain and/or light chain immunoglobulin library that can be genetically propagated, expressed, presented, and screened. One of ordinary skill in the art will readily recognize that the acceptor vector can be designed so as to produce constructs other than full-length human heavy chains and/or light chains. For example, in certain embodiments of the invention, the chassis may be designed to encode portions of a polypeptide encoding an antibody fragment or subunit of an antibody fragment, so that a sequence encoding an antibody fragment, or subunit thereof, is produced when the oligonucleotide cassette containing the CDR is recombined with the acceptor vector.

Thus, in certain embodiments, the invention provides a synthetic, preimmune human antibody repertoire the repertoire comprising:

(a) one or more selected human antibody heavy chain chassis (*i.e.*, amino acids 1 to 94 of the heavy chain variable region, using Kabat's definition);

(b) a CDRH3 repertoire (described more fully below), designed based on the human IGHD and IGHJ germline sequences, and the extraction of TN1 and N2 sequences from reference sets of human CDRH3 sequences, the CDRH3 repertoire comprising (i) a TN1 segment; (ii) a DH segment; (iii) an N2 segment; (iv) an H3-JH segment.

(c) one or more selected human antibody kappa and/or lambda light chain chassis; and

(d) a CDRL3 repertoire designed based on the human IGLV and IGLJ germline sequences, wherein "L" may be a kappa or lambda light chain.

The instant invention also provides methods for producing and using such libraries, as well as libraries comprising one or more immunoglobulin domains or antibody fragments. Design and synthesis of each component of the antibody libraries of the invention is provided in more detail below.

Design of Antibody Library Chassis Sequences

In certain embodiments, provided libraries are constructed from selected chassis sequences that are based on naturally occurring variable domain sequences (*e.g.*, IGHV and IGLV genes). The selection of such chassis sequences can be done arbitrarily, or through the definition of certain pre-determined criteria. For example, the Kabat database, an electronic database containing non-redundant rearranged antibody sequences, can be queried for those heavy and light chain germline sequences that are most frequently represented. An algorithm such as BLAST, or a more specialized tool such as SoDA (Volpe *et al.*, Bioinformatics, 2006, 22: 438-44, incorporated by reference in its entirety), can be used to compare rearranged antibody sequences with germline sequences (*e.g.*, using the V BASE2 database; see, for example, Retter *et al.*, Nucleic Acids Res., 2005, 33: D671-D674, incorporated by reference in its entirety), or similar collections of human V, D, and J genes, to identify germline families that are most frequently used to generate functional antibodies.

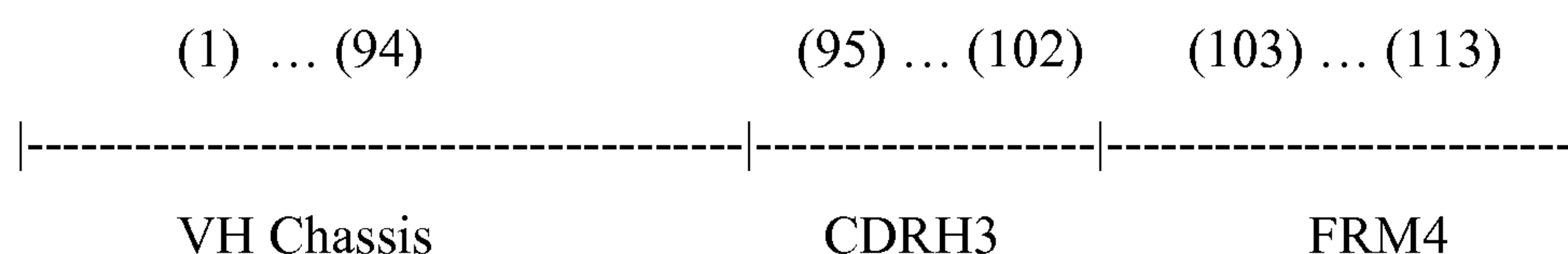
Several criteria can be utilized for the selection of chassis for inclusion in the libraries of the invention. For example, sequences that are known (or have been determined) to express poorly in yeast, or other organisms used in the invention (*e.g.*, bacteria, mammalian cells, fungi, or plants) can be excluded from the libraries. Chassis may also be chosen based on the representation of their corresponding germline genes in the peripheral blood of humans. In certain embodiments of the invention, it may be desirable to select chassis that correspond to germline sequences that are highly represented in the peripheral blood of humans. In some embodiments, it may be desirable to select chassis that correspond to germline sequences that are less frequently represented, for example, to increase the canonical diversity of the library. Therefore, chassis may be selected to produce libraries that represent the largest and most structurally diverse group of functional human antibodies.

In certain embodiments of the invention, less diverse chassis may be utilized, for example, if it is desirable to produce a smaller, more focused library with less chassis variability and greater CDR variability. In some embodiments of the invention, chassis may be selected based on both their expression in a cell of the invention (*e.g.*, a yeast cell) and the diversity of canonical structures represented by the selected sequences. One may therefore produce a library with a diversity of canonical structures that express well in a cell of the invention.

Design of Heavy Chain Chassis Sequences

The design and selection of heavy chain chassis sequences that can be used in the current invention is described in detail in U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379, each of which is incorporated by reference in its entirety, and is therefore described only briefly here.

In general, VH domains of the library comprise three components: (1) a VH “chassis,” which includes amino acids 1 to 94 (using Kabat numbering), (2) the CDRH3, which is defined herein to include the Kabat CDRH3 proper (positions 95-102), and (3) the FRM4 region, including amino acids 103 to 113 (Kabat numbering). The overall VH domain structure may therefore be depicted schematically (not to scale) as:

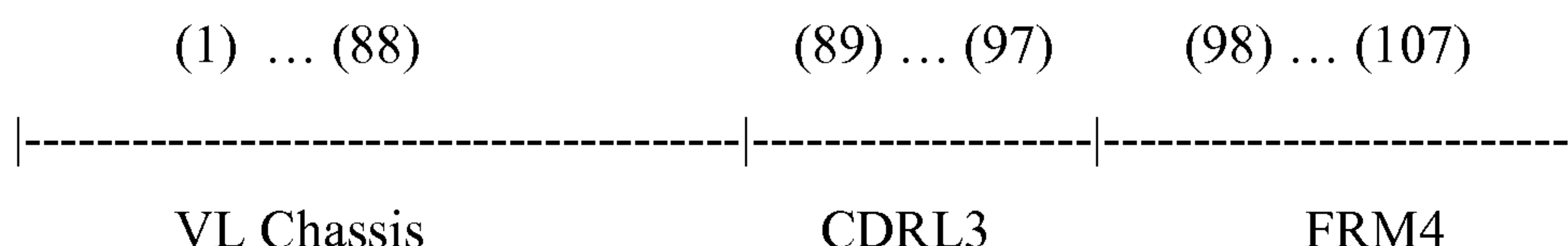


In certain embodiments of the invention, the VH chassis of the libraries may comprise from about Kabat residue 1 to about Kabat residue 94 of one or more of the following IGHV germline sequences: IGHV1-2, IGHV1-3, IGHV1-8, IGHV1-18, IGHV1-24, IGHV1-45, IGHV1-46, IGHV1-58, IGHV1-69, IGH8, IGH56, IGH100, IGHV3-7, IGHV3-9, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-43, IGHV3-48, IGHV3-49, IGHV3-53, IGHV3-64, IGHV3-66, IGHV3-72, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-28, IGHV4-31, IGHV4-34, IGHV4-39, IGHV4-59, IGHV4-61, IGHV4-B, IGHV5-51, IGHV6-1, and/or IGHV7-4-1. In some embodiments of the invention, a library may contain one or more of these sequences, one or more allelic variants of these sequences, or encode an amino acid sequence at least about 99.9%, 99.5%, 99%, 98.5%, 98%, 97.5%, 97%, 96.5%, 96%, 95.5%, 95%, 94.5%, 94%, 93.5%, 93%, 92.5%, 92%, 91.5%, 91%, 90.5%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 77.5%, 75%, 73.5%, 70%, 65%, 60%, 55%, or 50% identical to one or more of these sequences. One of ordinary skill in the art will recognize that given the chassis definition provided above, any IGHV-encoding sequence can be adapted for use as a chassis of the invention. As exemplified in U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379 (each incorporated by reference in its entirety), these chassis can also be varied, particularly by altering the amino acid residues in the CDRH1 and CDRH2 regions, further increasing the diversity of the library.

Design of Light Chain Chassis Sequences

The design and selection of light chain chassis sequences that can be used in the current invention is described in detail in U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379, each of which is incorporated by reference in its entirety, and is therefore described only briefly here. The light chain chassis of the invention may be based on kappa and/or lambda light chain sequences.

The VL domains of the library comprise three primary components: (1) a VL “chassis”, which includes amino acids 1 to 88 (using Kabat numbering), (2) the CDRL3, which is defined herein to include the Kabat CDRL3 proper (positions 89-97), and (3) the FRM4 region, including amino acids 98 to 107 (Kabat numbering). The overall VL domain structure may therefore be depicted schematically (not to scale) as:



In certain embodiments of the invention, the VL chassis of the libraries include one or more chassis based on IGKV germline sequences. In certain embodiments of the invention, the VL chassis of the libraries may comprise from about Kabat residue 1 to about Kabat residue 88 of one or more of the following IGKV germline sequences: IGKV1-05, IGKV1-06, IGKV1-08, IGKV1-09, IGKV1-12, IGKV1-13, IGKV1-16, IGKV1-17, IGKV1-27, IGKV1-33, IGKV1-37, IGKV1-39, IGKV1D-16, IGKV1D-17, IGKV1D-43, IGKV1D-8, IGK54, IGK58, IGK59, IGK60, IGK70, IGKV2D-26, IGKV2D-29, IGKV2D-30, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-07, IGKV3D-11, IGKV3D-20, IGKV4-1, IGKV5-2, IGKV6-21, and/or IGKV6D-41. In some embodiments of the invention, a library may contain one or more of these sequences, one or more allelic variants of these sequences, or encode an amino acid sequence at least about 99.9%, 99.5%, 99%, 98.5%, 98%, 97.5%, 97%, 96.5%, 96%, 95.5%, 95%, 94.5%, 94%, 93.5%, 93%, 92.5%, 92%, 91.5%, 91%, 90.5%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 77.5%, 75%, 73.5%, 70%, 65%, 60%, 55%, or 50% identical to one or more of these sequences.

In certain embodiments of the invention, the VL chassis of the libraries include one or more chassis based on IGλV germline sequences. In certain embodiments of the invention, the VL chassis of the libraries may comprise from about Kabat residue 1 to about Kabat residue 88 of one or more of the following IGλV germline sequences: IGλV3-1, IGλV3-21, IGλ44, IGλV1-40, IGλV3-19, IGλV1-51, IGλV1-44, IGλV6-57, IGλ11, IGλV3-25, IGλ53, IGλV3-10, IGλV4-69, IGλV1-47, IGλ41, IGλV7-43, IGλV7-

46, IG λ V5-45, IG λ V4-60, IG λ V10-54, IG λ V8-61, IG λ V3-9, IG λ V1-36, IG λ 48, IG λ V3-16, IG λ V3-27, IG λ V4-3, IG λ V5-39, IG λ V9-49, and/or IG λ V3-12. In some embodiments of the invention, a library may contain one or more of these sequences, one or more allelic variants of these sequences, or encode an amino acid sequence at least about 99.9%, 99.5%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to one or more of these sequences.

One of ordinary skill in the art will recognize that given the chassis definition provided above, any IGKV- or IG λ V-encoding sequence can be adapted for use as a chassis of the invention.

Design and Selection of TN1, DH, N2, and H3-JH Segments

The human germline repertoire contains at least six IGHJ genes (IGHJ1, IGHJ2, IGHJ3, IGHJ4, IGHJ5, and IGHJ6; included in Table 14, where the primary allele is designated “01,” and selected allelic variants are designated “02” or “03”), and at least 27 IGHD genes (Table 16, including allelic variants). In some embodiments, the invention comprises a library of CDRH3 polypeptide sequences, or polynucleotide sequences encoding CDRH3 sequences, the library comprising members of any of the theoretical segment pools disclosed herein.

A person of ordinary skill in the art will recognize that not every segment in a theoretical segment pool provided herein is necessary to produce a functional CDRH3 library of the invention. Therefore, in certain embodiments, a CDRH3 library of the invention will contain a subset of the segments of any of the theoretical segment pools described herein. For example, in certain embodiments of the invention, at least about 15, 30, 45, 60, 75, 90, 100, 105, 120, 135, 150, 165, 180, 195, 200, 210, 225, 240, 255, 270, 285, 300, 320, 340, 360, 380, 400, 420, 440, 460, 480, 500, 520, 540, 560, 580, 600, 620, 640, or 643 of the H3-JH segments of any of the theoretical segment pools provided herein, or generated by the methods described herein, are included in a library. In some embodiments of the invention, at least about 15, 30, 45, 60, 75, 90, 100, 105, 120, 135, 150, 165, 180, 195, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1111, 2000, 3000, 4000, 5000, 6000, 7000, 14000, 21000, 28000, 35000, 42000, 49000, 56000, 63000, or 68374 of the DH segments of any of the

theoretical segment pools provided herein, or generated by the methods described herein, are included in a library. In some embodiments of the invention, at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 141, 150, 160, 170, 180, 190, or 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 420, 424, 440, 460, 480, 500, 550, 600, 650, 700, 727, 750, 800, 850, 900, 950, or 1000 of the TN1 and/or N2 segments of any of the theoretical segment pools provided herein, or generated by the methods described herein, are included in a library. In certain embodiments, a library of the invention may contain less than a particular number of polynucleotide or polypeptide segments, where the number of segments is defined using any one of the integers provided above for the respective segment. In some embodiments of the invention, a particular numerical range is defined, using any two of the integers provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the integers provided, which define an upper and lower boundary, are contemplated.

In certain embodiments, the invention provides CDRH3 libraries comprising at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the segments from any of the theoretical segment pools provided herein. For example, the invention provides libraries comprising at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the TN1, DH, N2, and/or H3-JH segments from any of the theoretical segment pools provided herein. In some embodiments of the invention, a particular percentage range is defined, using any two of the percentages provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the percentages provided, which define an upper and lower boundary, are contemplated.

In some embodiments of the invention, at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the H3-JH, DH, TN1, and/or N2 segments in a CDRH3 library are H3-JH, DH, TN1, and/or N2 segments of any of the theoretical segment pools provided herein, or generated by the methods described herein,. In some embodiments of the invention, at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the H3-JH, DH, TN1, and/or N2

segments of antibodies isolated from a CDRH3 library (*e.g.*, by binding to a particular antigen and/or generic ligand through one or more rounds of selection) are H3-JH, DH, TN1, and/or N2 segments of any of the theoretical segment pools provided herein, or generated by methods described herein. In certain embodiments, a CDRH3 library of the invention may contain less than a particular percentage of H3-JH, DH, TN1, and/or N2 segments of any of the theoretical segment pools provided herein, or generated by the methods described herein, where the number of segments is defined using any one of the percentages provided above for the respective segment. In some embodiments of the invention, a particular percentage range is defined, using any two of the percentages provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the percentages provided, which define an upper and lower boundary, are contemplated.

One of ordinary skill in the art will appreciate, upon reading the disclosure herein, that given the TN1, DH, N2, and/or H3-JH segments of any of the theoretical segment pools provided herein, or generated by the methods described herein, similar TN1, DH, N2, and/or H3-JH segments, and corresponding CDRH3 libraries, could be produced which, while not 100% identical to those provided in terms of their sequences, may be functionally very similar. Such theoretical segment pools and CDRH3 libraries also fall within the scope of the invention. A variety of techniques well-known in the art could be used to obtain these additional sequences, including the mutagenesis techniques provided herein. Therefore, each of the explicitly enumerated embodiments of the invention can also be practiced using segments that share a particular percent identity to any of the segments of any of the theoretical segment pools provided herein, or generated by the methods described herein. For example, each of the previously described embodiments of the invention can be practiced using TN1, DH, N2, and/or H3-JH segments that are at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identical to the TN1, DH, N2, and/or H3-JH segments of any of the theoretical segment pools provided herein, or generated by the methods described herein.

In some embodiments, the invention provides libraries produced from one or more VH chassis sequences combined with one or more TN1 segments, one or more DH

segments, one or more N2 segments, and one or more H3-JH segments. In certain embodiments at least 1, 2, 5, 10, 20, 50, 75, or 100, of each chassis, TN1, DH, N2, or H3-JH segments are included in a library of the invention.

In some embodiments, the invention provides a method of selecting TN1, DH, N2, and H3-JH segments from a theoretical segment pool for inclusion in a synthetic CDRH3 library, comprising:

- (i) providing a theoretical segment pool containing one or more of TN1, DH, N2, and H3-JH segments;
- (ii) providing a reference set of CDRH3 sequences;
- (iii) utilizing the theoretical segment pool of (i) to identify the closest match(es) to each CDRH3 sequence in the reference set of (ii); and
- (iv) selecting segments from the theoretical segment pool for inclusion in a synthetic library.

In some embodiments, the selection process of (iv) can optionally involve any number of additional criteria, including the frequency of occurrence of the segments of (i) in the reference set of (ii); the corresponding segmental usage weights; and any physicochemical properties (see all numerical indices at www.genome.jp/aaindex/) of the segments (*e.g.*, hydrophobicity, alpha-helical propensity, and/or isoelectric point). Optionally, TN1 and/or N2 segments that do not occur in the theoretical segment pool of (i) but that are found in the reference set of (ii) may be identified and added to prospective theoretical segment pools to produce theoretical segment pools with increased TN1 and/or N2 diversity in the prospective theoretical segment pools and/or synthetic libraries of the invention.

Any characteristic or set of characteristics of the segments can be used to choose them for inclusion in the library, including for example one or more biological properties (*e.g.*, immunogenicity, stability, half-life) and/or one or more physicochemical properties such as the numerical indices provided above. In some embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more such properties is used to select segments for inclusion in a library of the invention.

Physiochemical properties included in the index provided above can include, for example, ANDN920101 alpha-CH chemical shifts (Andersen et al., 1992); ARGP820101

Hydrophobicity index (Argos et al., 1982); ARGP820102 Signal sequence helical potential (Argos et al., 1982); ARGP820103 Membrane-buried preference parameters (Argos et al., 1982); BEGF750101 Conformational parameter of inner helix (Beghin-Dirkx, 1975); BEGF750102 Conformational parameter of beta-structure (Beghin-Dirkx, 1975); BEGF750103 Conformational parameter of beta-turn (Beghin-Dirkx, 1975); BHAR880101 Average flexibility indices (Bhaskaran-Ponnuswamy, 1988); BIGC670101 Residue volume (Bigelow, 1967); BIOV880101 Information value for accessibility; average fraction 35% (Biou et al., 1988); BIOV880102 Information value for accessibility; average fraction 23% (Biou et al., 1988); BROCC820101 Retention coefficient in TFA (Browne et al., 1982); BROCC820102 Retention coefficient in HFBA (Browne et al., 1982); BULH740101 Transfer free energy to surface (Bull-Breese, 1974); BULH740102 Apparent partial specific volume (Bull-Breese, 1974); BUNA790101 alpha-NH chemical shifts (Bundi-Wuthrich, 1979); BUNA790102 alpha-CH chemical shifts (Bundi-Wuthrich, 1979); BUNA790103 Spin-spin coupling constants $3J_{\text{H}\alpha\text{NH}}$ (Bundi-Wuthrich, 1979); BURA740101 Normalized frequency of alpha-helix (Burgess et al., 1974); BURA740102 Normalized frequency of extended structure (Burgess et al., 1974); CHAM810101 Steric parameter (Charton, 1981); CHAM820101 Polarizability parameter (Charton-Charton, 1982); CHAM820102 Free energy of solution in water, kcal/mole (Charton-Charton, 1982); CHAM830101 The Chou-Fasman parameter of the coil conformation (Charton-Charton, 1983); CHAM830102 A parameter defined from the residuals obtained from the best correlation of the Chou-Fasman parameter of beta-sheet (Charton-Charton, 1983); CHAM830103 The number of atoms in the side chain labelled 1+1 (Charton-Charton, 1983); CHAM830104 The number of atoms in the side chain labelled 2+1 (Charton-Charton, 1983); CHAM830105 The number of atoms in the side chain labelled 3+1 (Charton-Charton, 1983); CHAM830106 The number of bonds in the longest chain (Charton-Charton, 1983); CHAM830107 A parameter of charge transfer capability (Charton-Charton, 1983); CHAM830108 A parameter of charge transfer donor capability (Charton-Charton, 1983); CHOC750101 Average volume of buried residue (Chothia, 1975); CHOC760101 Residue accessible surface area in tripeptide (Chothia, 1976); CHOC760102 Residue accessible surface area in folded protein (Chothia, 1976); CHOC760103 Proportion of residues 95% buried

(Chothia, 1976); CHOC760104 Proportion of residues 100% buried (Chothia, 1976); CHOP780101 Normalized frequency of beta-turn (Chou-Fasman, 1978a); CHOP780201 Normalized frequency of alpha-helix (Chou-Fasman, 1978b); CHOP780202 Normalized frequency of beta-sheet (Chou-Fasman, 1978b); CHOP780203 Normalized frequency of beta-turn (Chou-Fasman, 1978b); CHOP780204 Normalized frequency of N-terminal helix (Chou-Fasman, 1978b); CHOP780205 Normalized frequency of C-terminal helix (Chou-Fasman, 1978b); CHOP780206 Normalized frequency of N-terminal non helical region (Chou-Fasman, 1978b); CHOP780207 Normalized frequency of C-terminal non helical region (Chou-Fasman, 1978b); CHOP780208 Normalized frequency of N-terminal beta-sheet (Chou-Fasman, 1978b); CHOP780209 Normalized frequency of C-terminal beta-sheet (Chou-Fasman, 1978b); CHOP780210 Normalized frequency of N-terminal non beta region (Chou-Fasman, 1978b); CHOP780211 Normalized frequency of C-terminal non beta region (Chou-Fasman, 1978b); CHOP780212 Frequency of the 1st residue in turn (Chou-Fasman, 1978b); CHOP780213 Frequency of the 2nd residue in turn (Chou-Fasman, 1978b); CHOP780214 Frequency of the 3rd residue in turn (Chou-Fasman, 1978b); CHOP780215 Frequency of the 4th residue in turn (Chou-Fasman, 1978b); CHOP780216 Normalized frequency of the 2nd and 3rd residues in turn (Chou-Fasman, 1978b); CIDH920101 Normalized hydrophobicity scales for alpha-proteins (Cid et al., 1992); CIDH920102 Normalized hydrophobicity scales for beta-proteins (Cid et al., 1992); CIDH920103 Normalized hydrophobicity scales for alpha+beta-proteins (Cid et al., 1992); CIDH920104 Normalized hydrophobicity scales for alpha/beta-proteins (Cid et al., 1992); CIDH920105 Normalized average hydrophobicity scales (Cid et al., 1992); COHE430101 Partial specific volume (Cohn-Edsall, 1943); CRAJ730101 Normalized frequency of middle helix (Crawford et al., 1973); CRAJ730102 Normalized frequency of beta-sheet (Crawford et al., 1973); CRAJ730103 Normalized frequency of turn (Crawford et al., 1973); DAWD720101 Size (Dawson, 1972); DAYM780101 Amino acid composition (Dayhoff et al., 1978a); DAYM780201 Relative mutability (Dayhoff et al., 1978b); DESM900101 Membrane preference for cytochrome b: MPH89 (Degli Esposti et al., 1990); DESM900102 Average membrane preference: AMP07 (Degli Esposti et al., 1990); EISD840101 Consensus normalized hydrophobicity scale (Eisenberg, 1984); EISD860101 Solvation free energy (Eisenberg-McLachlan, 1986);

EISD860102 Atom-based hydrophobic moment (Eisenberg-McLachlan, 1986);
 EISD860103 Direction of hydrophobic moment (Eisenberg-McLachlan, 1986);
 FASG760101 Molecular weight (Fasman, 1976); FASG760102 Melting point (Fasman, 1976); FASG760103 Optical rotation (Fasman, 1976); FASG760104 pK-N (Fasman, 1976); FASG760105 pK-C (Fasman, 1976); FAUJ830101 Hydrophobic parameter π (Fauchere-Pliska, 1983); FAUJ880101 Graph shape index (Fauchere et al., 1988); FAUJ880102 Smoothed upslon steric parameter (Fauchere et al., 1988); FAUJ880103 Normalized van der Waals volume (Fauchere et al., 1988); FAUJ880104 STERIMOL length of the side chain (Fauchere et al., 1988); FAUJ880105 STERIMOL minimum width of the side chain (Fauchere et al., 1988); FAUJ880106 STERIMOL maximum width of the side chain (Fauchere et al., 1988); FAUJ880107 N.m.r. chemical shift of alpha-carbon (Fauchere et al., 1988); FAUJ880108 Localized electrical effect (Fauchere et al., 1988); FAUJ880109 Number of hydrogen bond donors (Fauchere et al., 1988); FAUJ880110 Number of full nonbonding orbitals (Fauchere et al., 1988); FAUJ880111 Positive charge (Fauchere et al., 1988); FAUJ880112 Negative charge (Fauchere et al., 1988); FAUJ880113 pK-a(RCOOH) (Fauchere et al., 1988); FINA770101 Helix-coil equilibrium constant (Finkelstein-Ptitsyn, 1977); FINA910101 Helix initiation parameter at position $i-1$ (Finkelstein et al., 1991); FINA910102 Helix initiation parameter at position $i, i+1, i+2$ (Finkelstein et al., 1991); FINA910103 Helix termination parameter at position $j-2, j-1, j$ (Finkelstein et al., 1991); FINA910104 Helix termination parameter at position $j+1$ (Finkelstein et al., 1991); GARJ730101 Partition coefficient (Garel et al., 1973); GEIM800101 Alpha-helix indices (Geisow-Roberts, 1980); GEIM800102 Alpha-helix indices for alpha-proteins (Geisow-Roberts, 1980); GEIM800103 Alpha-helix indices for beta-proteins (Geisow-Roberts, 1980); GEIM800104 Alpha-helix indices for alpha/beta-proteins (Geisow-Roberts, 1980); GEIM800105 Beta-strand indices (Geisow-Roberts, 1980); GEIM800106 Beta-strand indices for beta-proteins (Geisow-Roberts, 1980); GEIM800107 Beta-strand indices for alpha/beta-proteins (Geisow-Roberts, 1980); GEIM800108 Aperiodic indices (Geisow-Roberts, 1980); GEIM800109 Aperiodic indices for alpha-proteins (Geisow-Roberts, 1980); GEIM800110 Aperiodic indices for beta-proteins (Geisow-Roberts, 1980); GEIM800111 Aperiodic indices for alpha/beta-proteins (Geisow-Roberts, 1980); GOLD730101 Hydrophobicity factor (Goldsack-

Chalifoux, 1973); GOLD730102 Residue volume (Goldsack-Chalifoux, 1973); GRAR740101 Composition (Grantham, 1974); GRAR740102 Polarity (Grantham, 1974); GRAR740103 Volume (Grantham, 1974); GUYH850101 Partition energy (Guy, 1985); HOPA770101 Hydration number (Hopfinger, 1971), Cited by Charton-Charton (1982); HOPT810101 Hydrophilicity value (Hopp-Woods, 1981); HUTJ700101 Heat capacity (Hutchens, 1970); HUTJ700102 Absolute entropy (Hutchens, 1970); HUTJ700103 Entropy of formation (Hutchens, 1970); ISOY800101 Normalized relative frequency of alpha-helix (Isogai et al., 1980); ISOY800102 Normalized relative frequency of extended structure (Isogai et al., 1980); ISOY800103 Normalized relative frequency of bend (Isogai et al., 1980); ISOY800104 Normalized relative frequency of bend R (Isogai et al., 1980); ISOY800105 Normalized relative frequency of bend S (Isogai et al., 1980); ISOY800106 Normalized relative frequency of helix end (Isogai et al., 1980); ISOY800107 Normalized relative frequency of double bend (Isogai et al., 1980); ISOY800108 Normalized relative frequency of coil (Isogai et al., 1980); JANJ780101 Average accessible surface area (Janin et al., 1978); JANJ780102 Percentage of buried residues (Janin et al., 1978); JANJ780103 Percentage of exposed residues (Janin et al., 1978); JANJ790101 Ratio of buried and accessible molar fractions (Janin, 1979); JANJ790102 Transfer free energy (Janin, 1979); JOND750101 Hydrophobicity (Jones, 1975); JOND750102 pK (-COOH) (Jones, 1975); JOND920101 Relative frequency of occurrence (Jones et al., 1992); JOND920102 Relative mutability (Jones et al., 1992); JUKT750101 Amino acid distribution (Jukes et al., 1975); JUNJ780101 Sequence frequency (Jungck, 1978); KANM800101 Average relative probability of helix (Kanehisa-Tsong, 1980); KANM800102 Average relative probability of beta-sheet (Kanehisa-Tsong, 1980); KANM800103 Average relative probability of inner helix (Kanehisa-Tsong, 1980); KANM800104 Average relative probability of inner beta-sheet (Kanehisa-Tsong, 1980); KARP850101 Flexibility parameter for no rigid neighbors (Karplus-Schulz, 1985); KARP850102 Flexibility parameter for one rigid neighbor (Karplus-Schulz, 1985); KARP850103 Flexibility parameter for two rigid neighbors (Karplus-Schulz, 1985); KHAG800101 The Kerr-constant increments (Khanarian-Moore, 1980); KLEP840101 Net charge (Klein et al., 1984); KRIW710101 Side chain interaction parameter (Krigbaum-Rubin, 1971); KRIW790101 Side chain interaction parameter

(Krigbaum-Komoriya, 1979); KRIW790102 Fraction of site occupied by water (Krigbaum-Komoriya, 1979); KRIW790103 Side chain volume (Krigbaum-Komoriya, 1979); KYTJ820101 Hydropathy index (Kyte-Doolittle, 1982); LAWE840101 Transfer free energy, CHP/water (Lawson et al., 1984); LEVM760101 Hydrophobic parameter (Levitt, 1976); LEVM760102 Distance between C-alpha and centroid of side chain (Levitt, 1976); LEVM760103 Side chain angle theta(AAR) (Levitt, 1976); LEVM760104 Side chain torsion angle phi(AAAR) (Levitt, 1976); LEVM760105 Radius of gyration of side chain (Levitt, 1976); LEVM760106 van der Waals parameter R0 (Levitt, 1976) LEVM760107 van der Waals parameter epsilon (Levitt, 1976); LEVM780101 Normalized frequency of alpha-helix, with weights (Levitt, 1978); LEVM780102 Normalized frequency of beta-sheet, with weights (Levitt, 1978); LEVM780103 Normalized frequency of reverse turn, with weights (Levitt, 1978); LEVM780104 Normalized frequency of alpha-helix, unweighted (Levitt, 1978); LEVM780105 Normalized frequency of beta-sheet, unweighted (Levitt, 1978); LEVM780106 Normalized frequency of reverse turn, unweighted (Levitt, 1978); LEWP710101 Frequency of occurrence in beta-bends (Lewis et al., 1971); LIFS790101 Conformational preference for all beta-strands (Lifson-Sander, 1979); LIFS790102 Conformational preference for parallel beta-strands (Lifson-Sander, 1979); LIFS790103 Conformational preference for antiparallel beta-strands (Lifson-Sander, 1979); MANP780101 Average surrounding hydrophobicity (Manavalan-Ponnuswamy, 1978); MAXF760101 Normalized frequency of alpha-helix (Maxfield-Scheraga, 1976); MAXF760102 Normalized frequency of extended structure (Maxfield-Scheraga, 1976); MAXF760103 Normalized frequency of zeta R (Maxfield-Scheraga, 1976); MAXF760104 Normalized frequency of left-handed alpha-helix (Maxfield-Scheraga, 1976); MAXF760105 Normalized frequency of zeta L (Maxfield-Scheraga, 1976); MAXF760106 Normalized frequency of alpha region (Maxfield-Scheraga, 1976); MCMT640101 Refractivity (McMeekin et al., 1964), Cited by Jones (1975); MEEJ800101 Retention coefficient in HPLC, pH7.4 (Meek, 1980); MEEJ800102 Retention coefficient in HPLC, pH2.1 (Meek, 1980); MEEJ810101 Retention coefficient in NaClO4 (Meek-Rossetti, 1981); MEEJ810102 Retention coefficient in NaH2PO4 (Meek-Rossetti, 1981); MEIH800101 Average reduced distance for C-alpha (Meirovitch et al., 1980); MEIH800102 Average

reduced distance for side chain (Meirovitch et al., 1980); MEIH800103 Average side chain orientation angle (Meirovitch et al., 1980); MIYS850101 Effective partition energy (Miyazawa-Jernigan, 1985); NAGK730101 Normalized frequency of alpha-helix (Nagano, 1973); NAGK730102 Normalized frequency of beta-structure (Nagano, 1973); NAGK730103 Normalized frequency of coil (Nagano, 1973); NAKH900101 AA composition of total proteins (Nakashima et al., 1990); NAKH900102 SD of AA composition of total proteins (Nakashima et al., 1990); NAKH900103 AA composition of mt-proteins (Nakashima et al., 1990); NAKH900104 Normalized composition of mt-proteins (Nakashima et al., 1990); NAKH900105 AA composition of mt-proteins from animal (Nakashima et al., 1990); NAKH900106 Normalized composition from animal (Nakashima et al., 1990); NAKH900107 AA composition of mt-proteins from fungi and plant (Nakashima et al., 1990); NAKH900108 Normalized composition from fungi and plant (Nakashima et al., 1990); NAKH900109 AA composition of membrane proteins (Nakashima et al., 1990); NAKH900110 Normalized composition of membrane proteins (Nakashima et al., 1990); NAKH900111 Transmembrane regions of non-mt-proteins (Nakashima et al., 1990); NAKH900112 Transmembrane regions of mt-proteins (Nakashima et al., 1990); NAKH900113 Ratio of average and computed composition (Nakashima et al., 1990); NAKH920101 AA composition of CYT of single-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920102 AA composition of CYT2 of single-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920103 AA composition of EXT of single-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920104 AA composition of EXT2 of single-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920105 AA composition of MEM of single-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920106 AA composition of CYT of multi-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920107 AA composition of EXT of multi-spanning proteins (Nakashima-Nishikawa, 1992); NAKH920108 AA composition of MEM of multi-spanning proteins (Nakashima-Nishikawa, 1992); NISK800101 8 A contact number (Nishikawa-Ooi, 1980); NISK860101 14 A contact number (Nishikawa-Ooi, 1986); NOZY710101 Transfer energy, organic solvent/water (Nozaki-Tanford, 1971); OOBM770101 Average non-bonded energy per atom (Oobatake-Ooi, 1977);

OOBM770102 Short and medium range non-bonded energy per atom (Oobatake-Ooi, 1977); OOBM770103 Long range non-bonded energy per atom (Oobatake-Ooi, 1977); OOBM770104 Average non-bonded energy per residue (Oobatake-Ooi, 1977); OOBM770105 Short and medium range non-bonded energy per residue (Oobatake-Ooi, 1977); OOBM850101 Optimized beta-structure-coil equilibrium constant (Oobatake et al., 1985); OOBM850102 Optimized propensity to form reverse turn (Oobatake et al., 1985); OOBM850103 Optimized transfer energy parameter (Oobatake et al., 1985); OOBM850104 Optimized average non-bonded energy per atom (Oobatake et al., 1985); OOBM850105 Optimized side chain interaction parameter (Oobatake et al., 1985); PALJ810101 Normalized frequency of alpha-helix from LG (Palau et al., 1981); PALJ810102 Normalized frequency of alpha-helix from CF (Palau et al., 1981); PALJ810103 Normalized frequency of beta-sheet from LG (Palau et al., 1981); PALJ810104 Normalized frequency of beta-sheet from CF (Palau et al., 1981); PALJ810105 Normalized frequency of turn from LG (Palau et al., 1981); PALJ810106 Normalized frequency of turn from CF (Palau et al., 1981); PALJ810107 Normalized frequency of alpha-helix in all-alpha class (Palau et al., 1981); PALJ810108 Normalized frequency of alpha-helix in alpha+beta class (Palau et al., 1981); PALJ810109 Normalized frequency of alpha-helix in alpha/beta class (Palau et al., 1981); PALJ810110 Normalized frequency of beta-sheet in all-beta class (Palau et al., 1981); PALJ810111 Normalized frequency of beta-sheet in alpha+beta class (Palau et al., 1981); PALJ810112 Normalized frequency of beta-sheet in alpha/beta class (Palau et al., 1981); PALJ810113 Normalized frequency of turn in all-alpha class (Palau et al., 1981); PALJ810114 Normalized frequency of turn in all-beta class (Palau et al., 1981); PALJ810115 Normalized frequency of turn in alpha+beta class (Palau et al., 1981); PALJ810116 Normalized frequency of turn in alpha/beta class (Palau et al., 1981); PARJ860101 HPLC parameter (Parker et al., 1986); PLIV810101 Partition coefficient (Pliska et al., 1981); PONP800101 Surrounding hydrophobicity in folded form (Ponnuswamy et al., 1980); PONP800102 Average gain in surrounding hydrophobicity (Ponnuswamy et al., 1980); PONP800103 Average gain ratio in surrounding hydrophobicity (Ponnuswamy et al., 1980); PONP800104 Surrounding hydrophobicity in alpha-helix (Ponnuswamy et al., 1980); PONP800105 Surrounding hydrophobicity in

beta-sheet (Ponnuswamy et al., 1980); PONP800106 Surrounding hydrophobicity in turn (Ponnuswamy et al., 1980); PONP800107 Accessibility reduction ratio (Ponnuswamy et al., 1980); PONP800108 Average number of surrounding residues (Ponnuswamy et al., 1980); PRAM820101 Intercept in regression analysis (Prabhakaran-Ponnuswamy, 1982); PRAM820102 Slope in regression analysis x 1.0E1 (Prabhakaran-Ponnuswamy, 1982); PRAM820103 Correlation coefficient in regression analysis (Prabhakaran-Ponnuswamy, 1982); PRAM900101 Hydrophobicity (Prabhakaran, 1990); PRAM900102 Relative frequency in alpha-helix (Prabhakaran, 1990); PRAM900103 Relative frequency in beta-sheet (Prabhakaran, 1990); PRAM900104 Relative frequency in reverse-turn (Prabhakaran, 1990); PTIO830101 Helix-coil equilibrium constant (Ptitsyn-Finkelstein, 1983); PTIO830102 Beta-coil equilibrium constant (Ptitsyn-Finkelstein, 1983); QIAN880101 Weights for alpha-helix at the window position of -6 (Qian-Sejnowski, 1988); QIAN880102 Weights for alpha-helix at the window position of -5 (Qian-Sejnowski, 1988); QIAN880103 Weights for alpha-helix at the window position of -4 (Qian-Sejnowski, 1988); QIAN880104 Weights for alpha-helix at the window position of -3 (Qian-Sejnowski, 1988); QIAN880105 Weights for alpha-helix at the window position of -2 (Qian-Sejnowski, 1988); QIAN880106 Weights for alpha-helix at the window position of -1 (Qian-Sejnowski, 1988); QIAN880107 Weights for alpha-helix at the window position of 0 (Qian-Sejnowski, 1988); QIAN880108 Weights for alpha-helix at the window position of 1 (Qian-Sejnowski, 1988); QIAN880109 Weights for alpha-helix at the window position of 2 (Qian-Sejnowski, 1988); QIAN880110 Weights for alpha-helix at the window position of 3 (Qian-Sejnowski, 1988); QIAN880111 Weights for alpha-helix at the window position of 4 (Qian-Sejnowski, 1988); QIAN880112 Weights for alpha-helix at the window position of 5 (Qian-Sejnowski, 1988); QIAN880113 Weights for alpha-helix at the window position of 6 (Qian-Sejnowski, 1988); QIAN880114 Weights for beta-sheet at the window position of -6 (Qian-Sejnowski, 1988); QIAN880115 Weights for beta-sheet at the window position of -5 (Qian-Sejnowski, 1988); QIAN880116 Weights for beta-sheet at the window position of -4 (Qian-Sejnowski, 1988); QIAN880117 Weights for beta-sheet at the window position of -3 (Qian-Sejnowski, 1988); QIAN880118 Weights for beta-sheet at the window position of -2 (Qian-Sejnowski, 1988); QIAN880119 Weights for beta-sheet at the window

position of -1 (Qian-Sejnowski, 1988); QIAN880120 Weights for beta-sheet at the window position of 0 (Qian-Sejnowski, 1988); QIAN880121 Weights for beta-sheet at the window position of 1 (Qian-Sejnowski, 1988); QIAN880122 Weights for beta-sheet at the window position of 2 (Qian-Sejnowski, 1988); QIAN880123 Weights for beta-sheet at the window position of 3 (Qian-Sejnowski, 1988); QIAN880124 Weights for beta-sheet at the window position of 4 (Qian-Sejnowski, 1988); QIAN880125 Weights for beta-sheet at the window position of 5 (Qian-Sejnowski, 1988); QIAN880126 Weights for beta-sheet at the window position of 6 (Qian-Sejnowski, 1988); QIAN880127 Weights for coil at the window position of -6 (Qian-Sejnowski, 1988); QIAN880128 Weights for coil at the window position of -5 (Qian-Sejnowski, 1988); QIAN880129 Weights for coil at the window position of -4 (Qian-Sejnowski, 1988); QIAN880130 Weights for coil at the window position of -3 (Qian-Sejnowski, 1988); QIAN880131 Weights for coil at the window position of -2 (Qian-Sejnowski, 1988); QIAN880132 Weights for coil at the window position of -1 (Qian-Sejnowski, 1988); QIAN880133 Weights for coil at the window position of 0 (Qian-Sejnowski, 1988); QIAN880134 Weights for coil at the window position of 1 (Qian-Sejnowski, 1988); QIAN880135 Weights for coil at the window position of 2 (Qian-Sejnowski, 1988); QIAN880136 Weights for coil at the window position of 3 (Qian-Sejnowski, 1988); QIAN880137 Weights for coil at the window position of 4 (Qian-Sejnowski, 1988); QIAN880138 Weights for coil at the window position of 5 (Qian-Sejnowski, 1988); QIAN880139 Weights for coil at the window position of 6 (Qian-Sejnowski, 1988); RACS770101 Average reduced distance for C-alpha (Rackovsky-Scheraga, 1977); RACS770102 Average reduced distance for side chain (Rackovsky-Scheraga, 1977); RACS770103 Side chain orientational preference (Rackovsky-Scheraga, 1977); RACS820101 Average relative fractional occurrence in A0(i) (Rackovsky-Scheraga, 1982); RACS820102 Average relative fractional occurrence in AR(i) (Rackovsky-Scheraga, 1982); RACS820103 Average relative fractional occurrence in AL(i) (Rackovsky-Scheraga, 1982); RACS820104 Average relative fractional occurrence in EL(i) (Rackovsky-Scheraga, 1982); RACS820105 Average relative fractional occurrence in E0(i) (Rackovsky-Scheraga, 1982); RACS820106 Average relative fractional occurrence in ER(i) (Rackovsky-Scheraga, 1982); RACS820107 Average relative

fractional occurrence in A0(i-1) (Rackovsky-Scheraga, 1982); RACS820108 Average relative fractional occurrence in AR(i-1) (Rackovsky-Scheraga, 1982); RACS820109 Average relative fractional occurrence in AL(i-1) (Rackovsky-Scheraga, 1982); RACS820110 Average relative fractional occurrence in EL(i-1) (Rackovsky-Scheraga, 1982); RACS820111 Average relative fractional occurrence in E0(i-1) (Rackovsky-Scheraga, 1982); RACS820112 Average relative fractional occurrence in ER(i-1) (Rackovsky-Scheraga, 1982); RACS820113 Value of theta(i) (Rackovsky-Scheraga, 1982); RACS820114 Value of theta(i-1) (Rackovsky-Scheraga, 1982); RADA880101 Transfer free energy from chx to wat (Radzicka-Wolfenden, 1988); RADA880102 Transfer free energy from oct to wat (Radzicka-Wolfenden, 1988); RADA880103 Transfer free energy from vap to chx (Radzicka-Wolfenden, 1988); RADA880104 Transfer free energy from chx to oct (Radzicka-Wolfenden, 1988); RADA880105 Transfer free energy from vap to oct (Radzicka-Wolfenden, 1988); RADA880106 Accessible surface area (Radzicka-Wolfenden, 1988); RADA880107 Energy transfer from out to in(95%buried) (Radzicka-Wolfenden, 1988); RADA880108 Mean polarity (Radzicka-Wolfenden, 1988); RICJ880101 Relative preference value at N" (Richardson-Richardson, 1988); RICJ880102 Relative preference value at N' (Richardson-Richardson, 1988); RICJ880103 Relative preference value at N-cap (Richardson-Richardson, 1988); RICJ880104 Relative preference value at N1 (Richardson-Richardson, 1988); RICJ880105 Relative preference value at N2 (Richardson-Richardson, 1988); RICJ880106 Relative preference value at N3 (Richardson-Richardson, 1988); RICJ880107 Relative preference value at N4 (Richardson-Richardson, 1988); RICJ880108 Relative preference value at N5 (Richardson-Richardson, 1988); RICJ880109 Relative preference value at Mid (Richardson-Richardson, 1988); RICJ880110 Relative preference value at C5 (Richardson-Richardson, 1988); RICJ880111 Relative preference value at C4 (Richardson-Richardson, 1988); RICJ880112 Relative preference value at C3 (Richardson-Richardson, 1988); RICJ880113 Relative preference value at C2 (Richardson-Richardson, 1988); RICJ880114 Relative preference value at C1 (Richardson-Richardson, 1988); RICJ880115 Relative preference value at C-cap (Richardson-Richardson, 1988); RICJ880116 Relative preference value at C' (Richardson-Richardson, 1988);

RICJ880117 Relative preference value at C" (Richardson-Richardson, 1988);
ROBB760101 Information measure for alpha-helix (Robson-Suzuki, 1976);
ROBB760102 Information measure for N-terminal helix (Robson-Suzuki, 1976);
ROBB760103 Information measure for middle helix (Robson-Suzuki, 1976);
ROBB760104 Information measure for C-terminal helix (Robson-Suzuki, 1976);
ROBB760105 Information measure for extended (Robson-Suzuki, 1976); ROBB760106
Information measure for pleated-sheet (Robson-Suzuki, 1976); ROBB760107
Information measure for extended without H-bond (Robson-Suzuki, 1976);
ROBB760108 Information measure for turn (Robson-Suzuki, 1976); ROBB760109
Information measure for N-terminal turn (Robson-Suzuki, 1976); ROBB760110
Information measure for middle turn (Robson-Suzuki, 1976); ROBB760111 Information
measure for C-terminal turn (Robson-Suzuki, 1976); ROBB760112 Information measure
for coil (Robson-Suzuki, 1976); ROBB760113 Information measure for loop (Robson-
Suzuki, 1976); ROBB790101 Hydration free energy (Robson-Osguthorpe, 1979);
ROSG850101 Mean area buried on transfer (Rose et al., 1985); ROSG850102 Mean
fractional area loss (Rose et al., 1985); ROSM880101 Side chain hydrophathy, uncorrected
for solvation (Roseman, 1988); ROSM880102 Side chain hydrophathy, corrected for
solvation (Roseman, 1988); ROSM880103 Loss of Side chain hydrophathy by helix
formation (Roseman, 1988); SIMZ760101 Transfer free energy (Simon, 1976), Cited by
Charton-Charton (1982); SNEP660101 Principal component I (Sneath, 1966);
SNEP660102 Principal component II (Sneath, 1966); SNEP660103 Principal component
III (Sneath, 1966); SNEP660104 Principal component IV (Sneath, 1966); SUEM840101
Zimm-Bragg parameter s at 20 C (Sueki et al., 1984); SUEM840102 Zimm-Bragg
parameter $\sigma \times 1.0E4$ (Sueki et al., 1984); SWER830101 Optimal matching
hydrophobicity (Sweet-Eisenberg, 1983); TANS770101 Normalized frequency of alpha-
helix (Tanaka-Scheraga, 1977); TANS770102 Normalized frequency of isolated helix
(Tanaka-Scheraga, 1977); TANS770103 Normalized frequency of extended structure
(Tanaka-Scheraga, 1977); TANS770104 Normalized frequency of chain reversal R
(Tanaka-Scheraga, 1977); TANS770105 Normalized frequency of chain reversal S
(Tanaka-Scheraga, 1977); TANS770106 Normalized frequency of chain reversal D
(Tanaka-Scheraga, 1977); TANS770107 Normalized frequency of left-handed helix

(Tanaka-Scheraga, 1977); TANS770108 Normalized frequency of zeta R (Tanaka-Scheraga, 1977); TANS770109 Normalized frequency of coil (Tanaka-Scheraga, 1977); TANS770110 Normalized frequency of chain reversal (Tanaka-Scheraga, 1977); VASM830101 Relative population of conformational state A (Vasquez et al., 1983); VASM830102 Relative population of conformational state C (Vasquez et al., 1983); VASM830103 Relative population of conformational state E (Vasquez et al., 1983); VELV850101 Electron-ion interaction potential (Veljkovic et al., 1985); VENT840101 Bitterness (Venanzi, 1984); VHEG790101 Transfer free energy to lipophilic phase (von Heijne-Blomberg, 1979); WARP780101 Average interactions per side chain atom (Warne-Morgan, 1978); WEBA780101 RF value in high salt chromatography (Weber-Lacey, 1978); WERD780101 Propensity to be buried inside (Wertz-Scheraga, 1978); WERD780102 Free energy change of epsilon(i) to epsilon(ex) (Wertz-Scheraga, 1978); WERD780103 Free energy change of alpha(Ri) to alpha(Rh) (Wertz-Scheraga, 1978); WERD780104 Free energy change of epsilon(i) to alpha(Rh) (Wertz-Scheraga, 1978); WOEC730101 Polar requirement (Woese, 1973); WOLR810101 Hydration potential (Wolfenden et al., 1981); WOLS870101 Principal property value z1 (Wold et al., 1987); WOLS870102 Principal property value z2 (Wold et al., 1987); WOLS870103 Principal property value z3 (Wold et al., 1987); YUTK870101 Unfolding Gibbs energy in water, pH7.0 (Yutani et al., 1987); YUTK870102 Unfolding Gibbs energy in water, pH9.0 (Yutani et al., 1987); YUTK870103 Activation Gibbs energy of unfolding, pH7.0 (Yutani et al., 1987); YUTK870104 Activation Gibbs energy of unfolding, pH9.0 (Yutani et al., 1987); ZASB820101 Dependence of partition coefficient on ionic strength (Zaslavsky et al., 1982); ZIMJ680101 Hydrophobicity (Zimmerman et al., 1968); ZIMJ680102 Bulkiness (Zimmerman et al., 1968); ZIMJ680103 Polarity (Zimmerman et al., 1968); ZIMJ680104 Isoelectric point (Zimmerman et al., 1968); ZIMJ680105 RF rank (Zimmerman et al., 1968); AURR980101 Normalized positional residue frequency at helix termini N4'(Aurora-Rose, 1998); AURR980102 Normalized positional residue frequency at helix termini N''' (Aurora-Rose, 1998); AURR980103 Normalized positional residue frequency at helix termini N'' (Aurora-Rose, 1998); AURR980104 Normalized positional residue frequency at helix termini N'(Aurora-Rose, 1998); AURR980105 Normalized positional residue frequency at helix termini Nc (Aurora-

Rose, 1998); AURR980106 Normalized positional residue frequency at helix termini N1 (Aurora-Rose, 1998); AURR980107 Normalized positional residue frequency at helix termini N2 (Aurora-Rose, 1998); AURR980108 Normalized positional residue frequency at helix termini N3 (Aurora-Rose, 1998); AURR980109 Normalized positional residue frequency at helix termini N4 (Aurora-Rose, 1998); AURR980110 Normalized positional residue frequency at helix termini N5 (Aurora-Rose, 1998); AURR980111 Normalized positional residue frequency at helix termini C5 (Aurora-Rose, 1998); AURR980112 Normalized positional residue frequency at helix termini C4 (Aurora-Rose, 1998); AURR980113 Normalized positional residue frequency at helix termini C3 (Aurora-Rose, 1998); AURR980114 Normalized positional residue frequency at helix termini C2 (Aurora-Rose, 1998); AURR980115 Normalized positional residue frequency at helix termini C1 (Aurora-Rose, 1998); AURR980116 Normalized positional residue frequency at helix termini Cc (Aurora-Rose, 1998); AURR980117 Normalized positional residue frequency at helix termini C' (Aurora-Rose, 1998); AURR980118 Normalized positional residue frequency at helix termini C'' (Aurora-Rose, 1998); AURR980119 Normalized positional residue frequency at helix termini C''' (Aurora-Rose, 1998); AURR980120 Normalized positional residue frequency at helix termini C4' (Aurora-Rose, 1998); ONEK900101 Delta G values for the peptides extrapolated to 0 M urea (O'Neil-DeGrado, 1990); ONEK900102 Helix formation parameters (delta delta G) (O'Neil-DeGrado, 1990); VINM940101 Normalized flexibility parameters (B-values), average (Vihinen et al., 1994); VINM940102 Normalized flexibility parameters (B-values) for each residue surrounded by none rigid neighbours (Vihinen et al., 1994); VINM940103 Normalized flexibility parameters (B-values) for each residue surrounded by one rigid neighbours (Vihinen et al., 1994); VINM940104 Normalized flexibility parameters (B-values) for each residue surrounded by two rigid neighbours (Vihinen et al., 1994); MUNV940101 Free energy in alpha-helical conformation (Munoz-Serrano, 1994); MUNV940102 Free energy in alpha-helical region (Munoz-Serrano, 1994); MUNV940103 Free energy in beta-strand conformation (Munoz-Serrano, 1994); MUNV940104 Free energy in beta-strand region (Munoz-Serrano, 1994); MUNV940105 Free energy in beta-strand region (Munoz-Serrano, 1994)

WIMW960101 Free energies of transfer of AcWl-X-LL peptides from bilayer interface to water (Wimley-White, 1996); KIMC930101 Thermodynamic beta sheet propensity (Kim-Berg, 1993); MONM990101 Turn propensity scale for transmembrane helices (Monne et al., 1999); BLAM930101 Alpha helix propensity of position 44 in T4 lysozyme (Blaber et al., 1993); PARS000101 p-Values of mesophilic proteins based on the distributions of B values (Parthasarathy-Murthy, 2000); PARS000102 p-Values of thermophilic proteins based on the distributions of B values (Parthasarathy-Murthy, 2000); KUMS000101 Distribution of amino acid residues in the 18 non-redundant families of thermophilic proteins (Kumar et al., 2000); KUMS000102 Distribution of amino acid residues in the 18 non-redundant families of mesophilic proteins (Kumar et al., 2000); KUMS000103 Distribution of amino acid residues in the alpha-helices in thermophilic proteins (Kumar et al., 2000); KUMS000104 Distribution of amino acid residues in the alpha-helices in mesophilic proteins (Kumar et al., 2000); TAKK010101 Side-chain contribution to protein stability (kJ/mol) (Takano-Yutani, 2001); FODM020101 Propensity of amino acids within pi-helices (Fodje-Al-Karadaghi, 2002); NADH010101 Hydropathy scale based on self-information values in the two-state model (5% accessibility) (Naderi-Manesh et al., 2001); NADH010102 Hydropathy scale based on self-information values in the two-state model (9% accessibility) (Naderi-Manesh et al., 2001); NADH010103 Hydropathy scale based on self-information values in the two-state model (16% accessibility) (Naderi-Manesh et al., 2001); NADH010104 Hydropathy scale based on self-information values in the two-state model (20% accessibility) (Naderi-Manesh et al., 2001); NADH010105 Hydropathy scale based on self-information values in the two-state model (25% accessibility) (Naderi-Manesh et al., 2001); NADH010106 Hydropathy scale based on self-information values in the two-state model (36% accessibility) (Naderi-Manesh et al., 2001); NADH010107 Hydropathy scale based on self-information values in the two-state model (50% accessibility) (Naderi-Manesh et al., 2001); MONM990201 Averaged turn propensities in a transmembrane helix (Monne et al., 1999); KOEP990101 Alpha-helix propensity derived from designed sequences (Koehl-Levitt, 1999); KOEP990102 Beta-sheet propensity derived from designed sequences (Koehl-Levitt, 1999); CEDJ970101 Composition of amino acids in extracellular proteins (percent) (Cedano et al., 1997); CEDJ970102

Composition of amino acids in anchored proteins (percent) (Cedano et al., 1997); CEDJ970103 Composition of amino acids in membrane proteins (percent) (Cedano et al., 1997); CEDJ970104 Composition of amino acids in intracellular proteins (percent) (Cedano et al., 1997); CEDJ970105 Composition of amino acids in nuclear proteins (percent) (Cedano et al., 1997); FUKS010101 Surface composition of amino acids in intracellular proteins of thermophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010102 Surface composition of amino acids in intracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010103 Surface composition of amino acids in extracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010104 Surface composition of amino acids in nuclear proteins (percent) (Fukuchi-Nishikawa, 2001); FUKS010105 Interior composition of amino acids in intracellular proteins of thermophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010106 Interior composition of amino acids in intracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010107 Interior composition of amino acids in extracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010108 Interior composition of amino acids in nuclear proteins (percent) (Fukuchi-Nishikawa, 2001); FUKS010109 Entire chain composition of amino acids in intracellular proteins of thermophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010110 Entire chain composition of amino acids in intracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010111 Entire chain composition of amino acids in extracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001); FUKS010112 Entire chain composition of amino acids in nuclear proteins (percent) (Fukuchi-Nishikawa, 2001); AVBF000101 Screening coefficients gamma, local (Avbelj, 2000); AVBF000102 Screening coefficients gamma, non-local (Avbelj, 2000); AVBF000103 Slopes tripeptide, FDPB VFF neutral (Avbelj, 2000); AVBF000104 Slopes tripeptides, LD VFF neutral (Avbelj, 2000); AVBF000105 Slopes tripeptide, FDPB VFF no side (Avbelj, 2000); AVBF000106 Slopes tripeptide FDPB VFF all (Avbelj, 2000); AVBF000107 Slopes tripeptide FDPB PARSE neutral (Avbelj, 2000); AVBF000108 Slopes decapeptide, FDPB VFF neutral (Avbelj, 2000); AVBF000109 Slopes proteins, FDPB VFF neutral (Avbelj, 2000); YANJ020101 Side-chain conformation by gaussian evolutionary method (Yang et al., 2002); MITS020101

Amphiphilicity index (Mitaku et al., 2002); TSAJ990101 Volumes including the crystallographic waters using the ProtOr (Tsai et al., 1999); TSAJ990102 Volumes not including the crystallographic waters using the ProtOr (Tsai et al., 1999); COSI940101 Electron-ion interaction potential values (Cosic, 1994); PONP930101 Hydrophobicity scales (Ponnuswamy, 1993); WILM950101 Hydrophobicity coefficient in RP-HPLC, C18 with 0.1%TFA/MeCN/H2O (Wilce et al. 1995); WILM950102 Hydrophobicity coefficient in RP-HPLC, C8 with 0.1%TFA/MeCN/H2O (Wilce et al. 1995); WILM950103 Hydrophobicity coefficient in RP-HPLC, C4 with 0.1%TFA/MeCN/H2O (Wilce et al. 1995); WILM950104 Hydrophobicity coefficient in RP-HPLC, C18 with 0.1%TFA/2-PrOH/MeCN/H2O (Wilce et al. 1995); KUHL950101 Hydrophilicity scale (Kuhn et al., 1995); GUOD860101 Retention coefficient at pH 2 (Guo et al., 1986); JURD980101 Modified Kyte-Doolittle hydrophobicity scale (Juretic et al., 1998); BASU050101 Interactivity scale obtained from the contact matrix (Bastolla et al., 2005); BASU050102 Interactivity scale obtained by maximizing the mean of correlation coefficient over single-domain globular proteins (Bastolla et al., 2005); BASU050103 Interactivity scale obtained by maximizing the mean of correlation coefficient over pairs of sequences sharing the TIM barrel fold (Bastolla et al., 2005); SUYM030101 Linker propensity index (Suyama-Ohara, 2003); PUNT030101 Knowledge-based membrane-propensity scale from 1D_Helix in MPtopo databases (Punta-Maritan, 2003); PUNT030102 Knowledge-based membrane-propensity scale from 3D_Helix in MPtopo databases (Punta-Maritan, 2003); GEOR030101 Linker propensity from all dataset (George-Heringa, 2003); GEOR030102 Linker propensity from 1-linker dataset (George-Heringa, 2003); GEOR030103 Linker propensity from 2-linker dataset (George-Heringa, 2003); GEOR030104 Linker propensity from 3-linker dataset (George-Heringa, 2003); GEOR030105 Linker propensity from small dataset (linker length is less than six residues) (George-Heringa, 2003); GEOR030106 Linker propensity from medium dataset (linker length is between six and 14 residues) (George-Heringa, 2003); GEOR030107 Linker propensity from long dataset (linker length is greater than 14 residues) (George-Heringa, 2003); GEOR030108 Linker propensity from helical (annotated by DSSP) dataset (George-Heringa, 2003); GEOR030109 Linker propensity from non-helical (annotated by DSSP) dataset (George-Heringa, 2003); ZHOH040101 The stability scale

from the knowledge-based atom-atom potential (Zhou-Zhou, 2004); ZHOH040102 The relative stability scale extracted from mutation experiments (Zhou-Zhou, 2004); ZHOH040103 Buriability (Zhou-Zhou, 2004); BAEK050101 Linker index (Bae et al., 2005); HARY940101 Mean volumes of residues buried in protein interiors (Harpaz et al., 1994); PONJ960101 Average volumes of residues (Pontius et al., 1996); DIGM050101 Hydrostatic pressure asymmetry index, PAI (Di Giulio, 2005); WOLR790101 Hydrophobicity index (Wolfenden et al., 1979); OLSK800101 Average internal preferences (Olsen, 1980); KIDA850101 Hydrophobicity-related index (Kidera et al., 1985); GUYH850102 Apparent partition energies calculated from Wertz-Scheraga index (Guy, 1985); GUYH850103 Apparent partition energies calculated from Robson-Osguthorpe index (Guy, 1985); GUYH850104 Apparent partition energies calculated from Janin index (Guy, 1985); GUYH850105 Apparent partition energies calculated from Chothia index (Guy, 1985); ROSM880104 Hydropathies of amino acid side chains, neutral form (Roseman, 1988); ROSM880105 Hydropathies of amino acid side chains, pi-values in pH 7.0 (Roseman, 1988); JACR890101 Weights from the IFH scale (Jacobs-White, 1989); COWR900101 Hydrophobicity index, 3.0 pH (Cowan-Whittaker, 1990); BLAS910101 Scaled side chain hydrophobicity values (Black-Mould, 1991); CASG920101 Hydrophobicity scale from native protein structures (Casari-Sippl, 1992); CORJ870101 NNEIG index (Cornette et al., 1987); CORJ870102 SWEIG index (Cornette et al., 1987); CORJ870103 PRIFT index (Cornette et al., 1987); CORJ870104 PRILS index (Cornette et al., 1987); CORJ870105 ALTFT index (Cornette et al., 1987); CORJ870106 ALTLS index (Cornette et al., 1987); CORJ870107 TOTFT index (Cornette et al., 1987); CORJ870108 TOTLS index (Cornette et al., 1987); MIYS990101 Relative partition energies derived by the Bethe approximation (Miyazawa-Jernigan, 1999); MIYS990102 Optimized relative partition energies - method A (Miyazawa-Jernigan, 1999); MIYS990103 Optimized relative partition energies - method B (Miyazawa-Jernigan, 1999); MIYS990104 Optimized relative partition energies - method C (Miyazawa-Jernigan, 1999); MIYS990105 Optimized relative partition energies - method D (Miyazawa-Jernigan, 1999); ENGD860101 Hydrophobicity index (Engelman et al., 1986); and FASG890101 Hydrophobicity index (Fasman, 1989)

In some embodiments of the invention, degenerate oligonucleotides are used to synthesize one or more of the TN1, DH, N2, and/or H3-JH segments of the invention. In certain embodiments of the invention, the codon at or near the 5' end of the oligonucleotide encoding the H3-JH segment is a degenerate codon. Such degenerate codons may be the first codon from the 5' end, the second codon from the 5' end, the third codon from the 5' end, the fourth codon from the 5' end, the fifth codon from the 5' end, and/or any combination of the above. In some embodiments of the invention, one or more of the codons at or near the 5' and/or 3' ends of the DH segment are degenerate. Such degenerate codons may be the first codon from the 5' and/or 3' end(s), the second codon from the 5' and/or 3' end(s), the third codons from the 5' and/or 3' end(s), the fourth codon from the 5' and/or 3' end(s), the fifth codon from the 5' and/or 3' end(s), and/or any combination of the above. Degenerate codons used in each of the oligonucleotides encoding the segments may be selected for their ability to optimally recapitulate sequences in a theoretical segment pool and/or CDRH3 reference set.

In some embodiments, the invention provides methods of producing a theoretical segment pool of H3-JH segments, as described in the Examples. Theoretical segment pools generated utilizing NNN triplets, instead of or in addition to the NN doublets described in Example 5 also fall within the scope of the invention, as do synthetic libraries incorporating segments from these theoretical segment pools.

In some embodiments, the invention provides methods of producing a theoretical segment pool of DH segments, as described in the Examples. In particular, for example, the invention provides methods of producing a theoretical segment pool of DH segments described by the PYTHON program of Example 6. Example 6 describes the application of this program to produce the 68K theoretical segment pool (minimum length of DNA sequences after progressive deletions = 4 bases; and minimum length of peptide sequences for inclusion in the theoretical segment pool = 2). An alternative example is provided wherein the minimum length of the DNA sequences after progressive deletions was one base and the minimum length of the peptide sequence is one amino acid. It is also contemplated that other values could be used for these parameters. For example, the minimum length of the DNA sequences after progressive deletions could be set as 1, 2, 3,

4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15, and the minimum length of the peptide sequences in the theoretical segment pool could be set as 1, 2, 3, 4, or 5.

Design of CDRH3 Libraries Using the TN1, DH, N2, and H3-JH Segments

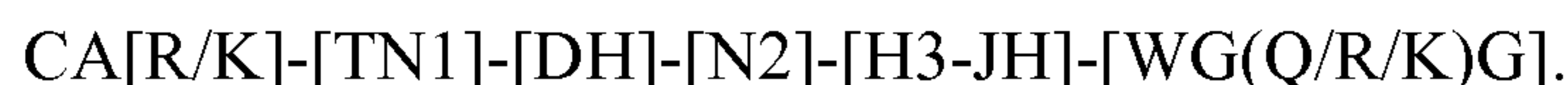
The CDRH3 libraries of the invention comprise TN1, DH, N2, and H3-JH segments. Thus, in certain embodiments of the invention, the overall design of the CDRH3 libraries can be represented by the following formula:



In certain embodiments of the invention, a synthetic CDRH3 repertoire is combined with selected VH chassis sequences and heavy chain constant regions, via homologous recombination. Therefore, in certain embodiments of the invention, it may be desirable to include DNA sequences flanking the 5' and 3' ends of the synthetic CDRH3 libraries, to facilitate homologous recombination between the synthetic CDRH3 libraries and vectors containing the selected chassis and constant regions. In certain embodiments, the vectors also contain a sequence encoding at least a portion of the non-truncated region of the IGHJ gene (*i.e.*, FRM4-JH). Thus, a polynucleotide encoding an N-terminal sequence (*e.g.*, CA(K/R/T)) may be added to the synthetic CDRH3 sequences, wherein the N-terminal polynucleotide is homologous with FRM3 of the chassis, while a polynucleotide encoding a C-terminal sequence (*e.g.*, WG(Q/R/K)G) may be added to the synthetic CDRH3, wherein the C-terminal polynucleotide is homologous with FRM4-JH. Although the sequence WG(Q/R)G is presented in this exemplary embodiment, additional amino acids, C-terminal to this sequence in FRM4-JH may also be included in the polynucleotide encoding the C-terminal sequence. The purpose of the polynucleotides encoding the N-terminal and C-terminal sequences, in this case, is to facilitate homologous recombination, and one of ordinary skill in the art would recognize that these sequences may be longer or shorter than depicted below. Accordingly, in certain embodiments of the invention, the overall design of the CDRH3 repertoire, including the sequences required to facilitate homologous recombination with the selected chassis, can be represented by the following formula (regions homologous with vector underlined):



In some embodiments of the invention, the CDRH3 repertoire can be represented by the following formula, which excludes the T residue presented in the schematic above:



References describing collections of V, D, and J genes include Scaviner *et al.*, Exp. Clin. Immunogenet., 1999, 16: 243 and Ruiz *et al.*, Exp. Clin. Immunogenet, 1999, 16: 173, each incorporated by reference in its entirety.

Although homologous recombination is one method of producing the libraries of the invention, a person of ordinary skill in the art will readily recognize that other methods of DNA assembly, such as ligation or site-specific recombination, and/or DNA synthesis, can also be used to produce the libraries of the invention.

CDRH3 Lengths

The lengths of the segments may also be varied, for example, to produce libraries with a particular distribution of CDRH3 lengths. In one embodiment of the invention, the H3-JH segments are about 0 to about 10 amino acids in length, the DH segments are about 0 to about 12 amino acids in length, the TN1 segments are about 0 to about 4 amino acids in length, and the N2 segments are about 0 to about 4 amino acids in length. In certain embodiments, the H3-JH segments are at least about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and/or 10 amino acids in length. In some embodiments, the DH segments are at least about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 amino acids in length. In certain embodiments, the TN1 segments are at least about 0, 1, 2, 3, or 4 amino acids in length. In some embodiments, the N2 amino acids are at least about 0, 1, 2, 3, or 4 amino acids in length. In certain embodiments of the invention, the CDRH3 is about 2 to about 35, about 2 to about 28, or about 5 to about 26 amino acids in length. In some embodiments, the CDRH3 is at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, and/or 35 amino acids in length. In some embodiments, the length of any of the segments or CDRH3s of the invention may be less than a particular number of amino acids, where the number of amino acids is defined using any one of the integers provided above for the respective segment or CDRH3. In certain embodiments of the invention, a particular numerical range is defined, using any two of the integers provided above as lower and upper boundaries of

the range, inclusive or exclusive. All combinations of the integers provided, which define an upper and lower boundary, are contemplated.

Design of CDRL3 Libraries

The design of CDRL3 libraries, and light chain sequences, is described in detail in U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379, each of which is incorporated by reference in its entirety, and is therefore only described briefly herein. Libraries described herein are designed according to similar principles, with three important differences, namely that the libraries of the current invention contain (1) variability in CDRL1 and CDRL2; (2) variability in the framework regions; and/or (3) variability in CDRL3 that is designed to produce light chain libraries with CDRL3s that closely resemble human germline-like CDRL3 sequences, as defined above (Table 1).

A CDRL3 library of the invention may be a VKCDR3 library and/or a V λ CDR3 library. In certain embodiments of the invention, patterns of occurrence of particular amino acids at defined positions within VL sequences are determined by analyzing data available in public or other databases, for example, the NCBI database (see, for example, WO/2009/036379). In certain embodiments of the invention, these sequences are compared on the basis of identity and assigned to families on the basis of the germline genes from which they are derived. The amino acid composition at each position of the sequence, in each germline family, may then be determined. This process is illustrated in the Examples provided herein.

Light Chains with Framework Variability

In some embodiments, the invention provides a library of light chain variable domains wherein the light chain variable domains are varied at one or more of framework positions 2, 4, 36, 46, 48, 49, and 66. In some embodiments, the invention provides a library of light chain variable domains comprising at least a plurality of light chain variable domains whose amino acid sequences are identical to one another except for substitutions at one or more of positions 2, 4, 36, 46, 48, 49, and 66. In certain embodiments, the invention provides a library of light chain variable domains comprising at least a plurality of light chain variable domains whose amino acid sequences are at

least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, and/or 99.5% to any of the light chain variable domain sequences disclosed herein, and further have substitutions at one or more of positions 2, 4, 36, 46, 48, 49, and 66. In some embodiments, the amino acids selected for inclusion in these positions are selected from amongst about the most 2, 3, 4, 5, 6, 7, 8, 9, and/or 10 most frequently occurring amino acids at the corresponding position in a reference set of light chain variable domains.

In some embodiments, the invention provides systems and methods of selecting framework positions to be varied in a light chain variable domain, comprising:

- (i) obtaining a reference set of light chain sequences, wherein the reference set contains light chain sequences with VL segments selected from the group consisting of sequences found in, or encoded by, a single IGVL germline gene and/or sequences found in, or encoded by, allelic variants of the single IGVL germline gene;
- (ii) determining which framework positions within the reference set have a degree of variability that is similar to the degree of variability occurring in one more CDR positions of the sequences in the reference set (*e.g.*, the variability in a framework position is at least about 70%, 80%, 90%, or 95%, 100%, or more of the variability found in a CDR position of the sequences in the reference set);
- (iii) determining the frequency of occurrence of amino acid residues for each of the framework positions identified in (ii);
- (iv) synthesizing light chain variable domain encoding sequences wherein the framework positions identified in (ii) are varied to include the 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 most frequently occurring amino acid residues (identified in (iii)) at the corresponding position.

One of ordinary skill in the art, reading the present disclosure will appreciate that the present invention provides analogous methods for developing framework variants of heavy chain sequences.

Light Chains with CDR1 and/or CDR2 Variability

In some embodiments, the invention provides a library of light chain variable domains wherein the light chain variable domains are varied at one or more of CDRL1 positions 28, 29, 30, 30A, 30B, 30E, 31, and 32 (Chothia-Lesk numbering scheme; Chothia and Lesk, J. Mol. Biol., 1987, 196: 901). In some embodiments, the invention provides a library of light chain variable domains wherein the light chain variable domains are varied at one or more of CDRL2 positions 50, 51, 53, and 55. In some embodiments, the amino acids selected for inclusion in these CDRL1 and/or CDRL2 positions are selected from amongst about the most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and/or 20 most frequently occurring amino acids at the corresponding position in a reference set of light chain variable domains.

In some embodiments, the invention provides systems and methods for selecting CDRL1 and/or CDRL2 positions to be varied in a light chain variable domain, comprising:

- (i) obtaining a reference set of light chain sequences, wherein the reference set contains light chain sequences with VL segments selected from the group consisting of sequences found in, or encoded by, a single IGVL germline gene and sequences found in, or encoded by, allelic variants of the single IGVL germline gene;
- (ii) determining which CDRL1 and/or CDRL2 positions are variable within the reference set;
- (iii) synthesizing light chain variable domain encoding sequences wherein the CDRL1 and/or CDRL2 positions identified in (ii) are varied to include the 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 most frequently occurring amino acid residues at the corresponding position.

One of ordinary skill in the art, reading the present disclosure will appreciate that the present invention provides analogous methods for developing CDRH2 and/or CDRH2 variants of heavy chain sequences.

Light Chain Sequences

In some embodiments, the invention provides a light chain library comprising one or more of any of the light chain sequences provided herein, for example, the polypeptide sequences of Table 3 and/or Table 4 and/or the polynucleotide sequences of Table 5, Table 6, and/or Table 7. A person of ordinary skill in the art will recognize that not every light chain sequence provided herein is necessary to produce a functional light chain library of the invention. Therefore, in certain embodiments, a light chain library of the invention will contain a subset of the sequences described above. For example, in certain embodiments of the invention, at least about 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 10^3 , 10^4 , and/or 10^5 of the light chain polynucleotide and/or polypeptide sequences provided herein are included in a library. In some embodiments, a library of the invention may contain less than a particular number of polynucleotide or polypeptide segments, where the number of segments is defined using any one of the integers provided above for the respective segment. In certain embodiments of the invention, a particular numerical range is defined, using any two of the integers provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the integers provided, which define an upper and lower boundary, are contemplated.

In certain embodiments, the invention provides light chain libraries comprising at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the sequences from any of the sets of light chain sequences provided herein. For example, the invention provides libraries comprising at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the light chain sequences provided in Table 3, Table 4, Table 5, Table 6, and/or Table 7. In some embodiments of the invention, a particular percentage range is defined, using any two of the percentages provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the percentages provided, which define an upper and lower boundary, are contemplated.

In some embodiments of the invention, at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the light chain sequences in a library are light chain sequences provided herein.

In certain embodiments of the invention, at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% of the light chain sequences isolated from a light chain library (*e.g.*, by binding to a particular antigen and/or generic ligand) are light chain sequences provided herein. In some embodiments, a light chain library of the invention may contain less than a particular percentage of light chain sequences provided herein, where the percentage of light chain sequences is defined using any one of the percentages provided above. In certain embodiments of the invention, a particular percentage range is defined, using any two of the percentages provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the percentages provided, which define an upper and lower boundary, are contemplated.

One of ordinary skill in the art will further recognize that given the light chain sequences provided herein, similar light chain sequences could be produced which share a designated level of overall sequence identity and/or one or more characteristic sequence elements described herein, which overall degree of sequence identity and/or characteristic sequence elements may confer common functional attributes. Those of ordinary skill in the art will be well familiar with a variety of techniques for preparing such related sequences, , including the mutagenesis techniques provided herein. Therefore, each of the explicitly enumerated embodiments of the invention can also be practiced using light chain sequences that share a particular percent identity to any of the light chain sequences provided herein. For example, each of the previously described embodiments of the invention can be practiced using light chain sequences that are at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identical to the light chain sequences provided herein. For example, in some embodiments, light chain libraries provided by the invention comprise light chain variable domains at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identical to the light chain sequences provided herein, with substitutions in one or more of framework positions 2, 4, 36, 46, 48, 49, and 66, CDRL1 positions 28, 29, 30, 30A, 30B, 30E, 31, and 32 (Chothia-Lesk numbering scheme), and/or CDRL2 positions 50, 51, 53, and 55.

In some embodiments, the invention provides systems and methods for varying positions within the portion of CDRL3s encoded by a particular IGVL germline gene, comprising:

- (i) obtaining a reference set of light chain sequences, wherein the reference set contains light chain sequences with VL segments originating from the same IGVL germline gene and/or its allelic variants;
- (ii) determining which amino acids occur at each of the CDRL3 positions in the reference set that are encoded by the IGVL gene (*i.e.*, positions 89-94, inclusive);
- (iii) synthesizing light chain variable domain encoding sequences wherein two positions in each light chain variable domain encoding sequence contain degenerate codons encoding the 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 most frequently occurring amino acid residues at the corresponding positions in the reference set.

As described in the examples, the degenerate codons of (iii) can be chosen to best reproduce the amino acid diversity contained in the reference set for each of the two positions varied in each light chain. Finally, while the methods and systems described above are described with respect to CDRL3, one of ordinary skill in the art will readily recognize that the same principles can be applied to CDRH1 and/or CDRH2 of the heavy chain, which are encoded entirely by the IGHV gene.

CDRL3 Lengths

In some embodiments, as an alternative or in addition to other features described herein, the present invention provides libraries in which lengths of CDRL3s may be varied. The present invention therefore provides, among other things, libraries with a particular distribution of CDRL3 lengths. Although CDRL3 libraries of lengths 8, 9, and 10 are exemplified, one of ordinary skill in the art will readily recognize that the methods described herein can be applied to produce light chains with CDRL3s of different lengths (*e.g.*, about 5, 6, 7, 11, 12, 13, 14, 15, and/or 16) that also fall within the scope of the

invention. In some embodiments, the length of any of the CDRL3s of the invention may be less than a particular number of amino acids, where the number of amino acids is defined using any one of the integers provided above. In some embodiments of the invention, a particular numerical range is defined, using any two of the integers provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the integers provided, which define an upper and lower boundary, are contemplated.

Synthetic Antibody Libraries

In some embodiments of the invention, provided libraries include one or more synthetic polynucleotides. In some embodiments, provided libraries may comprise synthetic polynucleotides selected from (a) heavy chain chassis polynucleotides; (b) light chain chassis polynucleotides; (c) CDR3 polynucleotides; (d) constant domain polynucleotides; and (e) combinations thereof. Those of ordinary skill in the art will appreciate that such synthetic polynucleotides may be linked to other synthetic or non-synthetic polynucleotides in provided libraries.

Synthetic polynucleotides provided herein may be prepared by any available method. For example, in some embodiments, synthetic polynucleotides can be synthesized by split pool DNA synthesis as described in Feldhaus *et al.*, Nucleic Acids Research, 2000, 28: 534; Omstein *et al.*, Biopolymers, 1978, 17: 2341; Brenner and Lerner, PNAS, 1992, 87: 6378, U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379 (each incorporated by reference in its entirety).

In some embodiments of the invention, segments representing the possible TN1, DH, N2, and JH diversity found in the human repertoire are synthesized *de novo* either as double-stranded DNA oligonucleotides, single-stranded DNA oligonucleotides representative of the coding strand, or single-stranded DNA oligonucleotides representative of the non-coding strand. Such sequences can then be introduced into a host cell along with an acceptor vector containing a chassis sequence and, in some cases a portion of FRM4 and a constant region. No primer-based PCR amplification from mammalian cDNA or mRNA or template-directed cloning steps from mammalian cDNA or mRNA need be employed.

Construction of Libraries by Yeast Homologous Recombination

In certain embodiments, the invention exploits the inherent ability of yeast cells to facilitate homologous recombination at high efficiency. The mechanism of homologous recombination in yeast and its applications are briefly described below (also see *e.g.*, U.S. Patent Nos. 6,406,863; 6,410,246; 6,410,271; 6,610,472; and 7,700,302, each of which is incorporated by reference in its entirety).

As an illustrative embodiment, homologous recombination can be carried out in, for example, *Saccharomyces cerevisiae*, which has genetic machinery designed to carry out homologous recombination with high efficiency. Exemplary *S. cerevisiae* strains include EM93, CEN.PK2, RM11-1a, YJM789, and BJ5465. This mechanism is believed to have evolved for the purpose of chromosomal repair, and is also called “gap repair” or “gap filling”. By exploiting this mechanism, mutations can be introduced into specific loci of the yeast genome. For example, a vector carrying a mutant gene can contain two sequence segments that are homologous to the 5' and 3' open reading frame (ORF) sequences of a gene that is intended to be interrupted or mutated. The vector may also encode a positive selection marker, such as a nutritional enzyme allele (*e.g.*, URA3) and/or an antibiotic resistant marker (*e.g.*, Geneticin / G418), flanked by the two homologous DNA segments. Other selection markers and antibiotic resistance markers are known to one of ordinary skill in the art.

In some embodiments of the invention, this vector (*e.g.*, a plasmid) is linearized and transformed into the yeast cells. Through homologous recombination between the plasmid and the yeast genome, at the two homologous recombination sites, a reciprocal exchange of the DNA content occurs between the wild type gene in the yeast genome and the mutant gene (including the selection marker gene(s)) that is flanked by the two homologous sequence segments. By selecting for the one or more selection markers, the surviving yeast cells will be those cells in which the wild-type gene has been replaced by the mutant gene (Pearson *et al.*, *Yeast*, 1998, 14: 391, incorporated by reference in its entirety). This mechanism has been used to make systematic mutations in all 6,000 yeast genes, or open reading frames (ORFs), for functional genomics studies. Because the exchange is reciprocal, a similar approach has also been used successfully to clone yeast

genomic DNA fragments into a plasmid vector (Iwasaki *et al.*, Gene, 1991, 109: 81, incorporated by reference in its entirety).

By utilizing the endogenous homologous recombination machinery present in yeast, gene fragments or synthetic oligonucleotides can also be cloned into a plasmid vector without a ligation step. In this application of homologous recombination, a target gene fragment (*i.e.*, the fragment to be inserted into a plasmid vector, *e.g.*, a CDR3) is obtained (*e.g.*, by oligonucleotides synthesis, PCR amplification, restriction digestion out of another vector, etc.). DNA sequences that are homologous to selected regions of the plasmid vector are added to the 5' and 3' ends of the target gene fragment. These homologous regions may be fully synthetic, or added via PCR amplification of a target gene fragment with primers that incorporate the homologous sequences. The plasmid vector may include a positive selection marker, such as a nutritional enzyme allele (*e.g.*, URA3), or an antibiotic resistance marker (*e.g.*, Geneticin / G418). The plasmid vector is then linearized by a unique restriction cut located in-between the regions of sequence homology shared with the target gene fragment, thereby creating an artificial gap at the cleavage site. The linearized plasmid vector and the target gene fragment flanked by sequences homologous to the plasmid vector are co-transformed into a yeast host strain. The yeast is then able to recognize the two stretches of sequence homology between the vector and target gene fragment and facilitate a reciprocal exchange of DNA content through homologous recombination at the gap. As a consequence, the target gene fragment is inserted into the vector without ligation.

The method described above has also been demonstrated to work when the target gene fragments are in the form of single stranded DNA, for example, as a circular M13 phage derived form, or as single stranded oligonucleotides (Simon and Moore, Mol. Cell Biol., 1987, 7: 2329; Ivanov *et al.*, Genetics, 1996, 142: 693; and DeMarini *et al.*, 2001, 30: 520., each incorporated by reference in its entirety). Thus, the form of the target that can be recombined into the gapped vector can be double stranded or single stranded, and derived from chemical synthesis, PCR, restriction digestion, or other methods.

Several factors may influence the efficiency of homologous recombination in yeast. For example, the efficiency of the gap repair is correlated with the length of the homologous sequences flanking both the linearized vector and the target gene. In certain

embodiments, about 20 or more base pairs may be used for the length of the homologous sequence, and about 80 base pairs may give a near-optimized result (Hua *et al.*, Plasmid, 1997, 38: 91; Raymond *et al.*, Genome Res., 2002, 12: 190, each incorporated by reference in its entirety). In certain embodiments of the invention, at least about 5, 10, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 187, 190, or 200 homologous base pairs may be used to facilitate recombination. In certain embodiments, between about 20 and about 40 base pairs are utilized. In addition, the reciprocal exchange between the vector and gene fragment is strictly sequence-dependent, *i.e.* it does not cause a frame shift. Therefore, gap-repair cloning assures the insertion of gene fragments with both high efficiency and precision. The high efficiency makes it possible to clone two, three, or more targeted gene fragments simultaneously into the same vector in one transformation attempt (Raymond *et al.*, Biotechniques, 1999, 26: 134, incorporated by reference in its entirety). Moreover, the nature of precision sequence conservation through homologous recombination makes it possible to clone selected genes or gene fragments into expression or fusion vectors for direct functional examination (El-Deiry *et al.*, Nature Genetics, 1992, 1: 4549; Ishioka *et al.*, PNAS, 1997, 94: 2449, each incorporated by reference in its entirety).

Libraries of gene fragments have also been constructed in yeast using homologous recombination. For example, a human brain cDNA library was constructed as a two-hybrid fusion library in vector pJG4-5 (Guidotti and Zervos, Yeast, 1999, 15: 715, incorporated by reference in its entirety). It has also been reported that a total of 6,000 pairs of PCR primers were used for amplification of 6,000 known yeast ORFs for a study of yeast genomic protein interactions (Hudson *et al.*, Genome Res., 1997, 7: 1169, incorporated by reference in its entirety). In 2000, Uetz *et al.* conducted a comprehensive analysis-of protein-protein interactions in *Saccharomyces cerevisiae* (Uetz *et al.*, Nature, 2000, 403: 623, incorporated by reference in its entirety). The protein-protein interaction map of the budding yeast was studied by using a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins (Ito *et al.*, PNAS, 2000, 97: 1143, incorporated by reference in its entirety), and the genomic protein

linkage map of Vaccinia virus was studied using this system (McCraith *et al.*, PNAS, 2000, 97: 4879, incorporated by reference in its entirety).

In certain embodiments of the invention, a synthetic CDR3 (heavy or light chain) may be joined by homologous recombination with a vector encoding a heavy or light chain chassis, a portion of FRM4, and a constant region, to form a full-length heavy or light chain. In certain embodiments of the invention, the homologous recombination is performed directly in yeast cells. In some embodiments, such a method comprises:

- (a) transforming into yeast cells:
 - (i) a linearized vector encoding a heavy or light chain chassis, a portion of FRM4, and a constant region, wherein the site of linearization is between the end of FRM3 of the chassis and the beginning of the constant region; and
 - (ii) a library of CDR3 insert nucleotide sequences that are linear and double stranded, wherein each of the CDR3 insert sequences comprises a nucleotide sequence encoding CDR3 and 5'- and 3'-flanking sequences that are sufficiently homologous to the termini of the vector of (i) at the site of linearization to enable homologous recombination to occur between the vector and the library of CDR3 insert sequences; and
- (b) allowing homologous recombination to occur between the vector and the CDR3 insert sequences in the transformed yeast cells, such that the CDR3 insert sequences are incorporated into the vector, to produce a vector encoding full-length heavy chain or light chain.

As specified above, CDR3 inserts may have a 5' flanking sequence and a 3' flanking sequence that are homologous to the termini of the linearized vector. When the CDR3 inserts and the linearized vectors are introduced into a host cell, for example, a yeast cell, the "gap" (the linearization site) created by linearization of the vector is filled by the CDR3 fragment insert through recombination of the homologous sequences at the 5' and 3' termini of these two linear double-stranded DNAs (*i.e.*, the vector and the insert). Through this event of homologous recombination, libraries of circular vectors encoding full-length heavy or light chains comprising variable CDR3 inserts is generated. Particular instances of these methods are presented in the Examples.

Subsequent analysis may be carried out to determine, for example, the efficiency of homologous recombination that results in correct insertion of the CDR3 sequences into the vectors. For example, PCR amplification of the CDR3 inserts directly from selected yeast clones may reveal how many clones are recombinant. In certain embodiments, libraries with minimum of about 90% recombinant clones are utilized. In certain embodiments libraries with a minimum of about 1%, 5% 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% recombinant clones are utilized. The same PCR amplification of selected clones may also reveal the insert size.

To verify the sequence diversity of the inserts in the selected clones, a PCR amplification product with the correct size of insert may be “fingerprinted” with restriction enzymes known to cut or not cut within the amplified region. From a gel electrophoresis pattern, it may be determined whether the clones analyzed are of the same identity or of the distinct or diversified identity. The PCR products may also be sequenced directly to reveal the identity of inserts and the fidelity of the cloning procedure, and to prove the independence and diversity of the clones.

Expression and Screening Systems

Libraries of polynucleotides generated by any of the techniques described herein, or other suitable techniques, can be expressed and screened to identify antibodies having desired structure and/or activity. Expression of the antibodies can be carried out, for example, using cell-free extracts (and *e.g.*, ribosome display), phage display, prokaryotic cells (*e.g.*, bacterial display), or eukaryotic cells (*e.g.*, yeast display). In certain embodiments of the invention, the antibody libraries are expressed in yeast.

In some embodiments, polynucleotides are engineered to serve as templates that can be expressed in a cell-free extract. Vectors and extracts as described, for example in U.S. Patent Nos. 5,324,637; 5,492,817; 5,665,563, (each incorporated by reference in its entirety) can be used and many are commercially available. Ribosome display and other cell-free techniques for linking a polynucleotide (*i.e.*, a genotype) to a polypeptide (*i.e.*, a phenotype) can be used, *e.g.*, ProfusionTM (see, *e.g.*, U.S. Patent Nos. 6,348,315; 6,261,804; 6,258,558; and 6,214,553, each incorporated by reference in its entirety).

Alternatively or additionally, polynucleotides of the invention can be expressed in an *E. coli* expression system, such as that described by Pluckthun and Skerra. (Meth. Enzymol., 1989, 178: 476; Biotechnology, 1991, 9: 273, each incorporated by reference in its entirety). Mutant proteins can be expressed for secretion in the medium and/or in the cytoplasm of the bacteria, as described by Better and Horwitz, Meth. Enzymol., 1989, 178: 476, incorporated by reference in its entirety. In some embodiments, the single domains encoding VH and VL are each attached to the 3' end of a sequence encoding a signal sequence, such as the ompA, phoA or pelB signal sequence (Lei *et al.*, J. Bacteriol., 1987, 169: 4379, incorporated by reference in its entirety). These gene fusions are assembled in a dicistronic construct, so that they can be expressed from a single vector, and secreted into the periplasmic space of *E. coli* where they will refold and can be recovered in active form. (Skerra *et al.*, Biotechnology, 1991, 9: 273, incorporated by reference in its entirety). For example, antibody heavy chain genes can be concurrently expressed with antibody light chain genes to produce antibodies or antibody fragments.

In some embodiments of the invention, antibody sequences are expressed on the membrane surface of a prokaryote, *e.g.*, *E. coli*, using a secretion signal and lipidation moiety as described, *e.g.*, in US2004/0072740; US2003/0100023; and US2003/0036092 (each incorporated by reference in its entirety).

Higher eukaryotic cells, such as mammalian cells, for example myeloma cells (*e.g.*, NS/O cells), hybridoma cells, Chinese hamster ovary (CHO), and human embryonic kidney (HEK) cells, can also be used for expression of the antibodies of the invention. Typically, antibodies expressed in mammalian cells are designed to be secreted into the culture medium, or expressed on the surface of the cell. Antibody or antibody fragments can be produced, for example, as intact antibody molecules or as individual VH and VL fragments, Fab fragments, single domains, or as single chains (scFv) (Huston *et al.*, PNAS, 1988, 85: 5879, incorporated by reference in its entirety).

Alternatively or additionally, antibodies can be expressed and screened by anchored periplasmic expression (APEX 2-hybrid surface display), as described, for example, in Jeong *et al.*, PNAS, 2007, 104: 8247 (incorporated by reference in its entirety) or by other anchoring methods as described, for example, in Mazor *et al.*, Nature Biotechnology, 2007, 25: 563 (incorporated by reference in its entirety).

In some embodiments of the invention, antibodies can be selected using mammalian cell display (Ho *et al.*, PNAS, 2006, 103: 9637, incorporated by reference in its entirety).

Screening of the antibodies derived from the libraries of the invention can be carried out by any appropriate means. For example, binding activity can be evaluated by standard immunoassay and/or affinity chromatography. Screening of antibodies of the invention for catalytic function, *e.g.*, proteolytic function can be accomplished using a standard assays, *e.g.*, the hemoglobin plaque assay as described in U.S. Patent No. 5,798,208 (incorporated by reference in its entirety). Determining the ability of candidate antibodies to bind therapeutic targets can be assayed *in vitro* using, *e.g.*, a BIACORETM instrument, which measures binding rates of an antibody to a given target or antigen based on surface plasmon resonance. *In vivo* assays can be conducted using any of a number of animal models and then subsequently tested, as appropriate, in humans. Cell-based biological assays are also contemplated.

One feature of the instant invention is the speed at which the antibodies of the library can be expressed and screened. In certain embodiments of the invention, the antibody library can be expressed in yeast, which have a doubling time of less than about 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 hours. In some embodiments, the doubling times are about 1 to about 3 hours, about 2 to about 4, about 3 to about 8 hours, about 3 to about 24, about 5 to about 24, about 4 to about 6 about 5 to about 22, about 6 to about 8, about 7 to about 22, about 8 to about 10 hours, about 7 to about 20, about 9 to about 20, about 9 to about 18, about 11 to about 18, about 11 to about 16, about 13 to about 16, about 16 to about 20, or about 20 to about 30 hours. In certain embodiments of the invention, an antibody library is expressed in yeast with a doubling time of about 16 to about 20 hours, about 8 to about 16 hours, or about 4 to about 8 hours. Thus, an antibody library of the instant invention can be expressed and screened in a matter of hours, as compared to previously known techniques which take several days to express and screen antibody libraries. A limiting step in the throughput of such screening processes in mammalian cells is typically the time required to iteratively regrow populations of isolated cells, which, in some cases, have doubling times greater than the doubling times of the yeast used in the current invention.

In certain embodiments of the invention, the composition of a library may be defined after one or more enrichment steps (for example by screening for antigen binding, binding to a generic ligand, or other properties). For example, a library with a composition comprising about x% sequences or libraries of the invention may be enriched to contain about 2x%, 3x%, 4x%, 5x%, 6x%, 7x%, 8x%, 9x%, 10x%, 20x%, 25x%, 40x%, 50x%, 60x%, 75x%, 80x%, 90x%, 95x%, or 99x% sequences or libraries of the invention, after one or more screening steps. In some embodiments of the invention, the sequences or libraries of the invention may be enriched about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 100-fold, 1,000-fold, or more, relative to their occurrence prior to the one or more enrichment steps. In certain embodiments of the invention, a library may contain at least a certain number of a particular type of sequence(s), such as CDRH3s, CDRL3s, heavy chains, light chains, or whole antibodies (e.g., at least about 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , 10^{14} , 10^{15} , 10^{16} , 10^{17} , 10^{18} , 10^{19} , or 10^{20}). In certain embodiments, these sequences may be enriched during one or more enrichment steps, to provide libraries comprising at least about 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , 10^{14} , 10^{15} , 10^{16} , 10^{17} , 10^{18} , or 10^{19} of the respective sequence(s).

Mutagenesis Approaches for Affinity Maturation

As described above, antibody leads can be identified through a selection process that involves screening the antibodies of a library of the invention for binding to one or more antigens, or for a biological activity. Coding sequences of these antibody leads may be further mutagenized *in vitro* or *in vivo* to generate secondary libraries with diversity introduced in the context of the initial antibody leads. Such mutagenized antibody leads can then be further screened for binding to target antigens or biological activity, *in vitro* or *in vivo*, following procedures similar to those used for the selection of the initial antibody lead from the primary library. Such mutagenesis and selection of primary antibody leads effectively mimics the affinity maturation process naturally occurring in a mammal that produces antibodies with progressive increases in the affinity to an antigen.

In some embodiments of the invention, only the CDRH3 region is mutagenized. In some embodiments of the invention, the whole variable region is mutagenized. In

some embodiments of the invention one or more of CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, and/ CDRL3 may be mutagenized. In some embodiments of the invention, “light chain shuffling” may be used as part of the affinity maturation protocol. In certain embodiments, this may involve pairing one or more heavy chains with a number of light chains, to select light chains that enhance the affinity and/or biological activity of an antibody. In certain embodiments of the invention, the number of light chains to which the one or more heavy chains can be paired is at least about 2, 5, 10, 100, 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , or 10^{10} . In certain embodiments of the invention, these light chains are encoded by plasmids. In some embodiments of the invention, the light chains may be integrated into the genome of the host cell.

Coding sequences of antibody leads may be mutagenized using any of wide variety of methods. Examples of methods of mutagenesis include, but are not limited to site-directed mutagenesis, error-prone PCR mutagenesis, cassette mutagenesis, and random PCR mutagenesis. Alternatively or additionally, oligonucleotides encoding regions with the desired mutations can be synthesized and introduced into the sequence to be mutagenized, for example, via recombination or ligation.

Site-directed mutagenesis or point mutagenesis may be used to gradually change the CDR sequences in specific regions. For example, this may be accomplished by using oligonucleotide-directed mutagenesis or PCR. For example, a short sequence of an antibody lead may be replaced with a synthetically mutagenized oligonucleotide in either the heavy chain or light chain region, or both. Such a method may not be efficient for mutagenizing large numbers of CDR sequences, but may be used for fine tuning of a particular lead to achieve higher affinity toward a specific target protein.

Cassette mutagenesis may alternatively or additionally be used to mutagenize the CDR sequences in specific regions. In a typical cassette mutagenesis, a sequence block, or a region, of a single template is replaced by a completely or partially randomized sequence. However, the maximum information content that can be obtained may be statistically limited by the number of random sequences of the oligonucleotides. Similar to point mutagenesis, this method may also be used for fine tuning of a particular lead to achieve higher affinity towards a specific target protein.

Error-prone PCR, or "poison" PCR, may be used to mutagenize the CDR sequences, for example, by following protocols described in U.S. Patent No. 6,153,745; Caldwell and Joyce, PCR Methods and Applications, 1992, 2: 28; Leung *et al.*, Technique, 1989, 1: 11; Shafikhani *et al.*, Biotechniques, 1997, 23: 304; and Stemmer *et al.*, PNAS, 1994, 91: 10747 (each of which is incorporated by reference in its entirety).

Conditions for error prone PCR may include, for example, (a) high concentrations of Mn^{2+} (*e.g.*, about 0.4 to about 0.6 mM) that efficiently induces malfunction of Taq DNA polymerase; and/or (b) a disproportionally high concentration of one nucleotide substrate (*e.g.*, dGTP) in the PCR reaction that causes incorrect incorporation of this high concentration substrate into the template and produces mutations. Alternatively or additionally, other factors such as, the number of PCR cycles, the species of DNA polymerase used, and the length of the template, may affect the rate of misincorporation of "wrong" nucleotides into the PCR product. Commercially available kits may be utilized for the mutagenesis of the selected antibody library, such as the "Diversity PCR random mutagenesis kit" (CLONTECHTM).

Primer pairs used in PCR-based mutagenesis may, in certain embodiments, include regions matched with the homologous recombination sites in the expression vectors. Such a design allows facile re-introduction of the PCR products back into the heavy or light chain chassis vectors, after mutagenesis, via homologous recombination.

Other PCR-based mutagenesis methods can also be used, alone or in conjunction with the error prone PCR described above. For example, the PCR amplified CDR segments may be digested with DNase to create nicks in the double stranded DNA. These nicks can be expanded into gaps by other exonucleases such as Bal 31. Gaps may then be filled by random sequences by using DNA Klenow polymerase at a low concentration of regular substrates dGTP, dATP, dTTP, and dCTP with one substrate (*e.g.*, dGTP) at a disproportionately high concentration. This fill-in reaction should produce high frequency mutations in the filled gap regions. Such methods of DNase digestion may be used in conjunction with error prone PCR to create a high frequency of mutations in the desired CDR segments.

CDR or antibody segments amplified from the primary antibody leads may also be mutagenized *in vivo* by exploiting the inherent ability of mutation in pre-B cells. The

Ig genes in pre-B cells are specifically susceptible to a high-rate of mutation. The Ig promoter and enhancer facilitate such high rate mutations in a pre-B cell environment while the pre-B cells proliferate. Accordingly, CDR gene segments may be cloned into a mammalian expression vector that contains a human Ig enhancer and promoter. Such a construct may be introduced into a pre-B cell line, such as 38B9, which allows the mutation of the VH and VL gene segments naturally in the pre-B cells (Liu and Van Ness, Mol. Immunol., 1999, 36: 461, incorporated by reference in its entirety). The mutagenized CDR segments can be amplified from the cultured pre-B cell line and re-introduced back into the chassis-containing vector(s) via, for example, homologous recombination.

In some embodiments, a CDR “hit” isolated from screening the library can be re-synthesized, for example using degenerate codons or trinucleotides, and re-cloned into the heavy or light chain vector using gap repair.

Other Variants of Polynucleotide Sequences of the Invention

In certain embodiments, the invention provides a polynucleotide that hybridizes with a polynucleotide taught herein, or that hybridizes with the complement of a polynucleotide taught herein. For example, an isolated polynucleotide that remains hybridized after hybridization and washing under low, medium, or high stringency conditions to a polynucleotide taught herein or the complement of a polynucleotide taught herein is encompassed by the present invention.

Exemplary low stringency conditions include hybridization with a buffer solution of about 30% to about 35% formamide, about 1 M NaCl, about 1% SDS (sodium dodecyl sulphate) at about 37°C, and a wash in about 1X to about 2X SSC (20X SSC=3.0 M NaCl/0.3 M trisodium citrate) at about 50°C to about 55°C.

Exemplary moderate stringency conditions include hybridization in about 40% to about 45% formamide, about 1 M NaCl, about 1% SDS at about 37°C, and a wash in about 0.5X to about 1X SSC at about 55°C to about 60°C.

Exemplary high stringency conditions include hybridization in about 50% formamide, about 1 M NaCl, about 1% SDS at about 37°C, and a wash in about 0.1X SSC at about 60°C to about 65°C.

Optionally, wash buffers may comprise about 0.1% to about 1% SDS.

The duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours.

Sublibraries and Larger Libraries Comprising Libraries or Sub-Libraries of the Invention

Libraries comprising combinations of the libraries described herein (*e.g.*, CDRH3 and CDRL3 libraries) are encompassed by the invention. Sublibraries comprising portions of the libraries described herein are also encompassed by the invention (*e.g.*, a CDRH3 library in a particular heavy chain chassis or a sub-set of the CDRH3 libraries, for example based on length).

Moreover, libraries containing one of the libraries or sublibraries of the invention also fall within the scope of the invention. For example, in certain embodiments of the invention, one or more libraries or sublibraries of the invention may be contained within a larger library (theoretical or physical), which may include sequences derived by other means, for example, non-human or human sequence derived by stochastic or sitewise-stochastic synthesis. In certain embodiments of the invention, at least about 1% of the sequences in a polynucleotide library may be those of the invention (*e.g.*, CDRH3 sequences, CDRL3 sequences, VH sequences, VL sequences), regardless of the composition of the other 99% of sequences. For the purposes of illustration only, one of ordinary skill in the art would readily recognize that a library containing 10^9 total members, where 10^7 members are members of the libraries of the invention (*i.e.*, 1%) would have utility, and that members of the libraries of the invention could be isolated from such a library. In some embodiments of the invention, at least about 0.001%, 0.01%, 0.1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% of the sequences in any polynucleotide library may be those of the invention, regardless of the composition of the other sequences. In some embodiments, the sequences of the invention may comprise about 0.001% to about 1%, about 1% to about 2%, about 2% to about 5%, about 5% to about 10%, about 10% to about 15%, about 15% to about 20%, about 20% to about 25%, about 25% to about 30%, about 30% to about

35%, about 35% to about 40%, about 40% to about 45%, about 45% to about 50%, about 50% to about 55%, about 55% to about 60%, about 60% to about 65%, about 65% to about 70%, about 70% to about 75%, about 75% to about 80%, about 80% to about 85%, about 85% to about 90%, about 90% to about 95%, or about 95% to about 99% of the sequences in any polynucleotide library, regardless of the composition of the other sequences. Thus, libraries more diverse than one or more libraries or sublibraries of the invention, but yet still comprising one or more libraries or sublibraries of the invention, in an amount in which the one or more libraries or sublibraries of the invention can be effectively screened and from which sequences encoded by the one or more libraries or sublibraries of the invention can be isolated, also fall within the scope of the invention.

Alternative Scaffolds

As would be evident to one of ordinary skill in the art, the CDRH3 and/or CDRL3 polypeptides provided by the invention may also be displayed on alternative scaffolds. Several such scaffolds have been shown to yield molecules with specificities and affinities that rival those of antibodies. Exemplary alternative scaffolds include those derived from fibronectin (*e.g.*, AdNectin), the β -sandwich (*e.g.*, iMab), lipocalin (*e.g.*, Anticalin), EETI-II/AGRP, BPTI/LACI-D1/ITI-D2 (*e.g.*, Kunitz domain), thioredoxin (*e.g.*, peptide aptamer), protein A (*e.g.*, Affibody), ankyrin repeats (*e.g.*, DARPin), γ B-crystallin/ubiquitin (*e.g.*, Affilin), CTLD₃ (*e.g.*, Tetranectin), and (LDLR-A module)₃ (*e.g.*, Avimers). Additional information on alternative scaffolds is provided, for example, in Binz *et al.*, Nat. Biotechnol., 2005 23: 1257 and Skerra, Current Opin. in Biotech., 2007 18: 295-304, each of which is incorporated by reference in its entirety.

Additional Embodiments of the Invention

Library Sizes

In some embodiments of the invention, a library comprises about 10^1 to about 10^{20} different polynucleotide or polypeptide sequences (encoding or comprising *e.g.*, antibodies, heavy chains, CDRH3s, light chains, and/or CDRL3s). In some embodiments, the libraries of the invention are designed to include at least about 10^1 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , 10^{14} , 10^{15} , 10^{16} , 10^{17} , 10^{18} , 10^{19} , or

10^{20} , or more different antibody, heavy chain, CDRH3, light chain, and/or CDRL3 polynucleotide or polypeptide sequences. In some embodiments, a library of the invention may contain less than a particular number of polynucleotide or polypeptide sequences, where the number of sequences is defined using any one of the integers provided above. In certain embodiments of the invention, a particular numerical range is defined, using any two of the integers provided above as lower and upper boundaries of the range, inclusive or exclusive. All combinations of the integers provided, which define an upper and lower boundary, are contemplated.

In some embodiments, the invention provides libraries wherein a fraction of the members of the library are members produced according to the methods, systems, and compositions provided herein. One important property of the libraries of the invention is that they favorably mimic certain aspects of the human preimmune repertoire, including length diversity and sequence diversity. One of ordinary skill in the art will readily recognize that libraries provided by the invention include libraries where a subset of the members of the library are members produced according to the methods, systems, and compositions provided herein. For example, a library containing 10^8 members wherein 10^6 members are produced according to the methods, systems, and compositions provided herein, would contain 1% sequences produced according to the methods, systems, and compositions provided herein. One of ordinary skill in the art would recognize that one or more of the 10^6 members could readily be isolated using screening techniques known in the art. Therefore, said libraries fall within the scope of the invention. More specifically, libraries comprising at least about 1%, 2.5%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% CDRH3, CDRL3, light chain, or heavy chain, and/or full-length antibody sequences provided herein fall within the scope of the invention. Libraries comprising at least about 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , 10^{14} , 10^{15} CDRH3, CDRL3, light chain, heavy chain, and/or full-length antibody sequences provided herein also fall within the scope of the invention.

Human Preimmune Set

In some embodiments, the invention comprises the set of 3,571 curated human preimmune antibody sequences contained within the HPS, their corresponding CDRH3 sequences (Appendix A), and/or a representation of these CDRH3 sequences (and/or TN1, DH, N2, and/or H3-JH segments thereof) in a computer readable format. In certain embodiments, the invention comprises a method of producing a CDRH3 library, the method comprising matching candidate segments (*i.e.*, TN1, DH, N2, and H3-JH) from a theoretical segment pool with CDRH3 sequences in the HPS and/or any other repertoire of CDRH3 sequences. In some embodiments, the invention comprises the candidate segments from the theoretical segment pools disclosed herein and/or the segments selected for inclusion in a physical library.

Embodiments

While the methods described herein demonstrate the production of theoretical segment pools of H3-JH and DH segments using a limited number of allelic variants, one of ordinary skill in the art will recognize that methods taught herein may be applied to any IGHJ and IGHD genes, including any other allelic variants and all non-human IGHJ and IGHD genes. Alternatively or additionally, methods described herein may be applied to any reference set of CDRH3 sequences, for example to extract additional TN1 and/or N2 segments. Alternatively or additionally, one of ordinary skill in the art will recognize that each of the described embodiments of the invention may be in polynucleotide or polypeptide form, within a vector, virus, or microorganism (*e.g.*, a yeast or bacteria). Furthermore, since the invention involves synthetic libraries that are fully enumerated, certain embodiments of the invention relate to any of the embodiments described above in a computer readable format, and uses thereof.

Non-human antibody libraries also fall within the scope of the invention.

The present disclosure describes the removal of sequences containing Cys residues, N-linked glycosylation motifs, deamidation motifs, and highly hydrophobic sequences from the libraries of the invention. One of ordinary skill in the art will recognize that one or more of these criteria (*i.e.*, not necessarily all) can be applied to remove undesirable sequences from any library of the invention. However, libraries

containing one or more of these types of sequences also fall within the scope of the invention. Other criteria can also be used; those described herein are not limiting.

In certain embodiments, the invention provides libraries in which the number of times a particular sequence is repeated within the library (either theoretical, synthetic, or physical realization) is limited. For example, in some embodiments, the invention provides libraries wherein the frequency of occurrence of any of the sequences in the library (*e.g.*, CDRH3, CDRL3, heavy chain, light chain, full-length antibody) is less than about 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the frequency of occurrence of any of the sequences in the library is less than a multiple of the frequency of occurrence of any other sequence in the library, for examples less than about 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 times the frequency of occurrence of any other sequence in the library.

In some embodiments, libraries are defined by the combinatorial diversity of the segments used to produce CDRH3 sequences, in particular the number of non-degenerate segment combinations that can be used to produce a particular CDRH3 sequence. In some embodiments, this metric may be calculated using, for example, a sample of about 2000, 5000, 10000, 20000, 50000, 100000, or more sequences from the CDRH3 library and “self-matching” using the segments used to generate the CDRH3 sequences of that library. In certain embodiments, the invention provides libraries wherein at least about 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% of the CDRH3 sequences in the library may be formed by a single combination of segments.

In certain embodiments of the invention, a statistical bootstrap analysis was used to generate CDRH3 reference sets. While it may be advantageous to use this method, it is not required for every embodiment of the invention.

In some embodiments, the invention provides methods and systems of selecting polynucleotides to encode polypeptides of the invention, comprising selecting polynucleotide segments lacking (or containing) certain restriction sites individually and/or after combinatorial concatenation with other segments (*e.g.*, see Example 9.3.7).

The exemplary libraries provided herein are not limiting and provided for exemplification only.

EXAMPLES

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

In general, the practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, recombinant DNA technology, PCR technology, immunology (especially, *e.g.*, antibody technology), expression systems (*e.g.*, yeast expression, cell-free expression, phage display, ribosome display, and PROFUSIONTM), and any necessary cell culture that are within the skill of the art and are explained in the literature. See, *e.g.*, Sambrook, Fritsch and Maniatis, *Molecular Cloning: Cold Spring Harbor Laboratory Press* (1989); *DNA Cloning*, Vols. 1 and 2, (D.N. Glover, Ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait, Ed. 1984); *PCR Handbook Current Protocols in Nucleic Acid Chemistry*, Beaucage, Ed. John Wiley & Sons (1999) (Editor); *Oxford Handbook of Nucleic Acid Structure*, Neidle, Ed., Oxford Univ Press (1999); *PCR Protocols: A Guide to Methods and Applications*, Innis *et al.*, Academic Press (1990); *PCR Essential Techniques: Essential Techniques*, Burke, Ed., John Wiley & Son Ltd (1996); *The PCR Technique: RT-PCR*, Siebert, Ed., Eaton Pub. Co. (1998); *Antibody Engineering Protocols (Methods in Molecular Biology)*, 510, Paul, S., Humana Pr (1996); *Antibody Engineering: A Practical Approach (Practical Approach Series, 169)*, McCafferty, Ed., Irl Pr (1996); *Antibodies: A Laboratory Manual*, Harlow *et al.*, C.S.H.L. Press, Pub. (1999); *Current Protocols in Molecular Biology*, eds. Ausubel *et al.*, John Wiley & Sons (1992); *Large-Scale Mammalian Cell Culture Technology*, Lubiniecki, A., Ed., Marcel Dekker, Pub., (1990); *Phage Display: A Laboratory Manual*, C. Barbas (Ed.), CSHL Press, (2001); *Antibody Phage Display*, P O'Brien (Ed.), Humana Press (2001); Border *et al.*, *Nature Biotechnology*, 1997, 15: 553; Border *et al.*, *Methods Enzymol.*, 2000, 328: 430; ribosome display as described by Pluckthun *et al.* in U.S. Patent No. 6,348,315, and ProfusionTM as described by Szostak *et al.* in U.S. Patent Nos. 6,258,558; 6,261,804; and 6,214,553; and bacterial periplasmic expression as described

in US20040058403A1. Each of the references cited in this paragraph is incorporated by reference in its entirety.

Further details regarding antibody sequence analysis using Kabat conventions and programs to analyze aligned nucleotide and amino acid sequences may be found, *e.g.*, in Johnson *et al.*, Methods Mol. Biol., 2004, 248: 11; Johnson *et al.*, Int. Immunol., 1998, 10: 1801; Johnson *et al.*, Methods Mol. Biol., 1995, 51: 1; Wu *et al.*, Proteins, 1993, 16: 1; and Martin, Proteins, 1996, 25: 130. Each of the references cited in this paragraph is incorporated by reference in its entirety.

Further details regarding antibody sequence analysis using Chothia conventions may be found, *e.g.*, in Chothia *et al.*, J. Mol. Biol., 1998, 278: 457; Morea *et al.*, Biophys. Chem., 1997, 68: 9; Morea *et al.*, J. Mol. Biol., 1998, 275: 269; Al-Lazikani *et al.*, J. Mol. Biol., 1997, 273: 927. Barre *et al.*, Nat. Struct. Biol., 1994, 1: 915; Chothia *et al.*, J. Mol. Biol., 1992, 227: 799; Chothia *et al.*, Nature, 1989, 342: 877; and Chothia *et al.*, J. Mol. Biol., 1987, 196: 901. Further analysis of CDRH3 conformation may be found in Shirai *et al.*, FEBS Lett., 1999, 455: 188 and Shirai *et al.*, FEBS Lett., 1996, 399: 1. Further details regarding Chothia analysis are described, for example, in Chothia *et al.*, Cold Spring Harb. Symp. Quant Biol., 1987, 52: 399. Each of the references cited in this paragraph is incorporated by reference in its entirety.

Further details regarding CDR contact considerations are described, for example, in MacCallum *et al.*, J. Mol. Biol., 1996, 262: 732, incorporated by reference in its entirety.

Further details regarding the antibody sequences and databases referred to herein are found, *e.g.*, in Tomlinson *et al.*, J. Mol. Biol., 1992, 227: 776, VBASE2 (Retter *et al.*, Nucleic Acids Res., 2005, 33: D671); BLAST (www.ncbi.nlm.nih.gov/BLAST/); CDHIT (bioinformatics.ljcrf.edu/cd-hi/); EMBOSS (www.hgmp.mrc.ac.uk/Software/EMBOSS/); PHYLIP (evolution.genetics.washington.edu/phylip.html); and FASTA (fasta.bioch.virginia.edu). Each of the references cited in this paragraph is incorporated by reference in its entirety.

LIGHT CHAIN LIBRARIES

Example 1. Light Chain Libraries with Framework and/or CDRL1 and/or CDRL2 Variability

Although the diversity in antibody sequences is concentrated in the CDRs, certain residues in the framework regions can also influence antigen recognition and/or modulate affinity (Queen *et al.*, Proc. Natl. Acad. Sci. USA, 1989, 86: 10029; Carter *et al.*, Proc. Natl. Acad. Sci. USA, 1992, 89: 4285, each incorporated by reference in its entirety). These residues have been cataloged and used to make framework substitutions that improve antibody affinity, for example, during the process of antibody humanization (*e.g.*, see the “Vernier” residues in Foote and Winter, J. Mol. Biol., 1992, 224: 487, incorporated by reference in its entirety). In the heavy chain, the Vernier residues include Kabat-numbered residues 2, 27-30, 47-49, 67, 69, 71, 73, 78, 93-94, and 103. In the light chain, the Vernier residues include Kabat residues 2, 4, 35-36, 46-49, 64, 66, 68-69, 71, and 98. The Vernier residue numbers are the same for kappa and lambda light chain sequences (*see* Table 4 in Chothia *et al.*, J. Mol. Biol., 1985, 186: 651, which is incorporated by reference in its entirety). Additionally, framework positions at the VL-VH interface may also influence affinity. In the heavy chain, the interface residues include Kabat residues 35, 37, 39, 45, 47, 91, 93, 95, 100, and 103 (Chothia *et al.*, J. Mol. Biol., 1985, 186: 651, incorporated by reference in its entirety). In the light chain, the interface residues include Kabat residues 34, 36, 38 44, 46, 87, 89, 91, 96, and 98.

The following procedure was used to select the framework residues to be varied and the amino acids to which they should be varied:

- a. A collection of human VK light chain DNA sequences was obtained from NCBI (*see* Appendix A of WO/2009/036379 for GI Nos.). These sequences were classified according to the germline origin of their VK germline segment.
- b. Patterns of variation at each of the Vernier and interface positions were examined as follows:
 - i. Equation 1 (from Makowski & Soares, Bioinformatics, 2003, 19: 483, incorporated by reference in its entirety) was used to calculate a

diversity index for the Vernier positions, interface positions, CDRL1, and CDRL2.

$$d = \frac{1}{N \sum p_i^2} \quad \text{Equation 1}$$

Here, d is the diversity index, N is 20, the total number of amino acid types, and p_i is the fraction of amino acid of type “i” at the position of interest. The sum is carried out over the 20 amino acid types. The parameter d will attain its minimum value of 0.05 or $1/20$, when a single amino acid type is observed at a given position: p_i is 1 for one type and zero for all the rest. Conversely, when all the amino acid types are equally probable (*e.g.*, p_i is 0.05 for all i), d will attain its maximum value of 1.0.

- ii. The diversity index for each of the Vernier and interface positions were compared to the diversity index for the positions in CDRL1 and CDRL2.
- iii. The interface positions were found to be relatively invariant, with d values very close to the minimum of 0.05, and were thus not altered. The Vernier residues with a diversity index comparable to or larger than that of the CDR positions (*i.e.*, at or above 0.07 for the particular example provided in Figure 1) were selected as candidates for variance (see Figure 1). The amino acid residues included in these positions were selected from amongst the two to three amino acids most frequently occurring in that position in the sequences in the collection of human VK light chains, for each particular VK germline.
- iv. Table 2 shows the positions selected for variance in each of nine exemplified light chain germlines. The alternative framework positions represent positions with a diversity index less than the primary framework positions, but where variability may still be incorporated to influence antigen binding.

v. The amino acid residues in the framework positions selected for variance were varied as follows (Table 3 provides the polypeptide sequences of these variants):

1. **Position 2:** Germline I was optionally changed to V.
2. **Position 4:** Germline M or L was optionally changed to L or M. In some embodiments, changes from M to L, but not the reverse, may be preferred, because M may undergo oxidation during production, processing, or storage, potentially altering the properties of the antibody.
3. **Position 36:** Germline Y was optionally changed to F and H.
4. **Position 46:** Germline L was optionally changed to V.
5. **Position 48:** Germline I was optionally changed to L.
6. **Position 49:** Germline Y was optionally changed to S, F, and H.
7. **Position 66:** Germline G was optionally changed to R and E.

One of ordinary skill in the art would readily recognize that the procedure outlined above could also be used to select positions to vary in V λ germline sequences, and that libraries containing V λ chains also fall within the scope of the invention.

In addition to the framework mutations, variability was also introduced into CDRL1 and CDRL2. This was performed by determining which residues in CDRL1 and CDRL2 were variable, within a particular germline, in the VK dataset used above and incorporating the most frequently occurring 2 to 4 variants into CDRL1 and CDRL2 in the synthetic libraries of the invention. With the exception of position 50 of CDRL2 of the VK1-5 germline, these alternatives did not arise from allelic variation. Table 3 shows the polypeptide sequences of nine light chain chassis and their framework and CDR L1/L2 variants for the currently exemplified embodiment of the invention. The amino acid residues in the CDRL1/L2 positions selected for variance were varied as follows (using the Chothia-Lesk numbering system; Chothia and Lesk, J. Mol. Biol., 1987, 196: 901):

1. **Position 28:** Germline S or G were optionally changed to G, A, or D.
2. **Position 29:** Germline V was optionally changed to I.
3. **Position 30:** Germline S was optionally changed to N, D, G, T, A, or R.
4. **Position 30A:** Germline H was optionally changed to Y
5. **Position 30B:** Germline S was optionally changed to R or T.
6. **Position 30E:** Germline Y was optionally changed to N.
7. **Position 31:** Germline S was optionally changed to D, R, I, N, or T.
8. **Position 32:** Germline Y or N were optionally changed to F, S, or D.
9. **Position 50:** Germline A, D, or G were optionally changed to G, S, E, K, or D.
10. **Position 51:** Germline G or A were optionally changed to A, S, or T.
11. **Position 53:** Germline S or N were optionally changed to N, H, S, K, or R.
12. **Position 55:** Germline E was optionally changed to A or Q.

Example 2. Light Chain Libraries with Enhanced Diversity in CDRL3

A variety of methods of producing light chain libraries are known in the art (*e.g.*, see U.S. Publication Nos. 2009/0181855, 2010/0056386, and WO/2009/036379). An analysis of clinically validated antibody sequences indicated that these sequences have very little deviation from germline-like VL-JL (where “L” can be a kappa or lambda germline sequence) rearrangements prior to somatic mutation (Figure 2). Here, a germline-like rearrangement is one where neither the V nor J portion differ from the respective germline genes and, for the purposes of this particular example, where the length of CDRL3 is restricted to 8, 9 or 10 amino acids (*see* U.S. Publication Nos.

2009/0181855, 2010/0056386, and WO/2009/036379). For the IGHJK1 gene, however, both WT (Trp-Thr) and RT (Arg-Thr) sequences (the first two N-terminal residues) are considered “germline-like” and so are full L3 rearrangements containing such sequences. Therefore, new light chain libraries were designed and constructed with the objectives of simultaneously (1) minimizing deviation from germline-like sequences, as defined above; and (2) generating maximal diversity. In particular, the overarching goal was to maximize the *type* of diversity that is indicated to be most favorable by clinically validated antibody sequences. In particular, the designed library sought to maximize the diversity of CDRL3 sequences that differ from length-matched germline sequences by two amino acids or fewer.

This was accomplished by utilizing a “jumping dimer” or “jumping trimer” approach to light chain oligonucleotide design. The jumping dimer approach involves the incorporation of degenerate codons at each of the six positions of CDRL3 encoded by the VL segment (L3-VL). At most two positions vary from germline in each individual L3-VL sequence, but the two positions do not have to be adjacent to one another. Thus, the total number of designed degenerate oligonucleotides synthesized per VL chassis is $6!/(4!2!)$, or fifteen (accounting for six of the most commonly occurring amino acids at the junction (position 96) between VL and JL for each kappa germline chassis (namely F, L, I, R, W, Y, and P; see U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379, each of which is incorporated by reference in its entirety, for more details on the junctional amino acids at position 96). The jumping trimer approach is analogous to the jumping dimer approach, but with three positions varying from germline in each individual L3-VL sequence, instead of two as in the jumping dimer. The degenerate codons selected for each position in the jumping dimer and trimer approaches were chosen to (1) to reproduce the diversity contained in the known repertoire of publicly available human VK sequences (see Appendix A of WO/2009/036379); and (2) to minimize or eliminate undesirable sequences within the CDRL3s of the resulting synthetic light chains, such as N-linked glycosylation motifs (NXS/NXT), Cys residues, stop codons, and deamidation-prone NG motifs. Table 4 shows the fifteen degenerate oligonucleotides encoding the VK1-39 CDRL3 sequences with a length of nine amino acids and F or Y as the junctional amino acid, and the corresponding degenerate

polypeptide sequences. Table 5, Table 6, and Table 7 provide the oligonucleotide sequences for each of the VK sequences of the exemplary jumping dimer and trimer libraries, for CDRL3 lengths of 8, 9, and 10, respectively, and the sequences for the corresponding CDRL3s.

The number of unique CDRL3 sequences within each germline library was then enumerated and compared to the number of unique CDRL3 sequences in a different light chain library, designated “VK-v1.0” (see Example 6.2 in US Publication No. 2009/0181855), for each of the three lengths. Table 8 provides the number of unique CDRL3 sequences in each of the respective germline libraries.

Figure 3 provides the percentage of sequences in the jumping dimer and VK-v1.0 libraries with CDRL3 length of nine amino acids that contain no mutations from germline-like sequences (Table 1) or 1, 2, 3, or 4 or fewer mutations from germline-like sequences. Naturally-occurring VK1-05 sequences are almost as likely to have Ser (germline amino acid type) as Pro at Kabat position 95, thus both residues (S and P) were incorporated in the synthetic libraries representing VK1-05 repertoires. However, as indicated in Table 1, only Ser was considered to be a germline-like residue at position 95 for the purposes of this analysis when the VK gene is VK1-05. The plot for VK3-20 is representative of the remaining chassis in the library for a length nine. All of the sequences in the VK1-05 library were within three amino acids of human germline sequences, and approximately 63% of the sequences were within two amino acids of human germline-like sequences. For the rest of the libraries, and as designed, 100% of the sequences were within two amino acids of human germline-like sequences; thus, over 95% of the sequences of length 9 in the jumping dimer library considered as a whole were within 2 amino acids of germline-like sequences. By comparison, only 16% of the members of the VK-v1.0 libraries of length nine amino acids are within two amino acids of the corresponding human germline-like sequences. For length 8, about 98% of the sequences in the jumping dimer libraries were within two amino acids of germline-like, versus about 19% for VK-v1.0. For length 10, more than 95% of the sequences of the jumping dimer library were within two amino acids of germline-like, versus about 8% for VK-v1.0.

In some embodiments, to concentrate the diversity in positions most likely to be solvent-exposed in the folded antibody, positions 89 and 90 (Kabat numbering) are not modified from germline – these are most often QQ, but the sequence is MQ for the VK2-28 chassis. Other VK germline genes have different sequences at positions 88-89, and the use of these genes as chassis also falls within the scope of the invention. For example, VK1-27 has QK, VK1-17 and VK1-6 both have LQ, and so on. The sequences in these positions are known in the art and can be obtained, for example, from Scaviner *et al.*, Exp. Clin. Immunogenet., 1999, 16: 234 (see Figure 2), which is incorporated by reference in its entirety.

CDRH3 LIBRARIES

The following examples describe methods and compositions useful for the design and synthesis of antibody libraries with improved CDRH3 sequences in comparison to libraries known in the art. The CDRH3 sequences of the invention have enhanced diversity in comparison to libraries known in the art, while retaining the character of human sequences, improving combinatorial efficiency of the synthetic CDRH3 segments, and/or improving the matching between synthetic CDRH3 sequences and human CDRH3 sequences in one or more reference sets.

Example 3. Generating a Curated Reference Set of Human Preimmune CDRH3 Sequences

A file containing approximately 84,000 human and mouse heavy chain DNA sequences was downloaded from the BLAST public resource (<ftp.ncbi.nih.gov/blast/db/FASTA/>; filename: igSeqNt.gz; download date: Aug 29, 2008). Of these approximately 84,000 sequences, approximately 34,000 sequences were identified as human heavy chain sequences based on analysis of the sequence header annotation. These sequences were then filtered as follows: First, all sequences were classified, via their VH-region, according to their corresponding (closest matched) VH germline. Sequences that were of an incorrect or insufficient length, or that could not be matched due to extensive mutation, were discarded. Second, any sequences containing

more than five mutations, at the DNA level, when compared to their corresponding germline VH sequence were also discarded. It was assumed, consistent with Rada and Milstein, EMBO J., 2001, 20: 4570, that mutations (or lack thereof) in the N-terminal portion of the variable region may be used as conservative surrogates for mutations (or lack thereof) in the C-terminal portion of the variable region, in particular in CDRH3. Therefore, selecting only sequences with five or fewer nucleotide mutations in VH, which is N-terminal to CDRH3 is highly likely to also select for CDRH3 sequences that are either lightly mutated or not mutated at all (*i.e.*, having preimmune character).

After translating the remaining DNA sequences into their amino acid counterparts, the appropriate reading frame containing the heavy chain germline amino acid sequence was identified and used to identify the sequences of the CDRs, including that of the CDRH3. The list of CDRH3 sequences obtained at this point was further filtered to eliminate members that did not differ from any other sequence in the set by at least three amino acids (after matching for length). This process yielded 11,411 CDRH3 sequences, with 3,571 sequences annotated as originating from healthy adults (“Healthy Preimmune Set” or “HPS”; *see* Appendix A for GI Nos.) and the other 7,840 sequences annotated as originating from individuals suffering from disease, of fetal origin, or of antigen-specific origin. The methods described below were then used to deconvolute each of the sequences in the HPS into the four segments that constitute the CDRH3: (1) TN1, (2) DH, (3) N2, and (4) H3-JH.

Example 4. Method to Match Segments from a Theoretical Segment Pool to CDRH3s in a Reference Set

This example describes the method used to identify the TN1, DH, N2, and H3-JH segments of the CDRH3s in the HPS. The currently exemplified approach to the design and synthesis of human CDRH3 sequences mimics the segmental V-D-J gene recombination processes by which the human immune system generates the preimmune CDRH3 repertoire. The matching method described here determines which TN1, DH, N2 and H3-JH segments have been used to produce a particular CDRH3 across a reference set of CDRH3s (*e.g.*, the HPS). This information is then used, optionally in conjunction with other information described below (*e.g.*, physicochemical properties), to determine

which TN1, DH, N2, and H3-JH segments from a theoretical segment pool (or segments extracted from the CDRH3 sequences in the reference set, in the case of the TN1 and N2) should be included in a synthetic CDRH3 library.

The inputs to the matching method are: (1) a reference set of CDRH3 sequences (*e.g.*, the human CDRH3 sequences in the HPS), and (2) a theoretical segment pool, containing a plurality of TN1, DH, N2 and/or H3-JH segments. Methods by which the members of the theoretical segment pool are generated are more fully described below. For each CDRH3 in the reference set, the matching method generates two outputs: (i) a list of the closest matched CDRH3 sequences that can be generated using the segments of the theoretical segment pool, and (ii) the one or more segment combinations from the theoretical segment pool that can be used to create these closest matched CDRH3 sequences.

The matching method was performed as follows: Each TN1 segment in the theoretical segment pool was aligned at its first amino acid with the first amino acid (position 95) of the CDRH3 sequence from the reference set. For each segment length, all (*i.e.*, one or more) of the segments returning the best matches are retained, and the remaining segments are discarded. The retained TN1 segments are then concatenated with all DH segments from the theoretical segment pool, to create [TN1]-[DH] segments. These segments are then aligned as above, and all the best matches for each of the [TN1]-[DH] segments are retained. The procedure is repeated with [TN1]-[DH]-[N2] and [TN1]-[DH]-[N2]-[H3-JH] segments until the length of the CDRH3 sequence from the reference set is identically recapitulated by the segment combinations from the theoretical segment pool. All segment combinations returning the best match to the CDRH3s in the reference set are retained as the output of the matching method.

Table 9 provides an example of the output of the matching method, specifically the output for four individual sequences from the HPS, using a theoretical segment pool designated “Theoretical Segment Pool 1,” or “TSP1”. TSP1 contains several theoretical segment pools, namely: 212 TN1 segments (Table 10), 1,111 DH segments (Table 11), 141 N2 segments (Table 12), and 285 H3-JH segments (Table 13). The CDRH3 sequence in Test Case 1 contains an identical match in TSP1 that is reached via a unique combination of the four segments. Test Cases 2.1 and 2.2 each return an identical match,

but via two distinct combinations that differ in the TN1 and DH segments. In Test Cases 3.1, 4.1, and 4.2, the closest matches are all a single amino acid away from the reference CDRH3, and can be reached via one (3.1) or two (4.1 and 4.2) combinations of segments from TSP1. This approach can be generalized to find all of the closest matches to any reference CDRH3 sequence within any theoretical segment pool and all combinations of the segments within the theoretical segment pool that can produce the reference CDRH3 sequence exactly and/or its closest matches.

Example 5. Deriving Theoretical Segment Pools of H3-JH Segments

In order to produce theoretical segment pools of H3-JH segments for consideration for inclusion in a synthetic CDRH3 library, the following method was applied to generate mutants of seven (IGHJ1-01, IGHJ2-01, IGHJ3-02, IGHJ4-02, IGHJ5-02, IGHJ6-02 and IGHJ6-03) of the twelve germline IGHJ sequences of Table 14. These seven alleles were chosen because they were among the most commonly occurring alleles in human sequences. Libraries where all sequences of Table 14 (some differing only in FRM4) are used to generate H3-JH and/or JH (*i.e.*, H3-JH and FRM4) also fall within the scope of the invention. The method is intended to simulate the creation of junctional diversity during the V-D-J recombination process *in vivo*, which occurs via enzyme-mediated addition and deletion of nucleotides to the germline gene segments. The method proceeds as follows, and results in a fully enumerated theoretical segment pool of H3-JH segments:

1. A pre-treatment was applied to the IGHJ genes that contain a partial codon consisting of two nucleotide bases at their 5' terminus (IGHJ3-02, IGHJ4-02, IGHJ5-02, IGHJ6-02 and IGHJ6-03), prior to the first nucleotide encoding the translation of the JH segment that produces the well-known JH framework regions. For example, the IGHJ3-02 gene contains an AT dinucleotide sequence prior to the first nucleotide encoding the translation of the JH segment that produces the JH framework region (Figure 4, top). All partial codons consisting of two nucleotide bases were completed, using all possible nucleotide doublets (*i.e.*, NN) at their two most 5' positions (Figure 4, top, second row for IGHJ3-02).

More specifically, the most 5' nucleotide in the germline sequence was mutated to N and an additional N was added 5' to that nucleotide.

2. IGHJ genes IGHJ1-01 (Figure 4, center) and IGHJ2-01 (Figure 4, bottom) contain zero and one nucleotide base(s) at their 5' termini, prior to the first nucleotide encoding the translation of the JH segment that produces the JH framework region. For these genes, the pre-treatment described in step 1 was not performed. Instead, the 5' doublets were mutated to NN (Figure 4, middle and bottom, second row of each). Therefore, after performing this step, each of the seven IGHJ genes enumerated above was converted to a variant with an NN doublet as its first two 5' positions.
3. The 5' codons of the sequences produced via steps 1 and 2 were then deleted, and the first two bases of the resulting DNA sequence were subsequently mutated to an NN doublet (Figure 4, rows 3-4 for all).
4. The 5' codons of the sequences produced in step 3 were then deleted, and the first two bases of the resulting DNA sequence were subsequently mutated to an NN doublet (Figure 4, rows 5-6 for all).
5. Each of the polynucleotide sequences generated by steps (1) – (4) were then translated, to obtain a theoretical segment pool consisting of 248 parent H3-JH polypeptide segments (Table 15) from the reading frame for each sequence that produced the JH framework region.
6. The parent H3-JH polypeptide segments were truncated at their N-termini, by removing one amino acid at a time until only the portion of the JH segment comprising FW4 remains (*i.e.*, an H3-JH segment with a length of zero amino acids).

The methods described above resulted in the production of a theoretical segment pool of 285 H3-JH segments (Table 13).

Example 6. Deriving Theoretical Segment Pools of DH Segments

Two theoretical pools of DH segments were generated, using one or more of two translation methods, designated “Translation Method 0” (TM0), or “Translation Method

1” (“TM1”), each performed in the three forward reading frames of 27 human germline IGHD DNA sequences or segments derived therefrom (Table 16).

The 1K DH Theoretical Segment Pool (1K DH)

TM1 was used to generate the “1K DH Theoretical Segment Pool” (“1K DH”; see the 1,111 DH segments of Table 11). In TM1, IGHD sequences that had a partial codon containing two untranslated bases after translation in any of the three forward reading frames were completed to produce a full codon only if the two bases could encode only a single amino acid upon completion. For example, a DNA sequence such as TTA-GCT-CG has two full codons that would be translated to LA, and a remaining partial codon (CG) that can only encode R, as any of CGA, CGC, CGG, or CGT will encode R. Thus, applying TM1 to this sequence will yield LAR. For sequences with partial codons that could encode more than one amino acid (*e.g.*, GA or AG), the partial codons were ignored. Applying TM1 to the 27 IGHD sequences of Table 16 generated a theoretical segment pool containing the 73 DH parent segments of Table 17 (some containing stop codons (“Z”) and unpaired Cys residues). These sequences were then progressively deleted at the amino acid level, at their N- and C-termini, until only two amino acids remained. Truncated segments were discarded if they contained a stop codon, unpaired Cys residues, N-linked glycosylation motifs, or deamidation motifs. This process yielded the 1,111 DH segments of Table 11.

The 68K DH Theoretical Segment Pool (68K DH)

The 27 IGHD genes and alleles of Table 16 were progressively deleted on either or both of their 5’ and 3’ ends until four bases remained, yielding 5,076 unique polynucleotide sequences of four or more nucleotides. These 5,076 sequences were subjected to systematic addition of 0, 1 and/or 2 N nucleotides their 5’ and/or 3’ ends. The resulting sequences were translated using TM0. In TM0, only full codons are translated; partial codons (*i.e.*, 1 or 2 bases) are ignored. This method yielded 68,374 unique DH polypeptide segments after elimination of segments with stop codons, unpaired Cys residues, Asn in the last or next to last position that can lead to N-linked glycosylation motifs, and deamidation motifs (the “68K DH Theoretical Segment Pool”).

Using the IGHD genes of Table 16 as an input for the PYTHON computer code provided below will reproduce the exact theoretical segment pool of 68,374 DH segments. There are two free parameters in this program: (1) the minimum length of the DNA sequences remaining after progressive deletions (4 bases in this example), and (2) the minimum length of the peptide sequences (2 amino acids in this example) acceptable for inclusion in the theoretical segment pool. These parameters can be changed to alter the output of the program. For example, changing the first parameter to one base and the second parameter to one amino acid would lead to a larger theoretical segment pool with 68,396 unique sequences, including 18 single-amino acid segments. DH segments progressively truncated to different lengths also fall within the scope of the invention; for example those truncated to 1, 2, 3, or 4 or more amino acids, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides prior to translation.

PYTHON Computer Program to Generate 68,374 DH Segments

```
import math, sys, string

class genes:
    name = 'x'
    seq = 'x'
    progdel = set()
class table:
    name = 'x'
    dna = 'x'
    dna_n = 20 * ['x']
    prot = 60 * ['x']
    uprot = set()

pepcod = { 'A':0, 'C':1, 'D':2, 'E': 3, 'F': 4, 'G': 5, 'H': 6,
'I':7, 'K':8, 'L': 9, 'M': 10,
'N':11, 'P':12, 'Q':13, 'R': 14, 'S': 15, 'T': 16, 'V':17, 'W':18,
'Y': 19, 'Z': 20}
codpep = 21 * ['']
codpep[0] = 'A'
codpep[1] = 'C'
codpep[2] = 'D'
codpep[3] = 'E'
codpep[4] = 'F'
codpep[5] = 'G'
codpep[6] = 'H'
codpep[7] = 'I'
```

```

codpep[8] = 'K'
codpep[9] = 'L'
codpep[10] = 'M'
codpep[11] = 'N'
codpep[12] = 'P'
codpep[13] = 'Q'
codpep[14] = 'R'
codpep[15] = 'S'
codpep[16] = 'T'
codpep[17] = 'V'
codpep[18] = 'W'
codpep[19] = 'Y'
# Z represents a stop codon
codpep[20] = 'Z'

bases = 'ACGT'

def translate_dna(sequence):
# Translation of input DNA sequence using standard genetic code
# Only full codons are considered with any remaining 1 or 2 bp
being ignored
# Z represents a stop codon
    code = {
        'ATA':'I', 'ATC':'I', 'ATT':'I', 'ATG':'M',
        'ACA':'T', 'ACC':'T', 'ACG':'T', 'ACT':'T',
        'AAC':'N', 'AAT':'N', 'AAA':'K', 'AAG':'K',
        'AGC':'S', 'AGT':'S', 'AGA':'R', 'AGG':'R',
        'CTA':'L', 'CTC':'L', 'CTG':'L', 'CTT':'L',
        'CCA':'P', 'CCC':'P', 'CCG':'P', 'CCT':'P',
        'CAC':'H', 'CAT':'H', 'CAA':'Q', 'CAG':'Q',
        'CGA':'R', 'CGC':'R', 'CGG':'R', 'CGT':'R',
        'GTA':'V', 'GTC':'V', 'GTG':'V', 'GTT':'V',
        'GCA':'A', 'GCC':'A', 'GCG':'A', 'GCT':'A',
        'GAC':'D', 'GAT':'D', 'GAA':'E', 'GAG':'E',
        'GGA':'G', 'GGC':'G', 'GGG':'G', 'GGT':'G',
        'TCA':'S', 'TCC':'S', 'TCG':'S', 'TCT':'S',
        'TTC':'F', 'TTT':'F', 'TTA':'L', 'TTG':'L',
        'TAC':'Y', 'TAT':'Y', 'TAA':'Z', 'TAG':'Z',
        'TGC':'C', 'TGT':'C', 'TGA':'Z', 'TGG':'W',
    }
    proteinseq = ''
    for n in range(0,len(sequence),3):
        if code.has_key(sequence[n:n+3]) == True:
            proteinseq += code[sequence[n:n+3]]
    return proteinseq

# main body starts here

```

```

# open input and output files
in1 = open(sys.argv[1], 'r')
out1 = open(sys.argv[2], 'w')

# read DNA sequences for input DH segments
data = in1.readlines()
nseg = len(data)

seqs = [ genes() for i in range(nseg) ]
for i in range(nseg):
    line = data[i]
    words = string.split(line)
    seqs[i].name = words[0]
    seqs[i].seq = words[1]
    seqs[i].progdel = set()

# Define here minimum length for DNA (4) and for protein (2)
minlen = 4
minp = 2

# Implement progressive base by base deletion from 5' or 3' or
both ends
alln = 0
for i in range(nseg):
    seq = seqs[i].seq
    lseq = len(seq)
    nt = ct = lseq
    for n in range(nt):
        for c in range(ct):
            nseq = seq[n:lseq-c]
            if (len(nseq) >= minlen):
                seqs[i].progdel.add(nseq)
    alln += len(seqs[i].progdel)

# Collect unique DNA sequences across all DH genes of origin and
ignore redundant ones
progdel = [ table() for i in range(alln) ]
n = 0
for i in range(nseg):
    for kk in seqs[i].progdel:
        unix = 1
        for j in range(n):
            if (kk == progdel[j].dna):
                unix = 0
                break
        if (unix == 1):
            progdel[n].name = seqs[i].name

```

```

        progdel[n].dna = kk
        n +=1

# Add none, 1 or 2 bp on one or both ends
# extras 20 + 20 * (21) = 20 + 420 = 440
# allocate memory for all variants
for i in range(n):
    progdel[i].dna_n = 440 * ['x']
    progdel[i].prot = 441 *3 * ['x']
    progdel[i].uprot = set()

# add 1 or 2 bases at each end of input segment
tot = 0
for i in range(n):
# Step over each unique DNA sequence
    k = 0
# One base on 5' end combined with 1 or 2 bases added to 3' end
    for l5 in range(4):
        progdel[i].dna_n[k] = bases[l5] + progdel[i].dna
        k +=1
        for l3 in range(4):
            progdel[i].dna_n[k] = bases[l5] + progdel[i].dna + bases[l3]
            k +=1
            for l3 in range(4):
                for m3 in range(4):
                    progdel[i].dna_n[k] = bases[l5] + progdel[i].dna +
bases[l3] + bases[m3]
                    k +=1
# One or two bases added to 3' only in this part
        for l3 in range(4):
            progdel[i].dna_n[k] = progdel[i].dna + bases[l3]
            k +=1
        for l3 in range(4):
            for m3 in range(4):
                progdel[i].dna_n[k] = progdel[i].dna + bases[l3] + bases[m3]
                k +=1
# Two bases on 5' end combined with 1 or 2 bp on 3' end
        for l5 in range(4):
            for m5 in range(4):
                progdel[i].dna_n[k] = bases[l5] + bases[m5] + progdel[i].dna
                k +=1
                for l3 in range(4):
                    progdel[i].dna_n[k] = bases[l5] + bases[m5] +
progdel[i].dna + bases[l3]
                    k +=1
                for l3 in range(4):
                    for m3 in range(4):

```

```

        progdel[i].dna_n[k] = bases[l5] + bases[m5] +
progdel[i].dna + bases[l3] + bases[m3]
        k +=1
# Now translate in all 3 forward reading frames
# Save unique peptide sequences
for fr in range(3):
    piece = progdel[i].dna
    piece = piece[fr:]
    tpiece = translate_dna(piece)
    progdel[i].prot[fr] = tpiece
    progdel[i].uprot.add(tpiece)

for k in range(440):
    piece = progdel[i].dna_n[k]
    piece = piece[fr:]
    tpiece = translate_dna(piece)
    progdel[i].uprot.add(tpiece)
    progdel[i].prot[3+440*fr +k] = tpiece
tot += len(progdel[i].uprot)

# Collect unique sequences with no ASN at last or next to last
position, no unpaired or consecutive CYS, no stops
unset = set()
segm = [ genes() for i in range(tot) ]
lux = 0
nn = 0
for i in range(n):
    k = 0
    for kk in progdel[i].uprot:
# Filter out sequences with undesired features, including length
being too short (under "minp" defined above)
        if (len(kk) < minp): continue
        if (kk[len(kk)-1] == "N" or kk[len(kk)-2] == "N"):
continue
        if (kk.count("Z") > 0 or kk.count("CC") >0 or
kk.count("C") % 2 >0): continue
        unset.add(kk)
        lux1 = len(unset)
        if (lux1 > lux):
            segm[nn].name = progdel[i].name + "_" +
str(nn)

            segm[nn].seq = kk
            nn += 1
        lux = lux1
    k +=1

# Print out unique peptide sequences that pass all the filters
for i in range(nn):

```

```
ou1.write("%s\t%s\n" % (segm[i].name, segm[i].seq))
```

Example 7. Deriving Theoretical Segment Pools of TN1 and N2 Segments

The libraries of this example are designed to, in some instances, have a greater diversity in their TN1 and N2 segments in comparison to other libraries known in the art. The diversity of the TN1 and N2 segments was increased by using the matching method described in Example 4 to deconvolute the CDRH3 sequences in the HPS into their constituent segments (*i.e.*, TN1, DH, N2, and H3-JH), followed by extraction of “novel” TN1 and N2 segments in the manner described below. For the purposes of the invention, “novel” TN1 and N2 segments are TN1 and N2 segments that do not appear in a theoretical segment pool that is matched to a reference set of CDRH3 sequences. Following is an example of the method used to extract novel TN1 and N2 segments from the HPS. This method can be generalized to extract novel TN1 and N2 segments from any reference set of CDRH3 sequences, using any theoretical segment pool containing TN1, DH, N2, and/or H3-JH segments.

Table 9 provides the matching results for the reference CDRH3 sequence ERTINWGWGVYAFDI (Test Cases 5.1-5.4) from the HPS, using Theoretical Segment Pool 1 (“TSP1”). The best matches to the reference CDRH3 are four CDRH3 sequences, each within three amino acids of the reference CDRH3 sequence. In each of these matches, the TN1, DH, N2 and H3-JH segments are of length 4, 3, 3 and 5 amino acids, respectively. Thus the reference CDRH3 can be deconvoluted into the following segments: ERTI-NWG-WGW-YAFDI (*i.e.* [TN1]-[DH]-[N2]-[H3-JH], respectively). The DH and H3-JH segments from the reference CDRH3, NWG and YAFDI respectively, are identically present in TSP1. However, the TN1 (ERTI) and N2 (WGW) segments from the reference CDRH3 are absent in TSP1 and match TSP1 segments with one or more amino acid mismatches. These “novel” TN1 and N2 segments are extracted from the reference CDRH3 and considered for inclusion prospective theoretical segment pools and/or synthetic libraries. Additional novel TN1 and N2 segments were accumulated by applying this analysis to all members of the HPS. In order to robustly identify TN1 and N2 sequences, the extraction was performed only for those CDRH3

sequences in which the DH and H3-JH segments in the reference CDRH3 and TSP1 cumulatively return no more than 3 amino acid mismatches, implying that the DH and H3-JH segments of the reference CDRH3 had been reliably assigned.

Example 8. Calculation of Segment Usage Weights

Segment usage weights were calculated for their utility in determining which segments from the theoretical segment pools (*e.g.*, TSP1 and TSP1 plus novel TN1 and N2 segments identified as described in Example 7) should be included in a synthetic library. Segment usage weights were obtained by utilization of the matching method described above and Equation 2:

$$\text{Equation 2} \quad w(i) = \frac{1}{S_m} \sum_{j=1}^{S_m} \frac{1}{g(j)} \sum_{k=1}^{g(j)} f_i(k)$$

where,

- $w(i)$: Weight for segment i . $0 \leq w(i) \leq 1$.
- S_m : Number of sequences (out of total S in the reference CDRH3 set) which contain one or more best matches with no more than m amino acid mismatches in the specified region of the reference CDRH3 sequence. Here, the mismatches are computed over the Kabat-CDRH3 region, but other fragments of the CDRH3 sequences may also be considered. A constant value of $m=3$ was used here, but other values may be used, or the value may depend on the length of the reference CDRH3 sequence.
- $g(j)$: Total number of degenerate segment combinations producing the best match to the reference CDRH3 sequence j .
- $f_i(k)$: Fractional amino acid identity of TN1, DH, N2 or H3-JH segment in degenerate match k , relative to the corresponding sequence fragment in the reference CDRH3 sequence j . The fractional amino acid identity equals zero if the segment does not appear in match k . Other definitions of f , such as amino acid similarity (*e.g.*, based on physicochemical properties of the amino acids such as hydrophobicity), instead of identity, may be also used.

The procedure for calculating segment usage weights will be further exemplified below. In each of these examples, the best match combinations from TSP1 are provided for a single CDRH3 sequence ($S_m = 1$) and the degeneracy (k) and fractional mismatch (f) dependent weight calculations are explained.

Example 8.1. Segment usage weights for Test Case 1 in Table 9

Refer to Test Case 1 in Table 9. The CDRH3 sequence RTAHHFDY is identically located in TSP1 ($f = 1$, subscripts dropped for simplicity) via a unique segment combination ($g = 1$). Table 18 provides the usage weights for the segments corresponding to the best match from TSP1 for the CDRH3 of Test Case 1.

Example 8.2. Segment usage weights for Test Cases 2.1 and 2.2 in Table 9

Refer to Test Cases 2.1 and 2.2 in Table 9. The CDRH3 sequence VGIVGAASY may be identically located in TSP1 ($f = 1$) via two distinct segment combinations ($g = 2$). Table 19 provides the usage weights for the segments corresponding to the best match from TSP1 for the CDRH3 of Test Cases 2.1 and 2.2.

Example 8.3. Segment usage weights for Test Case 3.1 in Table 9

Refer to Test Case 3.1 in Table 9. The CDRH3 sequence DRYSGHDLGY may be identically located in TSP1 via a unique segment combination ($g = 1$) with a single amino acid difference. As provided below, the TN1, N2 and H3-JH segments match the corresponding reference sequence fragments identically, while four of the five DH amino acids match identically.

Sequence from the HPS: DR-YSG~~H~~D-LG-Y

Nearest Neighbor in TSP1: DR-YSG~~Y~~D-LG-Y

Thus, here

$$f = 4/5 \text{ for the DH segment; and} \\ = 1 \text{ for the TN1, N2, and H3-JH segments (Table 20).}$$

Example 8.4. Matching of Test Cases 4.1 and 4.2 in Table 9.

Refer to test cases 4.1 and 4.2 in Table 9. The CDRH3 sequence GIAAADS~~N~~WLDP may be located in TSP1 via two distinct segment combinations ($g = 2$), each with a single amino acid difference. As provided below, the TN1, DH and N2 segments match the corresponding reference sequence fragments identically, while five of the six H3-JH amino acids match identically.

Sequence from HPS: (-) - GIAAA - D - SNWLDP

Nearest Neighbor in TSP1: (-) - GIAAA - D - SNW**F**DP

Sequence from HPS: G - IAAA - D - SNWLDP

Nearest Neighbor in TSP1: G - IAAA - D - SNW**F**DP

Here, (-) represents the “empty” TN1 segment.

Applying Equation 2 results in the segment usage weights provided in Table 21.

Example 8.5. Calculating the Segment usage weights for Test Cases 1 to 4.2 of Table 9

Extending the individual calculations described above to simultaneously include all of Test Cases 1 to 4.2 of Table 9 results in the segment usage weights of Table 22.

Example 8.6. Calculating the Segment usage weights for Test Cases 5.1 to 5.4 of Table 9

Refer to the CDRH3 sequence ERTINWGWGVYAFDI and the novel TN1 and N2 segments extracted from the CDRH3 sequence, in Example 7. In this case, the novel TN1 and N2 segments (ERTI and WGV respectively), and the DH and H3-JH segments from TSP1 (NWG and YAFDI respectively) are each assigned usage weights of unity.

Example 9. Selection of TN1, DH, N2 and JH Segments for Inclusion in Synthetic Libraries

Figure 5 provides the general method used for the design of synthetic CDRH3 libraries. The method uses as input: (1) a theoretical segment pool containing TN1, DH, N2, and H3-JH segments (*e.g.*, TSP plus novel TN1 and N2 segments); and (2) a collection of reference CDRH3 sequences (*e.g.*, the HPS). From these inputs, a particular subset of segments from the theoretical segment pool is selected for inclusion in a physical CDRH3 library.

First, the best matches to the CDRH3s of the HPS were obtained, from within the TSP1 set, with or without the novel TN1 and N2 segments, using the matching method described above. This data was then used to compute the segment usage weights via Equation 2. Segments were prioritized for inclusion in the physical library based on their relative frequency of occurrence in the CDRH3 sequences of the HPS (as indicated by the segment usage weights), as well as other factors (more fully described below), such as hydrophobicity, alpha-helical propensity, and expressibility in yeast.

Example 9.1. Exemplary Library Design (ELD-1)

ELD-1 uses the HPS and the segments from TSP1 1 (9.5×10^9 members) as inputs and produces an output of 100 TN1, 200 DH, 141 N2 and 100 H3-JH segments, each from TSP1, ranked in order by their usage weights in the HPS, to produce a library with theoretical complexity of 2.82×10^8 . The segments corresponding to ELD-1 are provided in Table 23. Note that here the combination of all of the segments (*i.e.*, TN1, DH, N2, and H3-JH), and the individual sets of segments (*i.e.*, TN1 only, DH only, N2 only, and H3-JH only) each constitute theoretical segment pools.

Example 9.2. Exemplary Library Design 2 (ELD-2)

The inputs for this design are the HPS and the segments from TSP1 plus the novel TN1 and N2 segments extracted from the HPS (Example 7). The outputs are (1) 200 DH and 100 H3-JH segments, each from TSP1; and (2) 100 TN1 and 200 N2 segments including TN1 and N2 segments originally in TSP1 and those extracted from the sequences in the HPS. Applying the method described in Example 7 to extract novel TN1

and N2 segments (*i.e.*, those not included in TSP1) resulted in the identification of 1,710 novel TN1 segments and 1,024 novel N2 segments. The segments corresponding to ELD-2 are provided in Table 24. Note that here the combination of all of the segments (*i.e.*, TN1, DH, N2, and H3-JH), and the individual sets of segments (*i.e.*, TN1 only, DH only, N2 only, and H3-JH only) each constitute theoretical segment pools. As in ELD-1, all segments in ELD-2 were selected for inclusion based solely on their usage weights in the HPS.

Example 9.3. Exemplary Library Design 3 (ELD-3)

The inputs for this design are identical to those for ELD-2. As in ELD-2, the outputs are (1) a set of 200 DH and 100 H3-JH segments,, each from TSP1; and (2) a set of 100 TN1 and 200 N2 segments, including TN1 and N2 segments originally in TSP1 and those extracted from the sequences in the HPS (Example 7). However, the approach used for the selection of the segments for ELD-3 differs in two respects. First, selected physicochemical properties of the segments (hydrophobicity, isoelectric point, and alpha-helix propensity) were used, in addition to the segment usage weights, to prioritize segments for inclusion in the physical library. Hydrophobicity was used to de-prioritize hydrophobic DH segments that are empirically over-represented in poorly expressed antibodies isolated from yeast-based libraries. Isoelectric point and propensity for alpha-helix formation were utilized to identify segments located in regions of physicochemical property space that were relatively unexplored in CDRH3 libraries known in the art (*e.g.*, U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379). Second, the segment usage weights were calculated via a bootstrap analysis of the HPS dataset. These methods are more fully described below. The segments corresponding to ELD-3 are provided in Table 25. Note that here the combination of all of the segments (*i.e.*, TN1, DH, N2, and H3-JH), and the individual sets of segments (*i.e.*, TN1 only, DH only, N2 only, and H3-JH only) each constitute theoretical segment pools.

Example 9.3.1. Generation of Segment Usage Weights via Bootstrap Analysis

Bootstrap analysis (Efron & Tibshirani, An Introduction to the Bootstrap, 1994 Chapman Hill, New York) is a widely used statistical procedure for estimating the

variability of a statistic of a given sample. This estimate is based on the value of the statistic calculated for several sub-samples, equal in size to the original sample and derived from it via sampling with replacement. Members of the original sample are chosen at random to form the sub-samples, and are typically included multiple times in each sub-sample (hence, “sampling with replacement”).

Here, the original sample is the HPS dataset with $n=3,571$ members and the statistic is the segment usage weight. One-thousand sub-samples, each with 3,571 members, were generated by randomly choosing sequences from the HPS dataset (no more than 10 repeats of a given sequence were allowed in each sub-sample). The matching method described above was then applied to each sub-sample, and the final segment usage weights were calculated as the average of the values obtained for the individual sub-samples. Average values derived via this bootstrap procedure are more robust than values calculated from the parent HPS dataset alone. Unless indicated otherwise, these average values of the 1,000 sub-samples were used in the selection of segments for ELD-3.

Example 9.3.2. Amino Acid Property Indices

The *AAindex* database, available online at www.genome.jp/aaindex/, provides more than 500 numerical indices representing various physicochemical and biochemical properties of amino acids and pairs of amino acids. These properties include hydrophobicity, electrostatic behavior, secondary structure propensities and other characteristics, with several indices often available for a given property. The following three indices were chosen by starting with the well-understood Kyte-Doolittle hydropathy index (KYJT820101) and adding the indices most numerically de-correlated from it and each other. They thus potentially describe non-overlapping regions of amino acid property space and were used for analysis and selection of the DH and H3-JH segments for ELD-3:

1. KYTJ820101 (hydropathy index)
2. LEVM780101(normalized frequency of alpha helix)
3. ZIMJ680104 (isoelectric point)

Example 9.3.3. Hydrophobic DH Segments are Over-Represented in Poorly Expressed Antibodies Isolated from Yeast-Based Libraries

Based on protein expression levels from approximately 1200 antibodies expressed in *S. cerevisiae*, antibodies were classified as either “Good” or “Poor” expressors. The CDRH3 sequence of each antibody in each category was examined to identify sequence features that correlated with the expression level. One such sequence feature is the hydrophobicity of the DH segments calculated using the KYTJ820101 index. Figure 6 provides the frequency of “Good” and “Poor” expressors as a function of the DH segment hydrophobicity (increasing to the right). The distribution expected from the synthetic library used to isolate these antibodies is also provided as a reference (“Design”). DH segments with the highest hydrophobicity values (far right of the plot) are over-represented (relative to the expectation based on the design) among “Poor” expressors and under-represented among “Good” expressors. Similarly, hydrophilic DH segments (far left) are over-represented among “Good” expressors and under-represented among “Poor” expressors. From this data, it was inferred that the overall expressibility of the antibodies of the library may be improved by synthesizing CDRH3 sequences with fewer hydrophobic DH segments.

Example 9.3.4. Selection of the 200 DH Segments for Inclusion in ELD-3

A set of 71 DH segments from TSP1 were designated as “core” DH segments for automatic inclusion in ELD-3. These segments had the following desirable properties:

1. Fifty-three of seventy-one were present within the top 7% of DH segments rank-ordered by segment usage weights from the bootstrap analysis.
2. Eighteen of seventy-one were present within the top 7% of DH segments rank-ordered by usage weights derived from antibodies isolated from libraries expressed in *S. cerevisiae*.

The remaining 1,040 segments were designated as “non-core.” To complete the set of 200 segments in ELD-3, 129 segments were chosen from the “non-core” pool of segments in the following manner:

1. Sixty-five segments were eliminated because they contain either (a) an Asn residue at the last or next-to-last position with the potential to form N-linked glycosylation motifs via combination with N2 amino acids or (b) the amino acid sequence NG, implicated in de-amidation.
2. Segments with higher than median values for the KYTJ820101 hydropathy index (median=2.9 for 1K DH) were eliminated from further consideration. In view of the known importance of Tyr for antigen recognition (Fellouse *et al.*, PNAS, 2004, 101: 12467; and Hofstadter *et al.*, J. Mol. Biol., 1999, 285: 805, each incorporated by reference in its entirety), segments containing at least one Tyr residue were retained unless located in the highest hydrophobicity quartile (KYTJ820101 value higher than 9.4). This eliminated 443 segments.
3. The final set of 129 segments was obtained by using an objective function that aimed to maximize the Euclidean distance, between the “core” and the remaining 443 “non-core” segments, in a multi-dimensional space defined by the following variables: (1) amino acid mismatches to nearest neighbor; and (2) values of the three physicochemical property indices.

Example 9.3.5. Selection of the 100 H3-JH Segments for Inclusion in ELD-3

One-hundred H3-JH segments were chosen for inclusion in ELD-3 in the following manner.

1. Twenty-eight H3-JH segments were selected after being experimentally validated in other libraries containing only these H3-JH segments (*see* U.S. Publication Nos. 2009/0181855 and 2010/0056386, and WO/2009/036379).
2. Fifty-seven segments were selected based on their presence within the top 25% of H3-JH segments rank-ordered by usage weights from the bootstrap analysis described above. These 57 H3-JH segments, plus the 28 H3-JH segments of (1) (*i.e.*, 85 segments total) were designated as the “core” H3-JH segments, which, like the core DH segments, were automatically included in ELD-3.
4. Fifteen additional segments were chosen by using an objective function that aimed to maximize the Euclidian distance, between the “core” and the

remaining 200 “non-core” segments, in a multi-dimensional space defined by the following variables: (1) amino acid mismatches to nearest neighbor; and (2) values of the three physicochemical property indices.

Example 9.3.6. Selection of 100 TN1 and 200 N2 segments for Inclusion in ELD-3

TN1 and N2 segments were extracted from the sequences in each sub-sample of the bootstrap procedure, and the 100 TN1 and 200 N2 segments with the highest average segment usage weights were chosen for inclusion into the library, after elimination of sequences with undesirable motifs, namely Cys and Asn residues.

Example 9.3.7. Selection of Nucleotide Sequences to Encode the Segments Chosen for Inclusion in ELD-3

Each of the polypeptide segments chosen for inclusion in the library must be back translated (polypeptide to DNA) into a corresponding oligonucleotide sequence. While a large number of oligonucleotides could possibly encode each polypeptide segment, due to the degeneracy of the genetic code, certain constraints were imposed to select oligonucleotides that were more desirable. First, since ELD-3 was expressed in yeast (*S. cerevisiae*), codons that are rarely used in yeast were avoided. For example, of the six possible codons for Arg, three: CGA, CGC and CGG are used to encode yeast proteins at rates of under 10% (see, for example, Nakamura et al., Nucleic Acids Res., 2000, 28:292), and therefore those three codons were avoided to the extent possible. Second, since many antibodies are produced in Chinese Hamster Ovary (CHO) cells (after discovery *e.g.*, in yeast), the CCG codon (encoding Pro) was also avoided, since it is rarely used by hamsters (Nakamura et al.)

A number of restriction enzymes are employed during the actual construction of the CDRH3 oligonucleotide library (see Example 10 of U.S. Pub. No. 2009/0181855). It is thus desirable to avoid the occurrence of recognition motifs for these restriction enzymes within the CDRH3 polynucleotide sequences. Codons are selected at the individual segment level to avoid introducing recognition motifs for restriction enzymes that may be used downstream. Since such motifs may also be generated by combinatorial assembly of the segments, the segment combinations are also checked and, whenever

possible, codons are changed to eliminate the occurrence of such motifs. Specifically, three restriction enzymes were used during the construction of the currently exemplified CDRH3 library: BsrDI, BbsI, and AvrII. The first two are type II enzymes with non-palindromic recognition sites. The reverse strand of the oligonucleotides encoding the segments was checked explicitly for recognition sites for these two enzymes. In particular, the reverse strands were checked for the motifs GCAATG and CATTGC (for BsrDI) and GAAGAC and GTCTTC (for BbsI). The recognition motif for AvrII is palindromic so the oligonucleotides were only checked for the sequence CCTAGG. However, AvrII is used only to treat TN1 segments, and thus it is not necessary to evaluate its presence in the other segments or their combinations.

An additional constraint that was imposed to improve engineering of the polypeptide to polynucleotide conversion was avoidance of consecutive runs of 6 or more of the same type of base, as this is believed to increase errors during solid phase oligonucleotide synthesis. Therefore, DNA sequences for the segments of ELD-3 were chosen to avoid such motifs. The DNA sequences for the ELD-3 segments are included, with the respective polypeptide sequences, in Table 25. One of ordinary skill in the art will readily recognize that these methods can also be applied to any other library, any restriction sites, any number of nucleotide repeats, and/or to avoid the occurrence of any codons considered undesirable in any organism.

Example 10. Matching of ELD-3 to Human CDRH3 Datasets and Clinically Relevant Antibodies

Among the objectives of the invention is to mimic the V-D-J recombination processes underlying the creation of the human CDRH3 repertoire *in vivo*, thereby increasing the diversity of the CDRH3 library in comparison to other libraries known in the art, while maintaining the human character of CDRH3. One measure of success is the extent to which collections of human reference CDRH3 sequences are represented identically, or via close matches (*e.g.*, less than about 5, 4, 3, or 2 amino acid differences) in any library of the invention. We evaluated this metric using two human CDRH3 sequence reference datasets, both non-overlapping with each other and the HPS: (1) a collection of 666 human CDRH3 sequences (Lee *et al.*, Immunogenetics, 2006, 57: 917;

“Lee-666”); and (2) a collection of 3,000 human CDRH3 sequences randomly chosen from the over 200,000 sequences disclosed in Boyd *et al.*, Science Translational Medicine, 2009, 1: 1-8 (“Boyd-3000”). The results of the random sample of the 3,000 human CDRH3 sequences from Boyd *et al.* was representative of the results of the same analysis as applied to all members of the Boyd *et al.* set (>200,000 CDRH3 sequences).

Figure 7 provides the percentage of CDRH3 sequences in two synthetic libraries, “LUA-141” and ELD-3, that match a sequence from the Lee-666 or Boyd-3000 sets with zero, one, two, three, or more than three amino acid mismatches. Here, “LUA-141” represents a library containing 212 TN1, 278 DH, 141 N2, and 28 H3-JH (*see* U.S. Publication No. 2009/0181855 for details). In particular, it is notable that ELD-3 exhibits a higher percentage of sequences (12.9% and 12.1% for the Lee-666 and Boyd-3000 sets, respectively) that identically match a reference CDRH3 sequence than LUA-141 (8.4% and 6.3% for the Lee-666 and Boyd-3000 sets, respectively). It is also notable that ELD-3 exhibits a higher cumulative percentage of human CDRH3 sequences found with no more than two amino acid mismatches (54.1% and 52.5% for the Lee-666 and Boyd-3000 sets, respectively) relative to LUA-141 (41.2% and 43.7% for the Lee-666 and Boyd-3000 sets, respectively).

Another metric by which antibody libraries can be evaluated is their ability to match “clinically relevant” reference CDRH3 sequences. Figure 8 demonstrates that ELD-3 returns better matches to clinically relevant CDRH3 sequences than the LUA-141 library. Specifically, ELD-3 matches 34 of 55 (62%) clinically validated antibodies within one amino acid, while the LUA-141 library only matches 20 of 55 (37%).

Example 11. Comparison of ELD-3 to LUA-141

ELD-3 has 73 TN1, 92 DH, 119 N2, and 28 H3-JH in common with LUA-141. Thus, 94.5% of the sequences in ELD-3 (4.0×10^8 members) are different from the LUA-141 library (2.3×10^8 members). Figure 9 demonstrates that the combinatorial efficiency of the segments in ELD-3 is greater than that of the segments in LUA-141. Specifically, the ELD-3 segments are more likely to yield a unique CDRH3 than the LUA-141 library segments. This is advantageous, because it allows one to synthesize libraries with increased CDRH3 diversity using fewer segments.

Figure 10 provides the amino acid compositions of the Kabat-CDRH3s of LUA-141, ELD-3, and Human CDRH3 sequences from the HPS.

Figure 11 provides the Kabat-CDRH3 length distribution of LUA-141, ELD-3, and Human CDRH3 sequences from the HPS.

CDRH3 Libraries Synthesized with Degenerate Oligonucleotides

Example 12. Further Increasing CDRH3 Diversity by Utilizing Degenerate Oligonucleotides

The methods described in this example extend the methods taught above, to produce CDRH3 libraries with more members than those of the libraries described above. In particular, one or two degenerate codons were introduced into the DH and or N2 polynucleotide segments, and (generally) no degenerate codon or one degenerate codon were introduced into the H3-JH segments. Segments with different numbers of degenerate codons are also contemplated; for example DH segments with 0, 1, 2, 3, 4, 5, 6, 7, 8, or more degenerate codons, and H3-JH segments with 0, 1, 2, 3, 4, 5, or more degenerate codons. This results in CDRH3 libraries containing greater than about 10^{11} (about 2×10^{11}) distinct CDRH3 amino acid sequences that closely reflect properties, such as length and composition among others, of a reference set of human CDRH3 sequences. As described below, the degenerate positions in the DH segments are usually, but not always, the very N- and/or C-terminal positions, or 5' and 3' end codons (*i.e.*, not necessarily only the first or last base), respectively, when considering the corresponding oligonucleotide sequences. Degenerate codons were similarly used to synthesize N2 segments. Two hundred of the TN1 segments were as described in ELD-3, although libraries with degenerate TN1 segments, or with alternative choices of TN1 segment sequences, fall within the scope of the invention. An additional one hundred TN1 segments complete the set of 300 TN1 segments for this library. The amino acid and nucleotide sequences are listed in Table 26. It is also possible to use mixtures of trinucleotides instead of, or in addition to, degenerate oligonucleotides in order to allow amino acid type variation at one or more selected positions within “base” or “seed” segment sequences (defined below).

Example 13. Selection of DH Segments for Synthesis by Degenerate Oligonucleotides

The segment usage weights were calculated for the 68K DH Theoretical Segment Pool by comparison to the sequences contained in Boyd *et al.* The DH segments with a length of three or more amino acids were ranked according to their segment usage weights (as described above), and the top 201 were designated as “seed” sequences. These seed sequences were then varied by selecting certain positions to incorporate degenerate codons. The positions selected for variance, the amino acids types to which they were varied, were determined by comparing the seed sequences to a reference set of 9,171 DH segments that were a subset of the 68K DH Theoretical Segment Pool. These 9,171 DH segments were selected because their segment usage weight in Boyd *et al.* was significant, meaning that the cumulative segment usage weight (Example 8) is at least 1.0.

Each of the 201 seed sequences was compared to each of the sequences in the reference set of 9,171 DH segments, and those of identical length and differing at a single position were further considered to inform possible variants of the seed. In this manner, the most variable position for each seed was identified and a set of candidate amino acid types was also identified for each position. Finally, a set of degenerate codons was considered, to identify the codon that most faithfully represented the set of candidate amino acid types for each particular position. Degenerate codons encoding stop codons, Cys residues, N-linked glycosylation motifs (*i.e.*, NXS or NXT, where X is any amino acid type), or deamidation motifs (NG) were eliminated from consideration. This process generated 149 unique degenerate oligonucleotide sequences, which collectively encode 3,566 unique polypeptide sequences. Alternative designs generated according to the same principles were also considered, and those having a larger diversity (in terms of the number of unique polypeptide sequences) and smaller RMAX values (see below) were given preference for inclusion in the libraries of the invention. However, it is also conceivable that different criteria could be used to select DH segments from the 68K DH Theoretical Segment Pool, and that libraries including DH segments selected by these different criteria would also fall within the scope of the invention.

Because not all degenerate oligonucleotides encode an identical number of polypeptides, the latter do not occur with uniformly identical weights over the entirety of a given theoretical segment pool (*i.e.*, TN1, DH, N2 and H3-JH) contained within a CDRH3 library of the invention. For example, an individual amino acid sequence X encoded by an oligonucleotide of total degeneracy 4 will have a “weight” of 1/4, while another individual amino acid sequence, Y, encoded by an oligonucleotide of degeneracy 6 will have a weight of 1/6. Moreover, certain amino acid sequences could be encoded by more than one degenerate oligonucleotide, so their weights will be the sum of the individual contributions by each oligonucleotide. Within a given theoretical segment pool, the ratio of the weight of the most heavily weighted polypeptide to that of the least heavily weighted one, RMAX, is an important design criterion that one would ideally like to minimize. The RMAX value may be defined by length, or overall for all of the segments of any given type (*i.e.*, all the DH segments, or all the H3-JH segments, and so on for the TN1, and/or the N2 segments). Table 27 lists the degenerate oligonucleotide sequences, while Table 28 lists the unique polypeptide sequences resulting from these oligonucleotides. These two tables include the DH dimer segments the design of which is detailed below.

Example 13.1. Selection of DH Dimer Segments

A different method was employed to design a set of degenerate oligonucleotides encoding DH dimer sequences. The method aimed to include all of the 45 dimer sequences in ELD-3 plus as many of the other 400 theoretically possible dimer sequences (*i.e.*, 20 residues possible in each of 2 positions = $20 * 20$), minus segments containing Asn (N) residues and excessively hydrophobic dimers (*i.e.*, any dimer combination comprising only F, I, L, M, and/or V residues). This design process ultimately yielded 35 degenerate oligonucleotide sequences encoding 213 unique peptide dimer sequences. As with the selection processes used for all of the other segments of the invention, one of ordinary skill in the art will readily recognize that other criteria could be used to select the DH dimer segments, and that libraries including these segments also fall within the scope of the invention.

Combining the DH dimer segments with the longer DH segments of Example 13, yielded the final set of DH segments of the currently exemplified library, encoded by a total of 184 oligonucleotides (35 encoding dimers and 149 encoding segments having three or more amino acids) versus the 200 oligonucleotides of ELD-3. The 184 oligonucleotides encode a total of 3,779 unique polypeptide sequences: 213 dimers and 3,566 longer segments of three amino acids or greater.

Example 14. Generation of Expanded N2 Diversity

As described above, ELD-3 contains 200 N2 segments. In the currently exemplified library, the empty N2 segment (*i.e.*, no N2, so that the DH segments are joined directly to the H3-JH segments) and monomer N2 segments were the same as in ELD-3. However, degenerate oligonucleotides were used to generate sets of two-, three-, and four-mers that not only recapitulated all of the corresponding sequences in ELD-3 but also resulted in additional diversity. As with the DH segments, these degenerate oligonucleotides were designed to eliminate Asn (in unsuitable positions) and Cys residues, and stop codons. More specifically, Asn residues were allowed at the first position of trimers and at the first or second position of tetramers whenever the subsequent amino acid was not Gly and the next amino acid was not Ser or Thr, thus avoiding deamidation or N-linked glycosylation motifs within the candidate N2 segments. The N2 theoretical segment pool for the currently exemplified library contains one zero-mer (*i.e.*, no N2 segment), 18 monomer, 279 dimer, 339 trimers, and 90 tetramer N2 amino acid sequences, or 727 segments in total. These amino acid sequences are encoded by 1, 18, 81, 36, and 10 oligonucleotides, respectively, for a total of 146 oligonucleotides. All but the first 19 oligonucleotides, those encoding the zero- and one-mers, are degenerate. Table 29 lists the 146 oligonucleotide sequences, while Table 30 lists the resulting 727 unique polypeptide sequences.

Example 15. Generation of Expanded H3-JH Diversity

Application of nucleotide-level progressive deletions on the 5' end of the human IGHJ polynucleotide segments down to the point where only the DNA sequence corresponding to FRM4 remained (*i.e.*, no H3-JH remained), followed by systematic 1-

or 2-bp completions on the same 5' end, resulted in 643 unique H3-JH peptide segments after translation ("643 H3-JH Set"). As done with the DH segments, it is possible to rank order each of the 643 segments by their usage weights obtained after comparison to the approximately 237,000 human sequences from Boyd *et al.*, and the top 200 individual sequences, from those devoid of the undesired motifs described above, were chosen to provide the set of H3-JH segments for the currently exemplified library.

In an alternatively exemplified embodiment, 46 of the 200 H3-JH segments were designed with a two-fold degenerate codon in the first position (*i.e.*, N-terminal or 5' end, respectively, at the peptide and oligonucleotide level), so that, overall, 200 oligonucleotides would encode 246 unique peptide sequences.

In yet other alternatively exemplified embodiments, further use of degenerate codons may be conceived to produce libraries encoded by 90, 100, 200 or more oligonucleotides representing up to 500 distinct polypeptide sequences. Preferably, but not necessarily, these up to 500 unique sequences could be a subset of the sequences in the 643 H3-JH reference set described above, or a subset of variants of these sequences. As exemplified above, H3-JH segments containing undesirable polypeptide motifs may be eliminated from the design. The oligonucleotide sequences for the JH segments are listed on Table 31, while the resulting unique polypeptide sequences are provided in Table 32. In Table 31, nucleotide sequences corresponding to the FRM4 region are also provided, but the "peptide length" value refers to the H3-JH portion only. For simplicity, only the H3-JH peptide sequences are included in Table 32.

Example 16. Extended Diversity Library Design (EDLD)

The TN1, DH, H3-JH, and N2 segments selected above, and provided in Tables 26 to 32, were combined to generate an Extended Diversity Library Design (EDLD) with theoretical diversity of about 2×10^{11} ($300 \text{ TN1} \times 3,779 \text{ DH} \times 727 \text{ N2} \times 246 \text{ H3-JH}$). The oligonucleotides encoding the selected segments were chosen according to the principles of Example 9.3.7.

Figures 12 - 15 illustrate certain characteristics of this design indicating, for example, that about 50% of the approximately 237,000 CDRH3 sequences in Boyd *et al.* may be recapitulated by library sequences with either one or no mismatches (*i.e.*, by

summing the “0” and “1” bins of Figure 12) . The theoretical length distributions (Figure 13) and amino acid compositions (Figure 14) of these libraries also match closely the respective characteristics observed in the same set of human CDRH3 sequences. Figure 15 shows the combinatorial efficiency of the Extended Diversity Library Design. Approximately 65% of the sequences appear only once in the design (*i.e.*, are generated via one non-degenerate combination of segments). Figure 8, previously presented, shows that the Extended Diversity Library Design outperforms both LUA-141 and ELD-3 in terms of matching to clinically relevant human antibody sequences.

Table 1. Germline-like sequences for eight of the VK chassis provided by the invention.

Germline	Junction	Germline-Like CDRL3 Sequence	SEQ ID NO	Germline	Junction	Germline-Like CDRL3 Sequence	SEQ ID NO
VK1-05	1	QQNSYSST		VK2-28	1	MQALQTPT	
VK1-05	2	QQNSYFT		VK2-28	2	MQALQTFT	
VK1-05	3	QQNSYLT		VK2-28	3	MQALQTLT	
VK1-05	4	QQNSYIT		VK2-28	4	MQALQTIT	
VK1-05	5	QQNSYRT		VK2-28	5	MQALQTRT	
VK1-05	6	QQNSYWT		VK2-28	6	MQALQTTWT	
VK1-05	7	QQNSYYT		VK2-28	7	MQALQTYT	
VK1-05	8	QQNSYSPT		VK2-28	8	MQALQTPPT	
VK1-05	9	QQNSYSFT		VK2-28	9	MQALQTPFT	
VK1-05	10	QQNSYSLT		VK2-28	10	MQALQTPLT	
VK1-05	11	QQNSYSIT		VK2-28	11	MQALQTPIT	
VK1-05	12	QQNSYSRT		VK2-28	12	MQALQTPRT	
VK1-05	13	QQNSYSWT		VK2-28	13	MQALQTPWT	
VK1-05	14	QQNSYSYT		VK2-28	14	MQALQTPYT	
VK1-05	15	QQNSYSPFT		VK2-28	15	MQALQTPPFT	
VK1-05	16	QQNSYSPLT		VK2-28	16	MQALQTPPLT	
VK1-05	17	QQNSYSPIIT		VK2-28	17	MQALQTPPIT	
VK1-05	18	QQNSYSPRT		VK2-28	18	MQALQTPPRT	
VK1-05	19	QQNSYSPWT		VK2-28	19	MQALQTPPWT	
VK1-05	20	QQNSYSPYT		VK2-28	20	MQALQTPPYT	
VK1-12	1	QQANSFPT		VK3-11	1	QQRSNWPT	
VK1-12	2	QQANSFFT		VK3-11	2	QQRSNWFT	
VK1-12	3	QQANSFLT		VK3-11	3	QQRSNWLT	
VK1-12	4	QQANSFIT		VK3-11	4	QQRSNWIT	
VK1-12	5	QQANSFRT		VK3-11	5	QQRSNWRT	
VK1-12	6	QQANSFWT		VK3-11	6	QQRSNNWT	
VK1-12	7	QQANSFYT		VK3-11	7	QQRSNWYT	
VK1-12	8	QQANSFPPT		VK3-11	8	QQRSNWPPT	
VK1-12	9	QQANSFPFT		VK3-11	9	QQRSNWPFT	
VK1-12	10	QQANSFPILT		VK3-11	10	QQRSNWPILT	
VK1-12	11	QQANSFPIT		VK3-11	11	QQRSNWPIT	
VK1-12	12	QQANSFPRT		VK3-11	12	QQRSNWPRT	
VK1-12	13	QQANSFPWT		VK3-11	13	QQRSNWPWT	
VK1-12	14	QQANSFPYT		VK3-11	14	QQRSNWPYT	
VK1-12	15	QQANSFPPFT		VK3-11	15	QQRSNWPPFT	
VK1-12	16	QQANSFPPLT		VK3-11	16	QQRSNWPPILT	
VK1-12	17	QQANSFPPIIT		VK3-11	17	QQRSNWPPIT	
VK1-12	18	QQANSFPPrT		VK3-11	18	QQRSNWPPRT	
VK1-12	19	QQANSFPpWT		VK3-11	19	QQRSNWPPWt	
VK1-12	20	QQANSFPpYT		VK3-11	20	QQRSNWPPYt	
VK1-33	1	QQYDNLPT		VK3-15	1	QQYNNWPT	
VK1-33	2	QQYDNLFT		VK3-15	2	QQYNNWFT	
VK1-33	3	QQYDNLILT		VK3-15	3	QQYNNWLt	
VK1-33	4	QQYDNLIT		VK3-15	4	QQYNNWIt	
VK1-33	5	QQYDNLRT		VK3-15	5	QQYNNWRT	

VK1-33	6	QQYDNLWT		VK3-15	6	QQYNNWWT	
VK1-33	7	QQYDNLYT		VK3-15	7	QQYNNWYT	
VK1-33	8	QQYDNLPPPT		VK3-15	8	QQYNNWPPT	
VK1-33	9	QQYDNLFFT		VK3-15	9	QQYNNWFFT	
VK1-33	10	QQYDNLPLT		VK3-15	10	QQYNNWPLT	
VK1-33	11	QQYDNLPIIT		VK3-15	11	QQYNNWPIIT	
VK1-33	12	QQYDNLPRIT		VK3-15	12	QQYNNWPRIT	
VK1-33	13	QQYDNLPPWT		VK3-15	13	QQYNNWPPWT	
VK1-33	14	QQYDNLPPYT		VK3-15	14	QQYNNWPPYT	
VK1-33	15	QQYDNLPPFFT		VK3-15	15	QQYNNWPPFFT	
VK1-33	16	QQYDNLPPPLT		VK3-15	16	QQYNNWPPPLT	
VK1-33	17	QQYDNLPPPIIT		VK3-15	17	QQYNNWPPPIIT	
VK1-33	18	QQYDNLPPPRIT		VK3-15	18	QQYNNWPPPRIT	
VK1-33	19	QQYDNLPPPWT		VK3-15	19	QQYNNWPPPWT	
VK1-33	20	QQYDNLPPPYT		VK3-15	20	QQYNNWPPPYT	
VK1-39	1	QQSYSTPT		VK3-20	1	QQYGSSPT	
VK1-39	2	QQSYSTFT		VK3-20	2	QQYGSSFT	
VK1-39	3	QQSYSTLIT		VK3-20	3	QQYGSSLIT	
VK1-39	4	QQSYSTIT		VK3-20	4	QQYGSSIT	
VK1-39	5	QQSYSTRIT		VK3-20	5	QQYGSSRIT	
VK1-39	6	QQSYSTWT		VK3-20	6	QQYGSSWT	
VK1-39	7	QQSYSTYT		VK3-20	7	QQYGSSYT	
VK1-39	8	QQSYSTPPPT		VK3-20	8	QQYGSSPPPT	
VK1-39	9	QQSYSTPFFT		VK3-20	9	QQYGSSPFFT	
VK1-39	10	QQSYSTPLT		VK3-20	10	QQYGSSPLT	
VK1-39	11	QQSYSTPIIT		VK3-20	11	QQYGSSPIIT	
VK1-39	12	QQSYSTPRIT		VK3-20	12	QQYGSSPRIT	
VK1-39	13	QQSYSTPWT		VK3-20	13	QQYGSSPWT	
VK1-39	14	QQSYSTPYT		VK3-20	14	QQYGSSPYT	
VK1-39	15	QQSYSTPPFFT		VK3-20	15	QQYGSSPPFFT	
VK1-39	16	QQSYSTPPPLT		VK3-20	16	QQYGSSPPPLT	
VK1-39	17	QQSYSTPPPIIT		VK3-20	17	QQYGSSPPPIIT	
VK1-39	18	QQSYSTPPPRIT		VK3-20	18	QQYGSSPPPRIT	
VK1-39	19	QQSYSTPPPWT		VK3-20	19	QQYGSSPPPWT	
VK1-39	20	QQSYSTPPPYT		VK3-20	20	QQYGSSPPPYT	
VK4-01	1	QQYYSTPT					
VK4-01	2	QQYYSTFT					
VK4-01	3	QQYYSTLIT					
VK4-01	4	QQYYSTIT					
VK4-01	5	QQYYSTRIT					
VK4-01	6	QQYYSTWT					
VK4-01	7	QQYYSTYT					
VK4-01	8	QQYYSTPPPT					
VK4-01	9	QQYYSTPFFT					
VK4-01	10	QQYYSTPLT					
VK4-01	11	QQYYSTPIIT					
VK4-01	12	QQYYSTPRIT					
VK4-01	13	QQYYSTPWT					
VK4-01	14	QQYYSTPYT					
VK4-01	15	QQYYSTPPFFT					

VK4-01	16	QYYSTPPLT							
VK4-01	17	QYYSTPPIT							
VK4-01	18	QYYSTPPRT							
VK4-01	19	QYYSTPPWT							
VK4-01	20	QYYSTPPYT							

Table 2. Summary of framework variants for exemplified light chain germlines.

Light Chain Germline	Number of Sequences Analyzed	Primary Framework Positions Selected for Variance	Alternative Framework Positions Selected for Variance
VK1-5	307	4, 49	46
VK1-12	113	4, 49	46, 66
VK1-33	188	4, 66	49
VK1-39	656	4, 49	46
VK2-28	275	46, 49	2, 4
VK3-11	375	4, 36	2, 49
VK3-15	202	4, 49	2, 48
VK3-20	867	4, 49	2, 48
VK4-1	368	4, 49	46, 66

Table 3. Polypeptide sequences of exemplified light chain chassis with variability in CDR11, CDR12, and frameworks. The Kabat numbers for segment boundaries are indicated. Here, L1 and L2 (in the “Category” column) indicate variability in CDR11 and CDR12, respectively, while “F” indicates a framework variant. Sequences designated with both L1 or L2 and F contain variability in both a CDR and framework region.

WO 2012/009568

PCT/US2011/044063

Name	Chassis	Category	FRM1: 1-23	CDR1: 24-34	FRM2: 35-49	CDR2: 50-56	FRM3: 57-88
VK1-39	VK1-39	Germline	DIQMTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.1	VK1-39	L1	DIQMTQSPSSLSASVG DRVITTC	RASQINSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.2	VK1-39	L1	DIQMTQSPSSLSASVG DRVITTC	RASQIDSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.3	VK1-39	L1	DIQMTQSPSSLSASVG DRVITTC	RASQISRYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.6	VK1-39	L2	DIQMTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	GASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.7	VK1-39	L2	DIQMTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	SASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.8	VK1-39	L2	DIQMTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	AASNLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.10	VK1-39	F	DIQLTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.11	VK1-39	FL1	DIQLTQSPSSLSASVG DRVITTC	RASQINSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.12	VK1-39	FL1	DIQLTQSPSSLSASVG DRVITTC	RASQIDSYLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.15	VK1-39	FL1	DIQLTQSPSSLSASVG DRVITTC	RASQISSFLN	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.17	VK1-39	FL2	DIQLTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	SASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.18	VK1-39	FL2	DIQLTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIY	AASNLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.20	VK1-39	F	DIQMTQSPSSLSASVG DRVITTC	RASQISSYLN	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.23	VK1-39	FL1	DIQMTQSPSSLSASVG DRVITTC	RASQISRYLN	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1-39.24	VK1-39	FL1	DIQMTQSPSSLSASVG DRVITTC	RASQISIYLN	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC

WO 2012/009568

PCT/US2011/044063

VK1 - 39.25	VK1 - 39	FL1	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSFLN	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 05	VK1 - 05	Germline	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.1	VK1 - 05	L1	DIQMTQSPSTLSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.5	VK1 - 05	L2	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIY	EASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.6	VK1 - 05	L2	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIY	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.7	VK1 - 05	L12	DIQMTQSPSTLSASVG DRVTTTC	RASQAIISSWLA	WYQQKPGKAPKL LIY	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.8	VK1 - 05	L12	DIQMTQSPSTLSASVG DRVTTTC	RASQSINSWLA	WYQQKPGKAPKL LIY	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.9	VK1 - 05	L12	DIQMTQSPSTLSASVG DRVTTTC	RASQSIGSWLA	WYQQKPGKAPKL LIY	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.10	VK1 - 05	F	DIQLTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.11	VK1 - 05	FL1	DIQLTQSPSTLSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.12	VK1 - 05	FL1	DIQLTQSPSTLSASVG DRVTTTC	RASQAIISSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.14	VK1 - 05	FL1	DIQLTQSPSTLSASVG DRVTTTC	RASQSIGSWLA	WYQQKPGKAPKL LIY	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.20	VK1 - 05	F	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIS	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.21	VK1 - 05	FL1	DIQMTQSPSTLSASVG DRVTTTC	RASQSINSWLA	WYQQKPGKAPKL LIS	DASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.25	VK1 - 05	FL2	DIQLTQSPSTLSASVG DRVTTTC	RASQSIGSWLA	WYQQKPGKAPKL LIY	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 05.26	VK1 - 05	FL2	DIQMTQSPSTLSASVG DRVTTTC	RASQSISSWLA	WYQQKPGKAPKL LIS	KASSLES	GVPSRFSGSGGTEFTLTIISSLQPDDEFATYYC
VK1 - 12	VK1 - 12	Germline	DIQMTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC

WO 2012/009568

PCT/US2011/044063

VK1 - 12.2	VK1 - 12	L1	DIQMTQSPSSVSASVG DRVTTTC	RASQIGSWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.3	VK1 - 12	L1	DIQMTQSPSSVSASVG DRVTTTC	RASQGIDSWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.4	VK1 - 12	L1	DIQMTQSPSSVSASVG DRVTTTC	RASQGISRWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.5	VK1 - 12	L2	DIQMTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	GASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.6	VK1 - 12	L2	DIQMTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	SASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.7	VK1 - 12	L2	DIQMTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	AASNLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.10	VK1 - 12	F	DIQLTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.11	VK1 - 12	FL1	DIQLTQSPSSVSASVG DRVTTTC	RASQDISSWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.14	VK1 - 12	FL1	DIQLTQSPSSVSASVG DRVTTTC	RASQGISRWLA	WYQQKPGKAPKL LIY	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.15	VK1 - 12	FL2	DIQLTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	GASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.16	VK1 - 12	FL2	DIQLTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	SASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.17	VK1 - 12	FL2	DIQLTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIY	AASNLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.20	VK1 - 12	F	DIQMTQSPSSVSASVG DRVTTTC	RASQGISSWLA	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.21	VK1 - 12	FL1	DIQMTQSPSSVSASVG DRVTTTC	RASQDISSWLA	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.23	VK1 - 12	FL1	DIQMTQSPSSVSASVG DRVTTTC	RASQGIDSWLA	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 12.24	VK1 - 12	FL1	DIQMTQSPSSVSASVG DRVTTTC	RASQGISRWLA	WYQQKPGKAPKL LIS	AASSLQS	GVPSRFSGSGGTDFTLTISSLQPEDFATYYC
VK1 - 33	VK1 - 33	Germline	DIQMTQSPSSLSASVG DRVTTTC	QASQDISNYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTFITISSLQPEDIATYYC

WO 2012/009568

PCT/US2011/044063

VK1 - 33.1	VK1 - 33	L1	DIQMTQSPSSLSASVG DRVTTTC	QASQDITNYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.2	VK1 - 33	L1	DIQMTQSPSSLSASVG DRVTTTC	QASQDIANYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.8	VK1 - 33	L2	DIQMTQSPSSLSASVG DRVTTTC	QASQDISNYLN	WYQQKPGKAPKL LIY	DASNLAT	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.10	VK1 - 33	F	DIQLTQSPSSLSASVG DRVTTTC	QASQDISNYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.13	VK1 - 33	FL1	DIQLTQSPSSLSASVG DRVTTTC	QASQDISNSLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.14	VK1 - 33	FL1	DIQLTQSPSSLSASVG DRVTTTC	QASQDISNFLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.17	VK1 - 33	FL2	DIQLTQSPSSLSASVG DRVTTTC	QASQDISNYLN	WYQQKPGKAPKL LIY	DASNLTQ	GVPSRFSGSGGTDFTTISSLQPEDIATYYC
VK1 - 33.20	VK1 - 33	F	DIQMTQSPSSLSASVG DRVTTTC	QASQDISNYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSRSGTDFTTISSLQPEDIATYYC
VK1 - 33.21	VK1 - 33	FL1	DIQMTQSPSSLSASVG DRVTTTC	QASQDITNYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSRSGTDFTTISSLQPEDIATYYC
VK1 - 33.22	VK1 - 33	FL1	DIQMTQSPSSLSASVG DRVTTTC	QASQDIANYLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSRSGTDFTTISSLQPEDIATYYC
VK1 - 33.23	VK1 - 33	FL1	DIQMTQSPSSLSASVG DRVTTTC	QASQDISNSLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSRSGTDFTTISSLQPEDIATYYC
VK1 - 33.24	VK1 - 33	FL1	DIQMTQSPSSLSASVG DRVTTTC	QASQDISNFLN	WYQQKPGKAPKL LIY	DASNLET	GVPSRFSGSRSGTDFTTISSLQPEDIATYYC
VK2 - 28	VK2 - 28	Germline	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHNSG YNYLD	WYLQKPGQSPQL LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2 - 28.1	VK2 - 28	L1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLLYSNG YNYLD	WYLQKPGQSPQL LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2 - 28.2	VK2 - 28	L1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHRNG YNYLD	WYLQKPGQSPQL LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2 - 28.3	VK2 - 28	L1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLLHTNG YNYLD	WYLQKPGQSPQL LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2 - 28.4	VK2 - 28	L1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHNSG NNYLD	WYLQKPGQSPQL LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC

WO 2012/009568

PCT/US2011/044063

VK2-28.5	VK2-28	L2	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQL LIY	LASNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.6	VK2-28	L2	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQL LIY	LGSHRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.10	VK2-28	F	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQV LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.11	VK2-28	L1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLLYSNG YNYLD	WYLQKPGQSPQV LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.15	VK2-28	FL2	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQV LIY	LASNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.17	VK2-28	FL2	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQV LIY	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.20	VK2-28	F	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG YNYLD	WYLQKPGQSPQL LIF	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK2-28.24	VK2-28	FL1	DIVMTQSPLSLPVTPG EPASISC	RSSQSLHHSNG NNYLD	WYLQKPGQSPQL LIF	LGSNRAS	GVPDRFSGSGGTDFTLKISRVEAEDVGVYYC
VK3-11	VK3-11	Germline	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.2	VK3-11	L1	EIVLTQSPATLSLSPG ERATLSC	RASQSVSRYLA	WYQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.3	VK3-11	L1	EIVLTQSPATLSLSPG ERATLSC	RASQSVSNYLA	WYQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.4	VK3-11	L2	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DSSNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.5	VK3-11	L2	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DTSNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.6	VK3-11	L2	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DASKRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.10	VK3-11	F	EIVMTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.13	VK3-11	FL1	EIVMTQSPATLSLSPG ERATLSC	RASQSVSNYLA	WYQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC
VK3-11.14	VK3-11	FL2	EIVMTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DSSNRAT	GIPARFSGSGGTDFTLTISSLEPEDFAVYYC

WO 2012/009568

PCT/US2011/044063

VK3 - 11.15	VK3 - 11	FL2	EIVMTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DTSNRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 11.16	VK3 - 11	FL2	EIVMTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WYQQKPGQAPRL LIY	DASKRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 11.20	VK3 - 11	F	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WFQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 11.21	VK3 - 11	FL1	EIVLTQSPATLSLSPG ERATLSC	RASQSISSYLA	WFQQKPGQAPRL LIY	DASNRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 11.24	VK3 - 11	FL2	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WFQQKPGQAPRL LIY	DSSNRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 11.25	VK3 - 11	FL2	EIVLTQSPATLSLSPG ERATLSC	RASQSVSSYLA	WFQQKPGQAPRL LIY	DTSNRAT	GIPARFSGSGGTDFTLT ISSLEPEDFAVYYC
VK3 - 15	VK3 - 15	Germline	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSNLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.1	VK3 - 15	L1	EIVMTQSPATLSVSPG ERATLSC	RASQSVGSNLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.6	VK3 - 15	L1	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSSLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.7	VK3 - 15	L2	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSNLA	WYQQKPGQAPRL LIY	DASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.8	VK3 - 15	L2	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSNLA	WYQQKPGQAPRL LIY	SASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.10	VK3 - 15	F	EIVLTQSPATLSVSPG ERATLSC	RASQSVSSNLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.11	VK3 - 15	FL1	EIVLTQSPATLSVSPG ERATLSC	RASQSVGSNLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.14	VK3 - 15	FL1	EIVLTQSPATLSVSPG ERATLSC	RASQSVSTNLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.16	VK3 - 15	FL1	EIVLTQSPATLSVSPG ERATLSC	RASQSVSSDLA	WYQQKPGQAPRL LIY	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.20	VK3 - 15	F	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSNLA	WYQQKPGQAPRL LIF	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.21	VK3 - 15	FL1	EIVMTQSPATLSVSPG ERATLSC	RASQSVGSNLA	WYQQKPGQAPRL LIF	GASTRAT	GIPARFSGSGGTEFTLT ISSLQSEDFAVYYC

WO 2012/009568

PCT/US2011/044063

VK3 - 15.25	VK3 - 15	FL1	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSDLA	WYQQKPGQAPRL LIF	GASTRAT	GIPDRFSGSGTGTEFTLT ISSLQSEDFAVYYC
VK3 - 15.26	VK3 - 15	FL1	EIVMTQSPATLSVSPG ERATLSC	RASQSVSSSLA	WYQQKPGQAPRL LIF	GASTRAT	GIPARFSGSGTGTEFTLT ISSLQSEDFAVYYC
VK3 - 20	VK3 - 20	Germline	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIY	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.1	VK3 - 20	L1	EIVLTQSPGTLSLSPG ERATLSC	RASQSVRSSYL A	WYQQKPGQAPRL LIY	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.4	VK3 - 20	L1	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSDYL A	WYQQKPGQAPRL LIY	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.7	VK3 - 20	L2	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIY	GASNRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.8	VK3 - 20	L2	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIY	GASRRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.10	VK3 - 20	F	EIVMTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIY	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.17	VK3 - 20	FL2	EIVMTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIY	GASNRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.20	VK3 - 20	F	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSSYL A	WYQQKPGQAPRL LIS	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK3 - 20.22	VK3 - 20	FL1	EIVLTQSPGTLSLSPG ERATLSC	RASQSVSSNYL A	WYQQKPGQAPRL LIS	GASSRAT	GIPDRFSGSGGTDFTTLT ISRLEPEDFAVYYC
VK4 - 01	VK4 - 01	Germline	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLYSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC
VK4 - 01.1	VK4 - 01	L1	DIVMTQSPDSLAVSLG ERATINC	KSSQSLLYSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC
VK4 - 01.2	VK4 - 01	L1	DIVMTQSPDSLAVSLG ERATINC	KSSQSILYSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC
VK4 - 01.3	VK4 - 01	L1	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLHSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC
VK4 - 01.4	VK4 - 01	L1	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLFFSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC
VK4 - 01.5	VK4 - 01	L1	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLYTSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGGTDFTTLT ISSLQAEDVAVYYC

VK4 - 01.7	VK4 - 01	L2	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLYSSN NKNYLA	WYQQKPGQPPKL LIY	WASSRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC
VK4 - 01.10	VK4 - 01	F	DIVLTQSPDSLAVSLG ERATINC	KSSQSVLYSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC
VK4 - 01.13	VK4 - 01	FL1	DIVLTQSPDSLAVSLG ERATINC	KSSQSVLHSSN NKNYLA	WYQQKPGQPPKL LIY	WASTRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC
VK4 - 01.17	VK4 - 01	FL2	DIVLTQSPDSLAVSLG ERATINC	KSSQSVLYSSN NKNYLA	WYQQKPGQPPKL LIY	WASSRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC
VK4 - 01.20	VK4 - 01	F	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLYSSN NKNYLA	WYQQKPGQPPKL LIS	WASTRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC
VK4 - 01.23	VK4 - 01	FL1	DIVMTQSPDSLAVSLG ERATINC	KSSQSVLHSSN NKNYLA	WYQQKPGQPPKL LIS	WASTRES	GVPDRFSGSGTDFTLTISSLQAEDVAVYYC

Table 4. Jumping dimer and trimer oligonucleotides for the VK1-39 sequences with CDRL3 length nine and F as the junctional amino acid. *i.e.*, The sequences depicted below occur between YYC and FGG, to yield: ...YYC-[89-97]-FGG... .

Name	Oligo	89	90	91	92	93	94	95	96	97
Jumping Dimer										
VK1-39_1	SWMSWMAGCTACA GTACTCCTTWCAC T	DEHLQV	DEHLQV	S	Y	S	T	P	FY	T
VK1-39_2	SWMCAAVNATAACA GTACTCCTTWCAC T	DEHLQV	Q	AEGIKLP QRTV	Y	S	T	P	FY	T
VK1-39_3	SWMCAAAAGCBHCA GTACTCCTTWCAC T	DEHLQV	Q	S	ADFHLPSV Y	S	T	P	FY	T
VK1-39_4	SWMCAAAAGCTACN HCACTCCTTWCAC T	DEHLQV	Q	S	Y	ADFHLNPS TVY	T	P	FY	T
VK1-39_5	SWMCAAAAGCTACA GTBHCCCTTWCAC T	DEHLQV	Q	S	Y	S	ADFHLP SVY	P	FY	T
VK1-39_6	CAGSWMVNATAACA GTACTCCTTWCAC T	Q	DEHLQV	AEGIKLP QRTV	Y	S	T	P	FY	T
VK1-39_7	CAGSWMAGCBHCA GTACTCCTTWCAC T	Q	DEHLQV	S	ADFHLPSV Y	S	T	P	FY	T
VK1-39_8	CAGSWMAGCTACN HCACTCCTTWCAC T	Q	DEHLQV	S	Y	ADFHLNPS TVY	T	P	FY	T
VK1-39_9	CAGSWMAGCTACA GTBHCCCTTWCAC	Q	DEHLQV	S	Y	S	ADFHLP SVY	P	FY	T

	T																		
VK1 - 39_10	CAGCAAVNABHCA GTACTCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_11	CAGCAAVNATACN HCACTCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_12	CAGCAAVNATACA GTBHCCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_13	CAGCAAAGCBHCN HCACTCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_14	CAGCAAAGCBHCA GTBHCCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_15	CAGCAAAGCTACN HCBHCCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
Jumping Trimer																			
VK1 - 39_10_0_9	CAGCAAVNABHCA GTACTCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_11_0_9	CAGCAAVNATACN HCACTCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_12_0_9	CAGCAAVNATACA GTBHCCCTTWCAC T	Q					Q		Q		Q		Q		Q		Q		
VK1 - 39_13_0_9	CAGCAAAGCBHCN HCACTCCTTWCAC	Q					Q		Q		Q		Q		Q		Q		

9	T																
VK1 - 39_14_0_ _	CAGCAAAGCBHCA GTBHCCCTTWCAC T	Q		Q		S		ADFHLPSV Y	S	ADFHLP SVY	P		FY		T		
VK1 - 39_15_0_ _	CAGCAAAGCTACN HCBHCCCTTWCAC T	Q		Q		S		Y	ADFHLNPS TVY	ADFHLP SVY	P		FY		T		
VK1 - 39_t1_0_ _	CAGCAAVNABHCN HCACTCCTTWCAC T	Q		Q		AEGIKLP QRTV		ADFHLPSV Y	ADFHLNPS TVY	T	P		FY		T		
VK1 - 39_t2_0_ _	CAGCAAVNABHCA GTBHCCCTTWCAC T	Q		Q		AEGIKLP QRTV		ADFHLPSV Y	S	ADFHLP SVY	P		FY		T		
VK1 - 39_t3_0_ _	CAGCAAVNATACN HCBHCCCTTWCAC T	Q		Q		AEGIKLP QRTV		Y	ADFHLNPS TVY	ADFHLP SVY	P		FY		T		
VK1 - 39_t4_0_ _	CAGCAAAGCBHCN HCBHCCCTTWCAC T	Q		Q		S		ADFHLPSV Y	ADFHLNPS TVY	ADFHLP SVY	P		FY		T		

Table 5. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDR13 length 8.

	Sequence of Synthesized Oligonucleotide										Portion of Oligonucleotide Corresponding to CDR13 Proper
Name											
Jumping Dimer											
VK1 - 05 1 0 8	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAA	TAGTTACTWCACTTTTGGCGGAGGACCAAG	SWMSWMTACAA								TAGTTACTWCACT
VK1 - 05 10 0 8	CCTGATGATTTTGCAACTTATTACTGCGCAGCAGBHCVRMAGTTACTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACTWCACT									TACTWCACT
VK1 - 05 11 0 8	CCTGATGATTTTGCAACTTATTACTGCGCAGCAGBHCAA	TMBCTACTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHCAA								TMBCTACTWCACT
VK1 - 05 12 0 8	CCTGATGATTTTGCAACTTATTACTGCGCAGCAGBHCAA	TAGTYWCTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHCAA								TAGTYWCTWCACT
VK1 - 05 13 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCTA	CTWCACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTA								CTWCACT
VK1 - 05 14 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAGTYWCTWCACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCTWCACT									
VK1 - 05 15 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAA	TMBCYWCTWCACTTTTGGCGGAGGACCAAG	CAGCAGTACAA								TMBCYWCTWCACT
VK1 - 05 2 0 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAA	TAGTTACTWCACTTTTGGCGGAGGACCAAG	SWMCAGBHCAA								TAGTTACTWCACT
VK1 - 05 3 0 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAGTTACTWCACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACTWCACT									
VK1 - 05 4 0 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAA	TMBCTACTWCACTTTTGGCGGAGGACCAAG	SWMCAGTACAA								TMBCTACTWCACT
VK1 - 05 5 0 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAA	TAGTYWCTWCACTTTTGGCGGAGGACCAAG	SWMCAGTACAA								TAGTYWCTWCACT
VK1 - 05 6 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAA	TAGTTACTWCACTTTTGGCGGAGGACCAAG	CAGSWMBHCAA								TAGTTACTWCACT
VK1 - 05 7 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTA	CVRMAGTTACTWCACTTTTGGCGGAGGACCAAG	CAGSWMTA								CVRMAGTTACTWCACT
VK1 - 05 8 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAA	TMBCTACTWCACTTTTGGCGGAGGACCAAG	CAGSWMTA								CAA
VK1 - 05 9 0 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAA	TAGTYWCTWCACTTTTGGCGGAGGACCAAG	CAGSWMTA								CA
VK1 - 05 1 1 8	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAA	TAGTTACMTCACTTTTGGCGGAGGACCAAG	SWMSWMTACAA								TAGTTACMTCACT
VK1 - 05 10 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTTACMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACMTCACT									TACMTCACT
VK1 - 05 11 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAA	TMBCTACMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCAA								TMBCTACMTCACT
VK1 - 05 12 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAA	TAGTYWCMTCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCAA								TAGTYWCMTCAC T
VK1 - 05 13 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCTA	CMTCAC TTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTA								CMTCAC T
VK1 - 05 14 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAGTYWCMTCAC TTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCMTCAC T									
VK1 - 05 15 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAA	TMBCYWCMTCAC TTTTGGCGGAGGACCAAG	CAGCAGTACAA								TMBCYWCMTCAC T
VK1 - 05 2 1 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAA	TAGTTACMTCACTTTTGGCGGAGGACCAAG	SWMCAGBHCAA								TAGTTACMTCACT
VK1 - 05 3 1 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAGTTACMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACMTCACT									
VK1 - 05 4 1 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAA	TMBCTACMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACAA								TMBCTACMTCACT
VK1 - 05 5 1 8	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAA	TAGTYWCMTCAC TTTTGGCGGAGGACCAAG	SWMCAGTACAA								TAGTYWCMTCAC T
VK1 - 05 6 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAA	TAGTTACMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHCAA								TAGTTACMTCACT
VK1 - 05 7 1 8	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTA	CVRMAGTTACMTCACTTTTGGCGGAGGACCAAG	CAGSWMTA								CVRMAGTTACMTCACT

Page 131 of 420

VK1-12	10	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNANHCAGTTTCTWCAC	TTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	11	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATNHC	TTTCTWCAC	TTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC
VK1-12	12	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATAGTYWC	TWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	13	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCANHCNHC	TTTCTWCAC	TTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC
VK1-12	14	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCANHCAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	15	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCAAAATNHC	YWC	TTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC
VK1-12	2	0	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGRNAAATAGTT	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	3	0	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCANHCAGTT	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	4	0	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATNHC	TTTCTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	5	0	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	6	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMRNAAATAGTT	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	7	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCANHCAGTT	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	8	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCAAATNHC	TTTCTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	9	0	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCAAATAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	1	1	8	CCTGAAGATTTTGCAACTTATTACTGTSWMSWMCAAATAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	10	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNANHCAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	11	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATNHC	TTTCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	12	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	13	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCANHCNHC	TTTCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	14	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCANHCAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	15	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGGCAAAATNHC	YWC	TTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC
VK1-12	2	1	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGRNAAATAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	3	1	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCANHCAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	4	1	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATNHC	TTTCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	5	1	8	CCTGAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	6	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMRNAAATAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	7	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCANHCAGTT	TCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	8	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCAAATNHC	TTTCMTCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	9	1	8	CCTGAAGATTTTGCAACTTATTACTGTCAGSWMGCAAATAGTYW	CTWCACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	1	2	8	CCTGAAGATTTTGCAACTTATTACTGTSWMSWMCAAATAGTT	TCWGGACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	10	2	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNANHCAGTT	TCWGGACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	11	2	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATNHC	TTTCWGGACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC
VK1-12	12	2	8	CCTGAAGATTTTGCAACTTATTACTGTCAGCAGRNAAATAGTYW	CTWGGACTTTTGGCGGAGGGACCAAG	TTTCTWCAC	TTTCTWCAC	TTTCTWCAC

VVK1 - 12	13	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAGCAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG																	
VVK1 - 12	14	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAGCAG	GCANHC	AGTY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	AGTY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	AGTY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCANHC	AGTY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG													
VVK1 - 12	15	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAGCAG	GCAAA	TNHC	Y	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCAAA	TNHC	Y	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCAAA	TNHC	Y	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAGCAG	GCAAA	TNHC	Y	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG									
VVK1 - 12	2	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	AGR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	AGR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	AGR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	AGR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	3	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	CAG	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	4	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	CAG	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	5	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	CAG	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAG	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	6	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWMR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWMR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWMR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWMR	NAAA	TAG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	7	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	8	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	CW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	9	2	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	1	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMS	W	M	GCAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	M	GCAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	M	GCAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	M	GCAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	
VVK1 - 12	10	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAGR	NANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	11	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAGR	NAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	12	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAGR	NAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAGR	NAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	13	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	TTCW	GGACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	14	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAG	GCANHC	AGTY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	AGTY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	AGTY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	AGTY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCANHC	AGTY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG
VVK1 - 12	15	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	CAG	GCAAA	TNHC	Y	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCAAA	TNHC	Y	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCAAA	TNHC	Y	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	GCAAA	TNHC	Y	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	2	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	2	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	SWMC	CAGR	NAAA	TAG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	7	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCANHC	AG	TT	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	8	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TNHC	T	C	CCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	9	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	10	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	11	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	12	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	13	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	14	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	15	3	8	CCTGAA	GATTTT	GCAACTTAT	TACTGT	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG	CAG	SWM	GCAAA	TAG	TY	WC	CCCTACTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 12	1	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	SWMS	W	TAC	GATAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	TAC	GATAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	TAC	GATAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	SWMS	W	TAC	GATAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	10	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	BHC	NHCAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	NHCAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	NHCAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	NHCAA	TCT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	11	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	BHC	GATNHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATNHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATNHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATNHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	12	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	BHC	GATAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	BHC	GATAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	13	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	TAC	NHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHC	CT	C	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	14	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					
VVK1 - 33	15	0	8	CCTGAA	GATATT	GCAACATAT	TACTGT	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG	CAG	CAG	TAC	NHCAA	TY	WC	TWCACTTTT	TGGCGG	AGGGAC	CCAAG					

VK1 - 33	2	0	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGBHCAGATAAATCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGATAAATCTCTWCACCT
VK1 - 33	3	0	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACNHCAAATCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAAATCTCTWCACCT
VK1 - 33	4	0	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACGATNHCCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACGATNHCCTCTWCACCT
VK1 - 33	5	0	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACGATAAATYWC TWACCTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAAATYWC TWACCT
VK1 - 33	6	0	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMBHCAGATAAATCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGATAAATCTCTWCACCT
VK1 - 33	7	0	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACNHCAAATCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAAATCTCTWCACCT
VK1 - 33	8	0	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACGATNHCCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCCTCTWCACCT
VK1 - 33	9	0	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACGATAAATYWC TWACCTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAAATYWC TWACCT
VK1 - 33	1	1	8	CCTGAAGATATTGCAACATATTACTGTSWMSWMTACGATAAATCTCMTCACTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAAATCTCMTCACT
VK1 - 33	10	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCNHCAAATCTCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAATCTCMTCACT
VK1 - 33	11	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCAGT NHCCTCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGT NHCCTCMTCACT
VK1 - 33	12	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCAGATAAATYWC MTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGATAAATYWC MTCACT
VK1 - 33	13	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACNHCNHCCTCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCMTCACT
VK1 - 33	14	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACNHCAAATYWC MTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCAAATYWC MTCACT
VK1 - 33	15	1	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACGATNHCYWC MTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACGATNHCYWC MTCACT
VK1 - 33	2	1	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGBHCAGATAAATCTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGATAAATCTCMTCACT
VK1 - 33	3	1	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACNHCAAATCTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAAATCTCMTCACT
VK1 - 33	4	1	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACGATNHCCTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATNHCCTCMTCACT
VK1 - 33	5	1	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACGATAAATYWC MTCACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAAATYWC MTCACT
VK1 - 33	6	1	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMBHCAGATAAATCTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGATAAATCTCMTCACT
VK1 - 33	7	1	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACNHCAAATCTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAAATCTCMTCACT
VK1 - 33	8	1	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACGATNHCCTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCCTCMTCACT
VK1 - 33	9	1	8	CCTGAAGATATTGCAACATATTACTGTCAGSWMTACGATAAATYWC MTCACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAAATYWC MTCACT
VK1 - 33	1	2	8	CCTGAAGATATTGCAACATATTACTGTSWMSWMTACGATAAATCTCWGGACCTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAAATCTCWGGACT
VK1 - 33	10	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCNHCAAATCTCWGGACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAATCTCWGGACT
VK1 - 33	11	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCAGT NHCCTCWGGACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGT NHCCTCWGGACT
VK1 - 33	12	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGBHCAGATAAATYWC WGGACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGATAAATYWC WGGACT
VK1 - 33	13	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACNHCNHCCTCWGGACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCWGGACT
VK1 - 33	14	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACNHCAAATYWC WGGACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCAAATYWC WGGACT
VK1 - 33	15	2	8	CCTGAAGATATTGCAACATATTACTGTCAGCAGTACGATNHCYWC WGGACCTTTTGGCGGAGGGACCAAG	CAGCAGTACGATNHCYWC WGGACT
VK1 - 33	2	2	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGBHCAGATAAATCTCWGGACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGATAAATCTCWGGACT
VK1 - 33	3	2	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACNHCAAATCTCWGGACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAAATCTCWGGACT
VK1 - 33	4	2	8	CCTGAAGATATTGCAACATATTACTGTSWMCAGTACGATNHCCTCWGGACCTTTTGGCGGAGGGACCAAG	SWMCAGTACGATNHCCTCWGGACT

VK1-33	5	2	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMCAGTACGATAATAAT	YWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAATAAT	YWCWGGACT
VK1-33	6	2	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMBHCGATAAATCT	CWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCGATAAATCT	CWGGACT
VK1-33	7	2	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACNHCAAATCT	CWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAAATCT	CWGGACT
VK1-33	8	2	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACGATNHCTC	WGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCTC	WGGACT
VK1-33	9	2	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACGATAAT	YWCWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAAT	YWCWGGACT
VK1-33	1	3	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMSWMTACGATAATCT	CCCCTACTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAATCT	CCCCTACT
VK1-33	10	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGBHCNHCAAATCT	CCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAATCT	CCCCTACT
VK1-33	11	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGBHCAGATNHCT	CCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGATNHCT	CCTCCTACT
VK1-33	12	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGBHCAGATAAT	YWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGATAAT	YWCCTACT
VK1-33	13	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGTACNHCHCTCCCT	ACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCHCTCCCT	ACT
VK1-33	14	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGTACNHCAAAT	YWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCAAAT	YWCCTACT
VK1-33	15	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGCAGTACGATNHCHCT	CCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACGATNHCHCT	CCCCTACT
VK1-33	2	3	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMCAGBHCAGATAATCT	CCCCTACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGATAATCT	CCCCTACT
VK1-33	3	3	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMCAGTACNHCAAATCT	CCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAAATCT	CCTCCTACT
VK1-33	4	3	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMCAGTACGATNHCT	CCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATNHCT	CCTCCTACT
VK1-33	5	3	8	CCTGAAGATATTGCAACATATTA	CTGT	SWMCAGTACGATAAT	YWCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAAT	YWCCTACT
VK1-33	6	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMBHCGATAAATCT	CCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMBHCGATAAATCT	CCTCCTACT
VK1-33	7	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACNHCAAATCT	CCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAAATCT	CCTCCTACT
VK1-33	8	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACGATNHCT	CCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCT	CCTCCTACT
VK1-33	9	3	8	CCTGAAGATATTGCAACATATTA	CTGT	CAGSWMTACGATAAT	YWCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAAT	YWCCTACT
VK1-39	1	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	SWMSWAGCTACAGTACTT	WCACCT
VK1-39	10	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	VNABHCAGTACTT
VK1-39	11	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	VNATACNHCACTT
VK1-39	12	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	VNATACAGTBHCT
VK1-39	13	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	AGCBHCNHCACTT
VK1-39	14	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	AGCBHCAGTBHCT
VK1-39	15	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGCAA	AGCTACNHCBHCT
VK1-39	2	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	SWMCAA	VNATACAGTACTT
VK1-39	3	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	SWMCAA	AGCBHCAGTACTT
VK1-39	4	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	SWMCAA	AGCTACNHCACTT
VK1-39	5	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	SWMCAA	AGCTACAGTBHCT
VK1-39	6	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTT	WCACCT
VK1-39	7	0	8	CCTGAAGATTTTGCAACTTA	CTACT	AGTACAGTACTT	WCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCAGTACTT	WCACCT

VK1-39	8	0	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWMAGCTACNHCACTTWCAC	TTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTTWCAC	TTTGGCGGAGGGACCAAG
VK1-39	9	0	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWMAGCTACAGTBHC	TTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHC	TTTGGCGGAGGGACCAAG
VK1-39	1	1	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMSWMAGCTACAGTACTMT	TTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTMT	TTTGGCGGAGGGACCAAG
VK1-39	10	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	11	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	12	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	13	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	14	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	15	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	2	1	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	3	1	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	4	1	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	5	1	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	6	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	7	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	8	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	9	1	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	1	2	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMSWM	TTTGGCGGAGGGACCAAG	SWMSWM	TTTGGCGGAGGGACCAAG
VK1-39	10	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	11	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	12	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	13	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	14	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	15	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGCAA	TTTGGCGGAGGGACCAAG	CAGCAA	TTTGGCGGAGGGACCAAG
VK1-39	2	2	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	3	2	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	4	2	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	5	2	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMC	TTTGGCGGAGGGACCAAG	SWMC	TTTGGCGGAGGGACCAAG
VK1-39	6	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	7	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	8	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	9	2	8	CCTGAAGATTTTGCAACTTACTACTGT	CAGSWM	TTTGGCGGAGGGACCAAG	CAGSWM	TTTGGCGGAGGGACCAAG
VK1-39	1	3	8	CCTGAAGATTTTGCAACTTACTACTGT	SWMSWM	TTTGGCGGAGGGACCAAG	SWMSWM	TTTGGCGGAGGGACCAAG

VK1 - 39	10	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAVNABHCAGTACTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTACTCCTACT
VK1 - 39	11	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAVNATACNHCACTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCACTCCTACT
VK1 - 39	12	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAVNATACAGTBHCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACAGTBHCCCTACT
VK1 - 39	13	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCBHCNHCACTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCNHCACTCCTACT
VK1 - 39	14	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCBHCAGTBHCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCAGTBHCCCTACT
VK1 - 39	15	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACNHCBHCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCTACNHCBHCCCTACT
VK1 - 39	2	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAVNATACAGTACTCCTACTTTTGGCGGAGGGACCAAG	SWMCAAVNATACAGTACTCCTACT
VK1 - 39	3	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCBHCAGTACTCCTACTTTTGGCGGAGGGACCAAG	SWMCAAAAGCBHCAGTACTCCTACT
VK1 - 39	4	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACNHCACTCCTACTTTTGGCGGAGGGACCAAG	SWMCAAAAGCTACNHCACTCCTACT
VK1 - 39	5	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACAGTBHCCCTACTTTTGGCGGAGGGACCAAG	SWMCAAAAGCTACAGTBHCCCTACT
VK1 - 39	6	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACAGTACTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTCCTACT
VK1 - 39	7	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCBHCAGTACTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCAGTACTCCTACT
VK1 - 39	8	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACNHCACTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTCCTACT
VK1 - 39	9	3	8	CCTGAAGATTTTGCAACTTACTACTGTGAGCAAAAGCTACAGTBHCCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCCTACT
VK2 - 28	1	0	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACTCCAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTTWCACCT
VK2 - 28	10	0	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNAMNACAGACTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTTWCACCT
VK2 - 28	11	0	8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGVNACTCSRMACTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTTWCACCT
VK2 - 28	12	0	8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGVNACTCCAGVBC	ATGCAGVNACTCCAGVBC
VK2 - 28	13	0	8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGGCAMNASRM	ATGCAGGCAMNASRMACTTWCACCT
VK2 - 28	14	0	8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGGCAMNACAGVBC	ATGCAGGCAMNACAGVBC
VK2 - 28	15	0	8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGGCAC	ATGCAGGCAC
VK2 - 28	2	0	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTTWCACCT
VK2 - 28	3	0	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTTWCACCT
VK2 - 28	4	0	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAC	DTSCAGGCAC
VK2 - 28	5	0	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAC	DTSCAGGCAC
VK2 - 28	6	0	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMVNACTCCAGACTTWCACCTTTTGGCGGAGGGACCAAG	ATGSWMVNACTCCAGACTTWCACCT
VK2 - 28	7	0	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCAMNACAGACTTWCACCTTTTGGCGGAGGGACCAAG	ATGSWMGCAMNACAGACTTWCACCT
VK2 - 28	8	0	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCAC	ATGSWMGCAC
VK2 - 28	9	0	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCAC	ATGSWMGCAC
VK2 - 28	1	1	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCAC	DTSSWMGCAC
VK2 - 28	10	1	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNAMNACAGACTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTTWCACCT
VK2 - 28	11	1	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNACTCSRMACTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTTWCACCT
VK2 - 28	12	1	8	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNACTCCAGVBC	ATGCAGVNACTCCAGVBC

VK2-28	13	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNASRM	ACTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRM	ACTMTCACT
VK2-28	14	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNACAG	VB	CMTCAC	TTTGGCGGAGGGACCAAG
VK2-28	15	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAC	TCSR	MVB	CMTCAC
VK2-28	2	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGVNA	CTCCAG	ACTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	3	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAMNACAG	ACTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAG	ACTMTCACT
VK2-28	4	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAC	TCSR	MACT	MTCACTTTTGGCGGAGGGACCAAG
VK2-28	5	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAC	TCCAG	VBC	MTCACTTTTGGCGGAGGGACCAAG
VK2-28	6	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMVNA	CTCAG	ACTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	7	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCAMNACAG	ACTMTCACTTTTGGCGGAGGGACCAAG	ATGSWMGCAMNACAG	ACTMTCACT
VK2-28	8	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCA	CTCSR	MACT	MTCACTTTTGGCGGAGGGACCAAG
VK2-28	9	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCA	CTCCAG	VBC	MTCACTTTTGGCGGAGGGACCAAG
VK2-28	1	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSSWMGCA	CTCCAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG
VK2-28	10	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	MACAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG
VK2-28	11	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	CTCSR	MACT	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	12	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	CTCAG	VBC	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	13	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNASRM	ACTWGG	ACTTTTGGCGGAGGGACCAAG	
VK2-28	14	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNACAG	VBC	WGGACTTTTGGCGGAGGGACCAAG	
VK2-28	15	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAC	TCSR	MVB	CMWGGACTTTTGGCGGAGGGACCAAG
VK2-28	2	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGVNA	CTCCAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG
VK2-28	3	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAMNACAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG	
VK2-28	4	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAC	TCSR	MACT	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	5	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSCAGGCAC	TCCAG	VBC	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	6	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMVNA	CTCCAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG
VK2-28	7	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCAMNACAG	ACTWGG	ACTTTTGGCGGAGGGACCAAG	
VK2-28	8	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCA	CTCSR	MACT	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	9	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGSWMGCA	CTCAG	VBC	WGGACTTTTGGCGGAGGGACCAAG
VK2-28	1	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CDTSSWMGCA	CTCCAG	ACTCCT	ACTTTTGGCGGAGGGACCAAG
VK2-28	10	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	MACAG	ACTCCT	ACTTTTGGCGGAGGGACCAAG
VK2-28	11	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	CTCSR	MACT	CCTACTTTTGGCGGAGGGACCAAG
VK2-28	12	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNA	CTCCAG	VBC	CCCTACTTTTGGCGGAGGGACCAAG
VK2-28	13	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNASRM	ACTCCT	ACTTTTGGCGGAGGGACCAAG	
VK2-28	14	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAMNACAG	VBC	CCCT	ACTTTTGGCGGAGGGACCAAG
VK2-28	15	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGGCAC	TCSR	MVB	CCCTACTTTTGGCGGAGGGACCAAG

VK2-28	2	3	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCCTACTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTACT
VK2-28	3	3	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTACTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTACT
VK2-28	4	3	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSRMACTCCTACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCSRMACTCCTACT
VK2-28	5	3	8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCAGVBCCCTACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCCAGVBCCCTACT
VK2-28	6	3	8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMVMVNACTCCAGACTCCTACTTTTGGCGGAGGGACCAAG	ATGSMVMVNACTCCAGACTCCTACT
VK2-28	7	3	8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMWGCAMNACAGACTCCTACTTTTGGCGGAGGGACCAAG	ATGSMWGCAMNACAGACTCCTACT
VK2-28	8	3	8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMWGCACCTCSRMACTCCTACTTTTGGCGGAGGGACCAAG	ATGSMWGCACCTCSRMACTCCTACT
VK2-28	9	3	8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMWGCACCTCAGVBCCCTACTTTTGGCGGAGGGACCAAG	ATGSMWGCACCTCAGVBCCCTACT
VK3-11	1	0	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWMAAGTAATTGGTWCACTTTTGGCGGAGGGACCAAG	SWMSWMAAGTAATTGGTWCACT
VK3-11	10	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTNHCAATTTGGTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCAATTTGGTWCACT
VK3-11	11	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTNHCTGGTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGTWCACT
VK3-11	12	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTAATYWC'TWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWC'TWCACT
VK3-11	13	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCNHC'TGGTWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHC'TGGTWCACT
VK3-11	14	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHC'AATYWC'TWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHC'AATYWC'TWCACT
VK3-11	15	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAAGTNHCYWC'TWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAGTNHCYWC'TWCACT
VK3-11	2	0	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGBHCAGTAATTGGTWCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGTWCACT
VK3-11	3	0	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGANHC'AATTTGGTWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHC'AATTTGGTWCACT
VK3-11	4	0	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTNHCTGGTWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTNHCTGGTWCACT
VK3-11	5	0	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTAATYWC'TWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTAATYWC'TWCACT
VK3-11	6	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMBHCAGTAATTGGTWCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTGGTWCACT
VK3-11	7	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGANHC'AATTTGGTWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHC'AATTTGGTWCACT
VK3-11	8	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAGTNHCTGGTWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAGTNHCTGGTWCACT
VK3-11	9	0	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAGTAATYWC'TWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAGTAATYWC'TWCACT
VK3-11	1	1	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWMAAGTAATTTGGMTCACTTTTGGCGGAGGGACCAAG	SWMSWMAAGTAATTTGGMTCACT
VK3-11	10	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCNHC'AATTTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHC'AATTTGGMTCACT
VK3-11	11	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTNHCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGMTCACT
VK3-11	12	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTAATYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCMTCACT
VK3-11	13	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCNHC'TGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHC'TGGMTCACT
VK3-11	14	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHC'AATYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHC'AATYWCMTCACT
VK3-11	15	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAAGTNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAGTNHCYWCMTCACT
VK3-11	2	1	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGBHCAGTAATTTGGMTCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTTGGMTCACT
VK3-11	3	1	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGANHC'AATTTGGMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHC'AATTTGGMTCACT
VK3-11	4	1	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTNHCTGGMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTNHCTGGMTCACT

VK3-11	5	1	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAAGTAATYWCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCMTCACT
VK3-11	6	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMBHCAGTAATTTGGMTCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTTGGMTCACT
VK3-11	7	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGANHCAATTGGMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGMTCACT
VK3-11	8	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAAGTNHCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCTGGMTCACT
VK3-11	9	1	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAAGTAATYWCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCMTCACT
VK3-11	1	2	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWMAAGTAATTGGWGGACTTTTGGCGGAGGGACCAAG	SWMSWMAAGTAATTGGWGGACT
VK3-11	10	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCNHCAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATTGGWGGACT
VK3-11	11	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTNHCTGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGWGGACT
VK3-11	12	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCWGGACT
VK3-11	13	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAGANHCHCTGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCHCTGGWGGACT
VK3-11	14	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAGANHCAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCWGGACT
VK3-11	15	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAAAGTNHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCWGGACT
VK3-11	2	2	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGBHCAGTAATTGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGWGGACT
VK3-11	3	2	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGANHCAATTGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGWGGACT
VK3-11	4	2	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAAGTNHCTGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCTGGWGGACT
VK3-11	5	2	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAAGTAATYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCWGGACT
VK3-11	6	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMBHCAGTAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTGGWGGACT
VK3-11	7	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGANHCAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGWGGACT
VK3-11	8	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAAGTNHCTGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCTGGWGGACT
VK3-11	9	2	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGAAAGTAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCWGGACT
VK3-11	1	3	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWMAAGTAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMSWMAAGTAATTGGCCTACT
VK3-11	10	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCNHCAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATTGGCCTACT
VK3-11	11	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTNHCTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTACT
VK3-11	12	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCAGTAATYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCCTACT
VK3-11	13	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCHCTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCHCTGGCCTACT
VK3-11	14	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCAATYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTACT
VK3-11	15	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAAAGTNHCYWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCCTACT
VK3-11	2	3	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGBHCAGTAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTACT
VK3-11	3	3	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGANHCAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGCCTACT
VK3-11	4	3	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAAGTNHCTGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCTGGCCTACT
VK3-11	5	3	8	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAAGTAATYWCCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCCCTACT
VK3-11	6	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMBHCAGTAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTGGCCTACT
VK3-11	7	3	8	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGANHCAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGCCTACT

VK3 - 11	8	3	8	CC	TG	AA	G	A	T	T	T	G	C	A	G	T	T	T	A	T	T	A	C	T	G	T	C	A	G	S	W	M	A	G	A	A	G	T	N	H	C	T	G	C	C	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	A	G	A	A	G	T	N	H	C	T	G	C	C	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	A	G	A	A	G	T	N	H	C	T	G	C	C	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	G	T	S	W	S	W	M	T	A	C	A	A	T	A	C	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W	M	T	A	C	A	A	T	A	C	T	T	T	T	G	C	G	G	A	G	G	A	C	C	A	A	G	C	A	G	S	W</
----------	---	---	---	----	----	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-----

VK3-15	10	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHC	CAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHC	CAATTGGWGGACT		
VK3-15	11	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATNHC	TGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATNHC	TGGWGGACT
VK3-15	12	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATAATYWCWGGACT		
VK3-15	13	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	NHCTGGWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	NHCTGGWGGACT		
VK3-15	14	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	CAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	CAATYWCWGGACT		
VK3-15	15	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTAC	CAATNHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	CAATNHCYWCWGGACT		
VK3-15	2	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGBHC	CAATAATTGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHC	CAATAATTGGWGGACT		
VK3-15	3	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTACNHC	CAATTGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHC	CAATTGGWGGACT		
VK3-15	4	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTAC	CAATNHC	TGGWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATNHC	TGGWGGACT
VK3-15	5	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTAC	CAATAATYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATAATYWCWGGACT		
VK3-15	6	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMBHC	CAATAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHC	CAATAATTGGWGGACT		
VK3-15	7	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHC	CAATTGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHC	CAATTGGWGGACT		
VK3-15	8	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATNHC	TGGWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATNHC	TGGWGGACT
VK3-15	9	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATAATYWCWGGACT		
VK3-15	1	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMSWMTAC	CAATAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMSWMTAC	CAATAATTGGCCTACT		
VK3-15	10	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHC	CAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHC	CAATTGGCCTACT		
VK3-15	11	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATNHC	TGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATNHC	TGGCCTACT
VK3-15	12	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATAATYWCCTCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATAATYWCCTCTACT		
VK3-15	13	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	NHCTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	NHCTGGCCTACT		
VK3-15	14	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	CAATYWCCTCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	CAATYWCCTCTACT		
VK3-15	15	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTAC	CAATNHCYWCCTCTACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	CAATNHCYWCCTCTACT		
VK3-15	2	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGBHC	CAATAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGBHC	CAATAATTGGCCTACT		
VK3-15	3	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTACNHC	CAATTGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHC	CAATTGGCCTACT		
VK3-15	4	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTAC	CAATNHC	TGGCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATNHC	TGGCCTACT
VK3-15	5	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	TSWMCAGTAC	CAATAATYWCCTCTACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATAATYWCCTCTACT		
VK3-15	6	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMBHC	CAATAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMBHC	CAATAATTGGCCTACT		
VK3-15	7	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHC	CAATTGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHC	CAATTGGCCTACT		
VK3-15	8	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATNHC	TGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATNHC	TGGCCTACT
VK3-15	9	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATNHC	TGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATNHC	TGGCCTACT
VK3-15	10	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATNHC	TGGCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATNHC	TGGCCTACT
VK3-20	1	0	8	CCTGAAGATTTTGCAGTGTA	TTACTGTSWMSWMTAC	GGAAGTAGTTWCAC	TTTGGCGGAGGGACCAAG	SWMSWMTAC	GGAAGTAGTTWCAC	TTTGGCGGAGGGACCAAG
VK3-20	10	0	8	CCTGAAGATTTTGCAGTGTA	TTACTGTCAGCAGBHC	BC	HACAGTAGTTWCAC	CAGCAGBHC	BC	HACAGTAGTTWCAC
VK3-20	11	0	8	CCTGAAGATTTTGCAGTGTA	TTACTGTCAGCAGBHC	CGGAVNCAGTTWCAC	TTTGGCGGAGGGACCAAG	CAGCAGBHC	CGGAVNCAGTTWCAC	TTTGGCGGAGGGACCAAG
VK3-20	12	0	8	CCTGAAGATTTTGCAGTGTA	TTACTGTCAGCAGBHC	CGGAAAGTBHCTWCAC	TTTGGCGGAGGGACCAAG	CAGCAGBHC	CGGAAAGTBHCTWCAC	TTTGGCGGAGGGACCAAG

VK3-20	13	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCVNCAGTTWCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACBHCVNCAGTTWCAC	TTTGGCGGAGGGACCAAG
VK3-20	14	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCAGTBHC	TWCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACBHCAGTBHC
VK3-20	15	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACGGAVNCBHC	TWCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACGGAVNCBHC
VK3-20	2	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGBHC	GGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	3	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG	CAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG
VK3-20	4	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACGGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	5	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACGGAAAGTBHC	TWCAC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	6	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMBHC	GGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	7	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACBHCAGTAG	TTTGGCGGAGGGACCAAG	CAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG
VK3-20	8	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACGGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	9	0	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACGGAAAGTBHC	TWCAC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	1	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMSWMTACGGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	10	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	BHCAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	11	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	GGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	12	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	GGAAAGTBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	13	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	14	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCAGTBHC	TTTGGCGGAGGGACCAAG	CAGTACBHCAGTBHC	TTTGGCGGAGGGACCAAG
VK3-20	15	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACGGAVNCBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	2	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGBHC	GGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	3	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG	CAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG
VK3-20	4	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACGGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	5	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMCAGTACGGAAAGTBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	6	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMBHC	GGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	7	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACBHCAGTAG	TTTGGCGGAGGGACCAAG	CAGTACBHCAGTAG	TTTGGCGGAGGGACCAAG
VK3-20	8	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACGGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	9	1	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGSWMTACGGAAAGTBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	1	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	SWMSWMTACGGAAAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	10	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	BHCAGTAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	11	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	GGAVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	12	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	GGAAAGTBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	13	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCVNCAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	14	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACBHCAGTBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG
VK3-20	15	2	8	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGTACGGAVNCBHC	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG	TTTGGCGGAGGGACCAAG

VK3 - 20	2	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAGBHC	BHCGGAAGTAGT	WGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHC	BHCGGAAGTAGT	WGGACT								
VK3 - 20	3	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAGTAC	BHCAGTAGT	WGGACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	BHCAGTAGT	WGGACT								
VK3 - 20	4	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAGTAC	BHCGAVNCAGT	WGGACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	BHCGAVNCAGT	WGGACT								
VK3 - 20	5	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAGTAC	BHCGAAGTBHC	WGGACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	BHCGAAGTBHC	WGGACT								
VK3 - 20	6	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	BHCGGAAGTAGT	WGGACTTTTGGCGGAGGGACCAAG	CAGSWM	BHCGGAAGTAGT	WGGACT								
VK3 - 20	7	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACBHCAGTAGT	WGGACTTTTGGCGGAGGGACCAAG	CAGSWM	TACBHCAGTAGT	WGGACT								
VK3 - 20	8	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACGGAVNCAGT	WGGACTTTTGGCGGAGGGACCAAG	CAGSWM	TACGGAVNCAGT	WGGACT								
VK3 - 20	9	2	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACGGAAAGTBHC	WGGACTTTTGGCGGAGGGACCAAG	CAGSWM	TACGGAAAGTBHC	WGGACT								
VK3 - 20	1	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMSWM	TACGGAAAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	SWMSWM	TACGGAAAGTAGT	CCTACT								
VK3 - 20	10	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	BHCBHCAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCBHCAGTAGT	CCTACT								
VK3 - 20	11	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	BHCGGAVNCAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCGGAVNCAGT	CCTACT								
VK3 - 20	12	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	BHCGGAAGTBHC	CCTACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCGGAAGTBHC	CCTACT								
VK3 - 20	13	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	TACBHCVNCAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGCAG	TACBHCVNCAGT	CCTACT								
VK3 - 20	14	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	TACBHCAGTBHCCC	TACTTTTGGCGGAGGGACCAAG	CAGCAG	TACBHCAGTBHCCC	TACT								
VK3 - 20	15	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGCAG	TACGGAVNCBHCCC	TACTTTTGGCGGAGGGACCAAG	CAGCAG	TACGGAVNCBHCCC	TACT								
VK3 - 20	2	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAG	BHCGGAAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	SWMCAG	BHCGGAAGTAGT	CCTACT								
VK3 - 20	3	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAG	TACBHCAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	SWMCAG	TACBHCAGTAGT	CCTACT								
VK3 - 20	4	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAG	TACGGAVNCAGT	CCTACTTTTGGCGGAGGGACCAAG	SWMCAG	TACGGAVNCAGT	CCTACT								
VK3 - 20	5	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	SWMCAG	TACGGAAAGTBHCCC	TACTTTTGGCGGAGGGACCAAG	SWMCAG	TACGGAAAGTBHCCC	TACT								
VK3 - 20	6	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	BHCGGAAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGSWM	BHCGGAAGTAGT	CCTACT								
VK3 - 20	7	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACBHCAGTAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGSWM	TACBHCAGTAGT	CCTACT								
VK3 - 20	8	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACGGAVNCAGT	CCTACTTTTGGCGGAGGGACCAAG	CAGSWM	TACGGAVNCAGT	CCTACT								
VK3 - 20	9	3	8	CCTGAAGATTTTGCAGTGTA	TACTGT	CAGSWM	TACGGAAAGTBHCCC	TACTTTTGGCGGAGGGACCAAG	CAGSWM	TACGGAAAGTBHCCC	TACT								
Jumping Trimer																			
VK1 - 05	t1	0	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCVRMMBCTA	CTWC	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCVRMMBCTA	CTWC	ACT						
VK1 - 05	t1	1	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCVRMMBCTA	CMT	CAC	TACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCVRMMBCTA	CMT	CA	CT				
VK1 - 05	t1	2	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCVRMMBCTA	CW	GGA	CTTTTGGCGGAGGGACCAAG	CAGCAG	BHCVRMMBCTA	CW	GGA	CT				
VK1 - 05	t1	3	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCVRMMBCTA	CY	CTA	CTTTTGGCGGAGGGACCAAG	CAGCAG	BHCVRMMBCTA	CY	CTA	CT				
VK1 - 05	t2	0	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCAA	TMBC	YWC	TWC	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCAA	TMBC	YWC	TWC	ACT		
VK1 - 05	t2	1	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCAA	TMBC	YWC	MT	C	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCAA	TMBC	YWC	MT	C	ACT
VK1 - 05	t2	2	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCAA	TMBC	YWC	WGG	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCAA	TMBC	YWC	WGG	ACT		
VK1 - 05	t2	3	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCAA	TMBC	YWC	YCT	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCAA	TMBC	YWC	YCT	ACT		
VK1 - 05	t3	0	8	CCTGATGATTTTGCAACTTA	TACTGC	CAGCAG	BHCVRMAGTY	WC	TWC	ACTTTTGGCGGAGGGACCAAG	CAGCAG	BHCVRMAGTY	WC	TWC	ACT				

Page 145 of 420

Page 146 of 420

VK2-28	t3	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNACTCSRMVB	CCCCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMVB	CCCCCTACT
VK2-28	t4	0	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNASRMVB	CTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNASRMVB	CTWCACCT
VK2-28	t4	1	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNASRMVB	CMTCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNASRMVB	CMTCACCT
VK2-28	t4	2	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNASRMVB	CGGACCTTTTGGCGGAGGGACCAAG	ATGCAGVNASRMVB	CGGACT
VK2-28	t4	3	8	GCTGAGGATGTTGGGGTTTATTACTG	CATGCAGVNASRMVB	CCCCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNASRMVB	CCCCCTACT
VK3-11	t1	0	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGTWCACCT	
VK3-11	t1	1	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGMTCACT	
VK3-11	t1	2	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCTGGWGGACCTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGWGGACT	
VK3-11	t1	3	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGCCTACT	
VK3-11	t2	0	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	TWCACTTTTGGCGGAGGGACCAAG	
VK3-11	t2	1	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	MTCACTTTTGGCGGAGGGACCAAG	
VK3-11	t2	2	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	WGGACCTTTTGGCGGAGGGACCAAG	
VK3-11	t2	3	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	CCCCCTACTTTTGGCGGAGGGACCAAG	
VK3-11	t3	0	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	TWCACTTTTGGCGGAGGGACCAAG	
VK3-11	t3	1	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	MTCACTTTTGGCGGAGGGACCAAG	
VK3-11	t3	2	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	WGGACCTTTTGGCGGAGGGACCAAG	
VK3-11	t3	3	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	CCCCCTACTTTTGGCGGAGGGACCAAG	
VK3-11	t4	0	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	TWCACTTTTGGCGGAGGGACCAAG	
VK3-11	t4	1	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	MTCACTTTTGGCGGAGGGACCAAG	
VK3-11	t4	2	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	WGGACCTTTTGGCGGAGGGACCAAG	
VK3-11	t4	3	8	CCTGAAAGATTTTGCAGTTTATTACT	GTG	CAGCBHCHCNHCAATYWC	CCCCCTACTTTTGGCGGAGGGACCAAG	
VK3-15	t1	0	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGTWCACCT	
VK3-15	t1	1	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGMTCACT	
VK3-15	t1	2	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCTGGWGGACCTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGWGGACT	
VK3-15	t1	3	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCTGGCCTACTTTTGGCGGAGGGACCAAG	CAGCBHCHCNHCTGGCCTACT	
VK3-15	t2	0	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCAATYWC	TWCACTTTTGGCGGAGGGACCAAG	
VK3-15	t2	1	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCAATYWC	MTCACTTTTGGCGGAGGGACCAAG	
VK3-15	t2	2	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCAATYWC	WGGACCTTTTGGCGGAGGGACCAAG	
VK3-15	t2	3	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCNHCAATYWC	CCCCCTACTTTTGGCGGAGGGACCAAG	
VK3-15	t3	0	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCAATNHC	YWC	TWCACTTTTGGCGGAGGGACCAAG
VK3-15	t3	1	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCAATNHC	YWC	MTCACTTTTGGCGGAGGGACCAAG
VK3-15	t3	2	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCAATNHC	YWC	WGGACCTTTTGGCGGAGGGACCAAG
VK3-15	t3	3	8	CAGTCTGAAGATTTTTCAGTTTATT	CTG	TCACTGTCAGCBHCHCAATNHC	YWC	CCCCCTACTTTTGGCGGAGGGACCAAG

VK3-15	t4	0	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHCYWC	TWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWC	TWCACT
VK3-15	t4	1	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHCYWC	MTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWC	MTCACT
VK3-15	t4	2	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHCYWC	GGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWC	GGACT
VK3-15	t4	3	8	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHCYWC	CCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWC	CCCTACT
VK3-20	t1	0	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCVNCAGTTWC	ACTTTTGGCGGAGGGACCAAG	TTWC
VK3-20	t1	1	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCVNCAGTMTCA	CTTTTGGCGGAGGGACCAAG	TMTC
VK3-20	t1	2	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCVNCAGTWGGA	CTTTTGGCGGAGGGACCAAG	TTWG
VK3-20	t1	3	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCVNCAGTCC	TACTTTTGGCGGAGGGACCAAG	TCTACT
VK3-20	t2	0	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCAGTBHC	TWCACTTTTGGCGGAGGGACCAAG	TTWCACT
VK3-20	t2	1	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCAGTBHC	MTCACTTTTGGCGGAGGGACCAAG	MTCACT
VK3-20	t2	2	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCAGTBHC	WGGACTTTTGGCGGAGGGACCAAG	WGGACT
VK3-20	t2	3	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	BCAGTBHC	CCCTACTTTTGGCGGAGGGACCAAG	CCCTACT
VK3-20	t3	0	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	CGGAVNCBHC	TWCACTTTTGGCGGAGGGACCAAG	TWCACT
VK3-20	t3	1	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	CGGAVNCBHC	MTCACTTTTGGCGGAGGGACCAAG	MTCACT
VK3-20	t3	2	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	CGGAVNCBHC	WGGACTTTTGGCGGAGGGACCAAG	WGGACT
VK3-20	t3	3	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGBHC	CGGAVNCBHC	CCCTACTTTTGGCGGAGGGACCAAG	CCCTACT
VK3-20	t4	0	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGTAC	BCVNCBHC	TWCACTTTTGGCGGAGGGACCAAG	TWCACT
VK3-20	t4	1	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGTAC	BCVNCBHC	MTCACTTTTGGCGGAGGGACCAAG	MTCACT
VK3-20	t4	2	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGTAC	BCVNCBHC	WGGACTTTTGGCGGAGGGACCAAG	WGGACT
VK3-20	t4	3	8	CCTGAAAGATTTTGCAGTGTTACTGT	CAGCAGTAC	BCVNCBHC	CCCTACTTTTGGCGGAGGGACCAAG	CCCTACT
VK4-01	t1	0	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCNHC	ACTTWCACTTTTGGCGGAGGGACCAAG	CTTWCACT
VK4-01	t1	1	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCNHC	ACTMTCACTTTTGGCGGAGGGACCAAG	CTMTCACT
VK4-01	t1	2	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCNHC	ACTWGGACTTTTGGCGGAGGGACCAAG	CTWGGACT
VK4-01	t1	3	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCNHC	ACTCC	TACTTTTGGCGGAGGGACCAAG
VK4-01	t2	0	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCAGTBHC	TWCACTTTTGGCGGAGGGACCAAG	TWCACT
VK4-01	t2	1	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCAGTBHC	MTCACTTTTGGCGGAGGGACCAAG	MTCACT
VK4-01	t2	2	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCAGTBHC	WGGACTTTTGGCGGAGGGACCAAG	WGGACT
VK4-01	t2	3	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	BCAGTBHC	CCCTACTTTTGGCGGAGGGACCAAG	CCCTACT
VK4-01	t3	0	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	CTACNHC	BC	TWCACT
VK4-01	t3	1	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	CTACNHC	BC	MTCACT
VK4-01	t3	2	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	CTACNHC	BC	WGGACT
VK4-01	t3	3	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGBHC	CTACNHC	BC	CCCTACT
VK4-01	t4	0	8	GCTGAAAGATGTGGCAGTTTATTACTGT	CAGCAGTAC	BCNHC	BC	TWCACT

VK4-01	t4	1	8	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBHCMTCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBHCMTCAC
VK4-01	t4	2	8	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBHCMTCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBHCMTCAC
VK4-01	t4	3	8	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBHCMTCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBHCMTCAC

Table 6. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDRL3 length 9.

	Name	Sequence of Synthesized Oligonucleotide										Portion of Oligonucleotide Corresponding to CDRL3 Proper
Jumping Dimer												
	VK1 - 05 1 0 9	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAAATAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAGTTAC	YCTTWCAC
	VK1 - 05 10 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTAC	YCTTWCAC
	VK1 - 05 11 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATMBC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATMBC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATMBC	TAC	YCTTWCAC
	VK1 - 05 12 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAG	TYWC	YCTTWCAC
	VK1 - 05 13 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBC	TAC	YCTTWCAC
	VK1 - 05 14 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAG	TYWC	YCTTWCAC
	VK1 - 05 15 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAAATM	BCYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACAAATM	BCYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGCAGTACAAATM	BCYWC	YCTTWCAC
	VK1 - 05 2 0 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAAATAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAG	TTAC	YCTTWCAC
	VK1 - 05 3 0 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAG	TTAC	YCTTWCAC
	VK1 - 05 4 0 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATM	BC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATM	BC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATM
	VK1 - 05 5 0 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAG	TYWC	YCTTWCAC
	VK1 - 05 6 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAAATAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMBHCAAATAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMBHCAAATAG	TTAC	YCTTWCAC
	VK1 - 05 7 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACVRMAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAG	TTAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAG	TTAC	YCTTWCAC
	VK1 - 05 8 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATM	BC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACAAATM	BC	TAC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACAAATM
	VK1 - 05 9 0 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAG	TYWC	YCTTWCAC	TTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAG	TYWC	YCTTWCAC
	VK1 - 05 1 1 9	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAG	TTAC	YCTMTCA
	VK1 - 05 10 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAG	TTAC	YCTMTCA
	VK1 - 05 11 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATM	BC	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATM	BC	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATM
	VK1 - 05 12 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAG	TYWC	YCTMTCA
	VK1 - 05 13 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMB	C	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMB	C	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMB
	VK1 - 05 14 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAG	TYWC	YCTMTCA
	VK1 - 05 15 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAAATM	BCYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACAAATM	BCYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGCAGTACAAATM	BCYWC	YCTMTCA
	VK1 - 05 2 1 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAG	TTAC	YCTMTCA
	VK1 - 05 3 1 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAG	TTAC	YCTMTCA
	VK1 - 05 4 1 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATM	BC	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATM	BC	TAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATM
	VK1 - 05 5 1 9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAG	TYWC	YCTMTCA	TTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAG	TYWC	YCTMTCA
	VK1 - 05 6 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGSWMBHCAAATAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGSWMBHCAAATAG	TTAC	YCTMTCA
	VK1 - 05 7 1 9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAG	TTAC	YCTMTCA	TTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAG	TTAC	YCTMTCA

VK1-05	8	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATMBCTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATMBCTACYCTMTCACT
VK1-05	9	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATAGTYWCYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAGTYWCYCTMTCACT
VK1-05	1	2	9	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAAATAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAGTTACYCTWGGACT
VK1-05	10	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACYCTWGGACT
VK1-05	11	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATMBCTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATMBCTACYCTWGGACT
VK1-05	12	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATAGTYWCYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAGTYWCYCTWGGACT
VK1-05	13	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTACYCTWGGACT
VK1-05	14	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAGTYWCYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYCTWGGACT
VK1-05	15	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAAATMBCYWCYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACAAATMBCYWCYCTWGGACT
VK1-05	2	2	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAAATAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAGTTACYCTWGGACT
VK1-05	3	2	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYCTWGGACT
VK1-05	4	2	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATMBCTACYCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATMBCTACYCTWGGACT
VK1-05	5	2	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATAGTYWCYCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAGTYWCYCTWGGACT
VK1-05	6	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAATAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHCAATAGTTACYCTWGGACT
VK1-05	7	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACVRMAGTTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTTACYCTWGGACT
VK1-05	8	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATMBCTACYCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATMBCTACYCTWGGACT
VK1-05	9	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATAGTYWCYCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAGTYWCYCTWGGACT
VK1-05	1	3	9	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAAATAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAGTTACYCTCCTACT
VK1-05	10	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACYCTCCTACT
VK1-05	11	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATMBCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATMBCTACYCTCCTACT
VK1-05	12	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCAAATAGTYWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCAAATAGTYWCYCTCCTACT
VK1-05	13	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTACYCTCCTACT
VK1-05	14	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAGTYWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYCTCCTACT
VK1-05	15	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAAATMBCYWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACAAATMBCYWCYCTCCTACT
VK1-05	2	3	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAAATAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGBHCAAATAGTTACYCTCCTACT
VK1-05	3	3	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYCTCCTACT
VK1-05	4	3	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATMBCTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATMBCTACYCTCCTACT
VK1-05	5	3	9	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATAGTYWCYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATAGTYWCYCTCCTACT
VK1-05	6	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCAATAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMBHCAATAGTTACYCTCCTACT
VK1-05	7	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACVRMAGTTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTTACYCTCCTACT
VK1-05	8	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATMBCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATMBCTACYCTCCTACT
VK1-05	9	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACAAATAGTYWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAGTYWCYCTCCTACT
VK1-12	1	0	9	CCTGAAGATTTTGCAACTTATTACTGTSWMSWMC AAAATAGTTTCCCTTWCACTTTTGGCGGAGGACCAAG	SWMSWMC AAAATAGTTTCCCTTWCACT

WO 2012/009568

PCT/US2011/044063

VK1 -12 12 10 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCAGTTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTTTCCCTTWCACT
VK1 -12 12 11 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATNHCTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATNHCTTCCCTTWCACT
VK1 -12 12 12 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATAGTYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATAGTYWCCCTTWCACT
VK1 -12 12 13 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCTTCCCTTWCACT
VK1 -12 12 14 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAGTYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCCCTTWCACT
VK1 -12 12 15 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATNHCYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCYWCCCTTWCACT
VK1 -12 12 2 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATAGTTTCCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGRNAAAATAGTTTCCCTTWCACT
VK1 -12 12 3 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAGTTTCCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTTWCACT
VK1 -12 12 4 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATNHCTTCCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATNHCTTCCCTTWCACT
VK1 -12 12 5 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATAGTYWCCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYWCCCTTWCACT
VK1 -12 12 6 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMRNAAAATAGTTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMRNAAAATAGTTTCCCTTWCACT
VK1 -12 12 7 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCANHCAGTTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTTWCACT
VK1 -12 12 8 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATNHCTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAATNHCTTCCCTTWCACT
VK1 -12 12 9 0 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATAGTYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAATAGTYWCCCTTWCACT
VK1 -12 12 1 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMSWMGCAAAATAGTTTCCCTMTCACT
VK1 -12 12 10 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTTTCCCTMTCACT
VK1 -12 12 11 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATNHCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATNHCTTCCCTMTCACT
VK1 -12 12 12 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATAGTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATAGTYWCCCTMTCACT
VK1 -12 12 13 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCTTCCCTMTCACT
VK1 -12 12 14 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAGTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCCCTMTCACT
VK1 -12 12 15 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATNHCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCYWCCCTMTCACT
VK1 -12 2 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMCAGRNAAAATAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGRNAAAATAGTTTCCCTMTCACT
VK1 -12 3 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCANHCAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTMTCACT
VK1 -12 4 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATNHCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATNHCTTCCCTMTCACT
VK1 -12 5 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATAGTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYWCCCTMTCACT
VK1 -12 6 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMRNAAAATAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTMTCACT
VK1 -12 7 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCANHCAGTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTMTCACT
VK1 -12 8 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATNHCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAATNHCTTCCCTMTCACT
VK1 -12 9 1 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATAGTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAATAGTYWCCCTMTCACT
VK1 -12 1 2 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGSWMGCAAAATAGTTTCCCTTWGGACTTTTGGCGGAGGGACCAAG	SWMSWMGCAAAATAGTTTCCCTTWGGACT
VK1 -12 10 2 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCAGTTTCCCTTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTTTCCCTTWGGACT
VK1 -12 11 2 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATNHCTTCCCTTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATNHCTTCCCTTWGGACT
VK1 -12 12 2 9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAAATAGTYWCCCTTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATAGTYWCCCTTWGGACT

VK1 - 12 13 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCANHCNHC	TTCCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCANHCNHC	TTCCCTWGGACT
VK1 - 12 14 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCANHCAGTYW	CCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYW	CCCTWGGACT
VK1 - 12 15 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCAAAATNH	CYCCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCAAAATNH	CYCCCTWGGACT
VK1 - 12 2 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGR	NAAATAGTTTCCCTWGGACTTTT	TGGCGGAGGGACCAAG	SWMCAGR	NAAATAGTTTCCCTWGGACT
VK1 - 12 3 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCANHCAGTTT	CCCTWGGACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTT	CCCTWGGACT
VK1 - 12 4 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCAAAATNH	CTTCCCTWGGACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCAAAATNH	CTTCCCTWGGACT
VK1 - 12 5 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCAAAATAGTYW	CCCTWGGACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYW	CCCTWGGACT
VK1 - 12 6 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMR	NAAATAGTTTCCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGSWMR	NAAATAGTTTCCCTWGGACT
VK1 - 12 7 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCANHCAGTTT	CCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTT	CCCTWGGACT
VK1 - 12 8 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCAAAATNH	CTTCCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCAAAATNH	CTTCCCTWGGACT
VK1 - 12 9 2 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCAAAATAGTYW	CCCTWGGACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCAAAATAGTYW	CCCTWGGACT
VK1 - 12 1 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMSWMGCAAAATAGTTT	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	SWMSWMGCAAAATAGTTT	CCCTCCTACT
VK1 - 12 10 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGR	NANHCAGTTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGR	NANHCAGTTTCCCTCCTACT
VK1 - 12 11 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGR	NAAATNHCTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGR	NAAATNHCTTCCCTCCTACT
VK1 - 12 12 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGR	NAAATAGTYWCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGR	NAAATAGTYWCCCTCCTACT
VK1 - 12 13 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCANHCNHC	TTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCANHCNHC	TTCCCTCCTACT
VK1 - 12 14 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCANHCAGTYW	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYW	CCCTCCTACT
VK1 - 12 15 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGCAGGCAAAATNH	CYCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGCAGGCAAAATNH	CYCCCTCCTACT
VK1 - 12 2 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGR	NAAATAGTTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	SWMCAGR	NAAATAGTTTCCCTCCTACT
VK1 - 12 3 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCANHCAGTTT	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTT	CCCTCCTACT
VK1 - 12 4 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCAAAATNH	CTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCAAAATNH	CTTCCCTCCTACT
VK1 - 12 5 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	SWMCAGGCAAAATAGTYW	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYW	CCCTCCTACT
VK1 - 12 6 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMR	NAAATAGTTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGSWMR	NAAATAGTTTCCCTCCTACT
VK1 - 12 7 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCANHCAGTTT	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTT	CCCTCCTACT
VK1 - 12 8 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCAAAATNH	CTTCCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCAAAATNH	CTTCCCTCCTACT
VK1 - 12 9 3 9	CCTGAAGATTTTGC	AACCTTATTACTGT	CAGSWMGCAAAATAGTYW	CCCTCCTACTTTT	TGGCGGAGGGACCAAG	CAGSWMGCAAAATAGTYW	CCCTCCTACT
VK1 - 33 1 0 9	CCTGAAGATATTGCA	ACATATTACTGT	SWMSWMT	ACGATAATCTCCCTTWCACTTTT	TGGCGGAGGGACCAAG	SWMSWMT	ACGATAATCTCCCTTWCACT
VK1 - 33 10 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGBHCNHC	AAATCTCCCTTWCACTTTT	TGGCGGAGGGACCAAG	CAGCAGBHCNHC	AAATCTCCCTTWCACT
VK1 - 33 11 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGBHC	GATNHCCCTCCCTTWCACTTTT	TGGCGGAGGGACCAAG	CAGCAGBHC	GATNHCCCTCCCTTWCACT
VK1 - 33 12 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGBHC	GATAATATYWCCCTTWCACTTTT	TGGCGGAGGGACCAAG	CAGCAGBHC	GATAATATYWCCCTTWCACT
VK1 - 33 13 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGTACNHC	NCCTCCCTTWCACTTTT	TGGCGGAGGGACCAAG	CAGCAGTACNHC	NCCTCCCTTWCACT
VK1 - 33 14 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGTACNHC	AAATYWCCCTTWCACTTTT	TGGCGGAGGGACCAAG	CAGCAGTACNHC	AAATYWCCCTTWCACT
VK1 - 33 15 0 9	CCTGAAGATATTGCA	ACATATTACTGT	CAGCAGTACNHC	AGTATAGTTT	TGGCGGAGGGACCAAG	CAGCAGTACNHC	AGTATAGTTT

VK1 - 33	2	0	9	CCTGAAGATATTGCAACATATTACTGTSWMCAGBHC	GATAAATCTCCCTTWCACTTTTGGCGGAGGACCAAG	SWMCAGBHC	GATAAATCTCCCTTWCACT		
VK1 - 33	3	0	9	CCTGAAGATATTGCAACATATTACTGTSWMCAGTAC	NHCAATCTCCCTTWCACTTTTGGCGGAGGACCAAG	SWMCAGTAC	NHCAATCTCCCTTWCACT		
VK1 - 33	4	0	9	CCTGAAGATATTGCAACATATTACTGTSWMCAGTAC	GATNHCCCTCCCTTWCACTTTTGGCGGAGGACCAAG	SWMCAGTAC	GATNHCCCTCCCTTWCACT		
VK1 - 33	5	0	9	CCTGAAGATATTGCAACATATTACTGTSWMCAGTAC	GATAATYWCCCTTWCACTTTTGGCGGAGGACCAAG	SWMCAGTAC	GATAATYWCCCTTWCACT		
VK1 - 33	6	0	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	BHC	GATAAATCTCCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	BHC	GATAAATCTCCCTTWCACT
VK1 - 33	7	0	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	NHCAATCTCCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	NHCAATCTCCCTTWCACT
VK1 - 33	8	0	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	GATNHCCCTCCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	GATNHCCCTCCCTTWCACT
VK1 - 33	9	0	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	GATAATYWCCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	GATAATYWCCCTTWCACT
VK1 - 33	1	1	9	CCTGAAGATATTGCAACATATTACTGTSWMSWM	TAC	GATAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	SWMSWM	TAC	GATAATCTCCCTMTCACT
VK1 - 33	10	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	NHCAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	NHCAATCTCCCTMTCACT
VK1 - 33	11	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	GATNHCCCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	GATNHCCCTCCCTMTCACT
VK1 - 33	12	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	GATAATYWCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	GATAATYWCCCTMTCACT
VK1 - 33	13	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCCCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCCCTCCCTMTCACT
VK1 - 33	14	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCAATYWCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCAATYWCCCTMTCACT
VK1 - 33	15	1	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCCCTMTCACT
VK1 - 33	2	1	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	BHC	GATAAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAG	BHC	GATAAATCTCCCTMTCACT
VK1 - 33	3	1	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	TAC	NHCAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAG	TAC	NHCAATCTCCCTMTCACT
VK1 - 33	4	1	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	TAC	NHCCCTCCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAG	TAC	NHCCCTCCCTMTCACT
VK1 - 33	5	1	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	TAC	GATAATYWCCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAG	TAC	GATAATYWCCCTMTCACT
VK1 - 33	6	1	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	BHC	GATAAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWM	BHC	GATAAATCTCCCTMTCACT
VK1 - 33	7	1	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	NHCAATCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	NHCAATCTCCCTMTCACT
VK1 - 33	8	1	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	GATNHCCCTCCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	GATNHCCCTCCCTMTCACT
VK1 - 33	9	1	9	CCTGAAGATATTGCAACATATTACTGTCAGSWM	TAC	GATAATYWCCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWM	TAC	GATAATYWCCCTMTCACT
VK1 - 33	1	2	9	CCTGAAGATATTGCAACATATTACTGTSWMSWM	TAC	GATAATCTCCCTWGGACTTTTGGCGGAGGACCAAG	SWMSWM	TAC	GATAATCTCCCTWGGACT
VK1 - 33	10	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	NHCAATCTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	NHCAATCTCCCTWGGACT
VK1 - 33	11	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	GATNHCCCTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	GATNHCCCTCCCTWGGACT
VK1 - 33	12	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	BHC	GATAATYWCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	BHC	GATAATYWCCCTWGGACT
VK1 - 33	13	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCCCTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCCCTCCCTWGGACT
VK1 - 33	14	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCAATYWCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCAATYWCCCTWGGACT
VK1 - 33	15	2	9	CCTGAAGATATTGCAACATATTACTGTCAGCAG	TAC	NHCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAG	TAC	NHCCCTWGGACT
VK1 - 33	2	2	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	BHC	GATAAATCTCCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAG	BHC	GATAAATCTCCCTWGGACT
VK1 - 33	3	2	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	TAC	NHCAATCTCCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAG	TAC	NHCAATCTCCCTWGGACT
VK1 - 33	4	2	9	CCTGAAGATATTGCAACATATTACTGTSWMCAG	TAC	GATNHCCCTCCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAG	TAC	GATNHCCCTCCCTWGGACT

VK1 - 33	5	2	9	CCTGAAGATATTGCAACATATTA	CTGTSWMCAGTACGATAA	TYWCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAA	TYWCCCTWGGACT						
VK1 - 33	6	2	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMBHCAGATAA	TCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGATAA	TCCTCCTWGGACT						
VK1 - 33	7	2	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTACNHCAA	TCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAA	TCCTCCTWGGACT						
VK1 - 33	8	2	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTACGATNHCC	CTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCC	CTCCCTWGGACT						
VK1 - 33	9	2	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTACGATAA	TYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAA	TYWCCCTWGGACT						
VK1 - 33	1	3	9	CCTGAAGATATTGCAACATATTA	CTGTSWMSWMTACGATAA	TCCTCCTACTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAA	TCCTCCTCCTACT						
VK1 - 33	10	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGBHCNHCAA	TCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAA	TCCTCCTCCTACT						
VK1 - 33	11	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGBHC	GATNHCCCTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	GATNHCCCTCCCTCCTACT						
VK1 - 33	12	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGBHC	GATAA	TYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	GATAA	TYWCCCTCCTACT				
VK1 - 33	13	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGTAC	NHCCCTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCCCTCCCTCCTACT						
VK1 - 33	14	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGTAC	NHCAA	TYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCAA	TYWCCCTCCTACT				
VK1 - 33	15	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGCAGTAC	GATNHCCYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	GATNHCC	YWCCTCCTACT					
VK1 - 33	2	3	9	CCTGAAGATATTGCAACATATTA	CTGTSWMCAGBHC	GATAA	TCCTCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGBHC	GATAA	TCCTCCTCCTACT				
VK1 - 33	3	3	9	CCTGAAGATATTGCAACATATTA	CTGTSWMCAGTAC	NHCAA	TCCTCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	NHCAA	TCCTCCTCCTACT				
VK1 - 33	4	3	9	CCTGAAGATATTGCAACATATTA	CTGTSWMCAGTAC	GATNHCCCTCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	GATNHCC	CTCCCTCCTACT					
VK1 - 33	5	3	9	CCTGAAGATATTGCAACATATTA	CTGTSWMCAGTAC	GATAA	TYWCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTAC	GATAA	TYWCCCTCCTACT				
VK1 - 33	6	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMBHC	GATAA	TCCTCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMBHC	GATAA	TCCTCCTCCTACT				
VK1 - 33	7	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTAC	NHCAA	TCCTCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	NHCAA	TCCTCCTCCTACT				
VK1 - 33	8	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTAC	GATNHCCCTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	GATNHCC	CTCCCTCCTACT					
VK1 - 33	9	3	9	CCTGAAGATATTGCAACATATTA	CTGTCAGSWMTAC	GATAA	TYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMTAC	GATAA	TYWCCCTCCTACT				
VK1 - 39	1	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTSWMSWMA	GCTACAGTAC	TCCTTWCACTTTTGGCGGAGGGACCAAG	SWMSWMAGCTAC	AGTAC	TCCTTWCACT				
VK1 - 39	10	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	VNABHCAGTAC	TCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	VNABHC	AGTAC	TCCTTWCACT			
VK1 - 39	11	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	VNATA	CNHC	ACTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	VNATA	CNHC	ACTCCTTWCACT		
VK1 - 39	12	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	VNATA	CAGTBH	CCCC	TTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	VNATA	CAGTBH	CCCC	TTWCACT
VK1 - 39	13	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	AAGCBHC	NHC	ACTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	AAGCBHC	NHC	ACTCCTTWCACT		
VK1 - 39	14	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	AAGCBHC	AGTBH	CCC	TTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	AAGCBHC	AGTBH	CCC	TTWCACT
VK1 - 39	15	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGCAA	AAGCTAC	NHCBH	CCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAA	AAGCTAC	NHCBH	CCCTTWCACT		
VK1 - 39	2	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTSWMC	AAVNATACAGTAC	CTCC	TTWCACTTTTGGCGGAGGGACCAAG	SWMCAA	VNATA	CAGTAC	CTCC	TTWCACT	
VK1 - 39	3	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTSWMC	AAAGCBHC	AGTAC	CTCC	TTWCACTTTTGGCGGAGGGACCAAG	SWMCAA	AGCBHC	AGTAC	CTCC	TTWCACT
VK1 - 39	4	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTSWMC	AAAGCTA	CNHC	ACTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAA	AGCTA	CNHC	ACTCCTTWCACT		
VK1 - 39	5	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTSWMC	AAAGCTAC	AGTBH	CCC	TTWCACTTTTGGCGGAGGGACCAAG	SWMCAA	AGCTAC	AGTBH	CCC	TTWCACT
VK1 - 39	6	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGSWM	VNATACAGTAC	CTCC	TTWCACTTTTGGCGGAGGGACCAAG	CAGSWM	VNATAC	AGTAC	CTCC	TTWCACT	
VK1 - 39	7	0	9	CCTGAAGATTTTGCAACTTA	CTACTGTGTCAGSWM	AGCBHC	AGTAC	CTCC	TTWCACTTTTGGCGGAGGGACCAAG	CAGSWM	AGCBHC	AGTAC	CTCC	TTWCACT

VK1 - 39 8 0 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACNHCACTCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTCCTTWCACCTCCTTWCACCT
VK1 - 39 9 0 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACAGTBHCCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCCCTTWCACCT
VK1 - 39 1 1 9	CCTGAAGATTTTGCAACTTACTACTGTSWMSWMAGCTACAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTCCTMTCACT
VK1 - 39 10 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNABHCAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTACTCCTMTCACT
VK1 - 39 11 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNATACNHCACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCACTCCTMTCACT
VK1 - 39 12 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNATACAGTBHCCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACAGTBHCCCCTMTCACT
VK1 - 39 13 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCBHCNHCACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCNHCACTCCTMTCACT
VK1 - 39 14 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCBHCAGTBHCCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCAGTBHCCCCTMTCACT
VK1 - 39 15 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCTACNHCBHCCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCTACNHCBHCCCCTMTCACT
VK1 - 39 2 1 9	CCTGAAGATTTTGCAACTTACTACTGTSWMCAAVNATACAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAAVNATACAGTACTCCTMTCACT
VK1 - 39 3 1 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAAGCBHCAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCMAAGCBHCAGTACTCCTMTCACT
VK1 - 39 4 1 9	CCTGAAGATTTTGCAACTTACTACTGTSWMAAGCTACNHCACTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMAAGCTACNHCACTCCTMTCACT
VK1 - 39 5 1 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAAGCTACAGTBHCCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMAAGCTACAGTBHCCCCTMTCACT
VK1 - 39 6 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMVNATACAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTCCTMTCACT
VK1 - 39 7 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCBHCAGTACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCAGTACTCCTMTCACT
VK1 - 39 8 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACNHCACTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTCCTMTCACT
VK1 - 39 9 1 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACAGTBHCCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCCCTMTCACT
VK1 - 39 1 2 9	CCTGAAGATTTTGCAACTTACTACTGTSMWSWMAGCTACAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTCCTWGGACT
VK1 - 39 10 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNABHCAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTACTCCTWGGACT
VK1 - 39 11 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNATACNHCACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCACTCCTWGGACT
VK1 - 39 12 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAVNATACAGTBHCCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNATACAGTBHCCCCTWGGACT
VK1 - 39 13 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCBHCNHCACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCNHCACTCCTWGGACT
VK1 - 39 14 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCBHCAGTBHCCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCBHCAGTBHCCCCTWGGACT
VK1 - 39 15 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGCAAAAGCTACNHCBHCCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAAAGCTACNHCBHCCCCTWGGACT
VK1 - 39 2 2 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAVNATACAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCMAVNATACAGTACTCCTWGGACT
VK1 - 39 3 2 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAAGCBHCAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCMAAGCBHCAGTACTCCTWGGACT
VK1 - 39 4 2 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAAGCTACNHCACTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCMAAGCTACNHCACTCCTWGGACT
VK1 - 39 5 2 9	CCTGAAGATTTTGCAACTTACTACTGTSMWMAAGCTACAGTBHCCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCMAAGCTACAGTBHCCCCTWGGACT
VK1 - 39 6 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMVNATACAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTCCTWGGACT
VK1 - 39 7 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCBHCAGTACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCAGTACTCCTWGGACT
VK1 - 39 8 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACNHCACTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTCCTWGGACT
VK1 - 39 9 2 9	CCTGAAGATTTTGCAACTTACTACTGTAGSWMAGCTACAGTBHCCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCCCTWGGACT
VK1 - 39 1 3 9	CCTGAAGATTTTGCAACTTACTACTGTSMWSWMAGCTACAGTACTCCTCCTACTTTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTCCTCCTACT

VK1 - 39	10	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAVNABHCAGTACTCCTCCTACTTTTGGCGGAGGGACCCAAG	CAGCAAVNABHCAGTACTCCTCCTACT
VK1 - 39	11	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAVNATAACNHCAGTACTCCTCCTACTTTTGGCGGAGGGACCCAAG	CAGCAAVNATAACNHCAGTACTCCTCCTACT
VK1 - 39	12	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAVNATAACAGTBHCCCTCCTACTACTTTTGGCGGAGGGACCCAAG	CAGCAAVNATAACAGTBHCCCTCCTACT
VK1 - 39	13	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAAAGCBHCNHCAGTCTCCTCCTACTCTTTTGGCGGAGGGACCCAAG	CAGCAAAAGCBHCNHCAGTCTCCTCCTACT
VK1 - 39	14	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAAAGCBHCAGTBHCCCTCCTACTACTTTTGGCGGAGGGACCCAAG	CAGCAAAAGCBHCAGTBHCCCTCCTACT
VK1 - 39	15	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGCAAAAGCTACNHCBHCCCTCCTACTACTTTTGGCGGAGGGACCCAAG	CAGCAAAAGCTACNHCBHCCCTCCTACT
VK1 - 39	2	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTSWMCAAVNATAACAGTACTCCTCCTACTTTTGGCGGAGGGACCCAAG	SWMCAAVNATAACAGTACTCCTCCTACT
VK1 - 39	3	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTSWMCAAAGCBHCAGTACTCCTCCTACTTTTGGCGGAGGGACCCAAG	SWMCAAAAGCBHCAGTACTCCTCCTACT
VK1 - 39	4	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTSWMCAAAGCTACNHCAGTCTCCTCCTACTTTTGGCGGAGGGACCCAAG	SWMCAAAAGCTACNHCAGTCTCCTCCTACT
VK1 - 39	5	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTSWMCAAAGCTACAGTBHCCCTCCTACTACTTTTGGCGGAGGGACCCAAG	SWMCAAAAGCTACAGTBHCCCTCCTACT
VK1 - 39	6	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGSWMVNATAACAGTACTCCTCCTACTCTTTTGGCGGAGGGACCCAAG	CAGSWMVNATAACAGTACTCCTCCTACT
VK1 - 39	7	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGSWMAGCBHCAGTACTCCTCCTACTACTTTTGGCGGAGGGACCCAAG	CAGSWMAGCBHCAGTACTCCTCCTACT
VK1 - 39	8	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGSWMAGCTACNHCAGTCTCCTCCTACTTTTGGCGGAGGGACCCAAG	CAGSWMAGCTACNHCAGTCTCCTCCTACT
VK1 - 39	9	3	9	CCTGAAGATTTTGCAACTTACTACTGTGTCAGSWMAGCTACAGTBHCCCTCCTACTACTTTTGGCGGAGGGACCCAAG	CAGSWMAGCTACAGTBHCCCTCCTACT
VK2 - 28	1	0	9	GCTGAGGATGTTGGGGTTTATTACTGTGCDTSSWMGCACTCCAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	DTSSWMGCACTCCAGACTCCTTWCACT
VK2 - 28	10	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNACAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGVNAMNACAGACTCCTTWCACT
VK2 - 28	11	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCSRMACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGVNACTCSRMACTCCTTWCACT
VK2 - 28	12	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCCAGVBCCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGVNACTCCAGVBCCCTTWCACT
VK2 - 28	13	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCAMNASRMACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGGCAMNASRMACTCCTTWCACT
VK2 - 28	14	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCAMNACAGVBCCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGGCAMNACAGVBCCCTTWCACT
VK2 - 28	15	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCACTCSRMBVCCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGCAGGCACTCSRMBVCCCTTWCACT
VK2 - 28	2	0	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	DTSCAGVNACTCCAGACTCCTTWCACT
VK2 - 28	3	0	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	DTSCAGGCAMNACAGACTCCTTWCACT
VK2 - 28	4	0	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSRMACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	DTSCAGGCACCTCSRMACTCCTTWCACT
VK2 - 28	5	0	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCAGVBCCCTTWCACCTTTTGGCGGAGGGACCCAAG	DTSCAGGCACCTCCAGVBCCCTTWCACT
VK2 - 28	6	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGSWMVNACTCCAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGSWMVNACTCCAGACTCCTTWCACT
VK2 - 28	7	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGSWMGCAMNACAGACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGSWMGCAMNACAGACTCCTTWCACT
VK2 - 28	8	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGSWMGCACCTCSRMACTCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGSWMGCACCTCSRMACTCCTTWCACT
VK2 - 28	9	0	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGSWMGCACCTCCAGVBCCCTTWCACCTTTTGGCGGAGGGACCCAAG	ATGSWMGCACCTCCAGVBCCCTTWCACT
VK2 - 28	1	1	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACCTCCAGACTCCTMTACCTTTTGGCGGAGGGACCCAAG	DTSSWMGCACCTCCAGACTCCTMTCACT
VK2 - 28	10	1	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNACAGACTCCTMTACCTTTTGGCGGAGGGACCCAAG	ATGCAGVNAMNACAGACTCCTMTCACT
VK2 - 28	11	1	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCSRMACTCCTMTACCTTTTGGCGGAGGGACCCAAG	ATGCAGVNACTCSRMACTCCTMTCACT
VK2 - 28	12	1	9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCCAGVBCCCTMTCACTTTTGGCGGAGGGACCCAAG	ATGCAGVNACTCCAGVBCCCTMTCACT

WO 2012/009568

PCT/US2011/044063

VK2 - 28	13	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMACTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTMTCACT
VK2 - 28	14	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNACAGVBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCTMTCACT
VK2 - 28	15	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCACCTCSRMBVBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCACCTCSRMBVBCCCTMTCACT
VK2 - 28	2	1	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTMTCACT
VK2 - 28	3	1	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTMTCACT
VK2 - 28	4	1	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSRMACTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCSRMACTCCTMTCACT
VK2 - 28	5	1	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCAGVBCCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCCAGVBCCCTMTCACT
VK2 - 28	6	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMVNACTCCAGACTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSWMVNACTCCAGACTCCTMTCACT
VK2 - 28	7	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCAMNACAGACTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSWMGCAMNACAGACTCCTMTCACT
VK2 - 28	8	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCACCTCSRMACTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSWMGCACCTCSRMACTCCTMTCACT
VK2 - 28	9	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCACCTCCAGVBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSWMGCACCTCCAGVBCCCTMTCACT
VK2 - 28	1	2	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACCTCCAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSSWMGCACCTCCAGACTCCTWGGACT
VK2 - 28	10	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNAMNACAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCTWGGACT
VK2 - 28	11	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCSRMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCTWGGACT
VK2 - 28	12	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCCAGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCTWGGACT
VK2 - 28	13	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTWGGACT
VK2 - 28	14	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNACAGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCTWGGACT
VK2 - 28	15	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCACCTCSRMBVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCACCTCSRMBVBCCCTWGGACT
VK2 - 28	2	2	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTWGGACT
VK2 - 28	3	2	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTWGGACT
VK2 - 28	4	2	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSRMACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCSRMACTCCTWGGACT
VK2 - 28	5	2	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCAGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCCAGVBCCCTWGGACT
VK2 - 28	6	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMVNACTCCAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSWMVNACTCCAGACTCCTWGGACT
VK2 - 28	7	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCAMNACAGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSWMGCAMNACAGACTCCTWGGACT
VK2 - 28	8	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCACCTCSRMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSWMGCACCTCSRMACTCCTWGGACT
VK2 - 28	9	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGSWMGCACCTCCAGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSWMGCACCTCCAGVBCCCTWGGACT
VK2 - 28	1	3	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACCTCCAGACTCCTCCTACTTTTGGCGGAGGGACCAAG	DTSSWMGCACCTCCAGACTCCTCCTACT
VK2 - 28	10	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNAMNACAGACTCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCTCCTACT
VK2 - 28	11	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCSRMACTCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCTCCTACT
VK2 - 28	12	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCCAGVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCTCCTACT
VK2 - 28	13	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMACTCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTCCTACT
VK2 - 28	14	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNACAGVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCTCCTACT
VK2 - 28	15	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCACCTCSRMBVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGGCACCTCSRMBVBCCCTCCTACT

VK2-28	2	3	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTCCTACTTTTGGCGGAGGGACCCAAG	DTSCAGVNACTCCAGACTCCTCCTACT
VK2-28	3	3	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTCCTACTTTTGGCGGAGGGACCCAAG	DTSCAGGCAMNACAGACTCCTCCTACT
VK2-28	4	3	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSRMACTCCTCCTACTTTTGGCGGAGGGACCCAAG	DTSCAGGCACCTCSRMACTCCTCCTACT
VK2-28	5	3	9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCAGVBCCCTCCTACTTTTGGCGGAGGGACCCAAG	DTSCAGGCACCTCCAGVBCCCTCCTACT
VK2-28	6	3	9	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMVNACTCCAGACTCCTCCTACTTTTGGCGGAGGGACCCAAG	ATGSWMVNACTCCAGACTCCTCCTACT
VK2-28	7	3	9	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCAMNACAGACTCCTCCTACTTTTGGCGGAGGGACCCAAG	ATGSWMGCAMNACAGACTCCTCCTACT
VK2-28	8	3	9	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCACCTCSRMACTCCTCCTACTTTTGGCGGAGGGACCCAAG	ATGSWMGCACCTCSRMACTCCTCCTACT
VK2-28	9	3	9	GCTGAGGATGTTGGGGTTTATTACTGCAATGSWMGCACCTCCAGVBCCCTCCTACTTTTGGCGGAGGGACCCAAG	ATGSWMGCACCTCCAGVBCCCTCCTACT
VK3-11	1	0	9	CCTGAAGATTTGCAGTTTATTACTGTSMWSWMAGAAAGTAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	SWMSWMAGAAAGTAAATTGGCCCTTWCACT
VK3-11	10	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCNHCAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCNHCAAATTGGCCCTTWCACT
VK3-11	11	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCAGTNHCTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTNHCTGGCCCTTWCACT
VK3-11	12	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCAGTAAATYWCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTAAATYWCCCTTWCACT
VK3-11	13	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGANHCNHCCTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCNHCCTGGCCCTTWCACT
VK3-11	14	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGANHCAAATYWCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCAAATYWCCCTTWCACT
VK3-11	15	0	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGAAAGTNHCYWCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGAAAGTNHCYWCCTTWCACT
VK3-11	2	0	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGBHCAGTAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	SWMCAGBHCAGTAAATTGGCCCTTWCACT
VK3-11	3	0	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGAGANHCAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	SWMCAGAGANHCAAATTGGCCCTTWCACT
VK3-11	4	0	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGAGAAGTNHCTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	SWMCAGAGAAGTNHCTGGCCCTTWCACT
VK3-11	5	0	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGAGAAGTAAATYWCCCTTWCACTTTTGGCGGAGGGACCCAAG	SWMCAGAGAAGTAAATYWCCCTTWCACT
VK3-11	6	0	9	CCTGAAGATTTGCAGTTTATTACTGTCSGWSMBHCAGTAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGSWSMBHCAGTAAATTGGCCCTTWCACT
VK3-11	7	0	9	CCTGAAGATTTGCAGTTTATTACTGTCSGWSMAGANHCAAATTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGANHCAAATTGGCCCTTWCACT
VK3-11	8	0	9	CCTGAAGATTTGCAGTTTATTACTGTCSGWSMAGAAGTNHCTGGCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGAAGTNHCTGGCCCTTWCACT
VK3-11	9	0	9	CCTGAAGATTTGCAGTTTATTACTGTCSGWSMAGAAGTAAATYWCCCTTWCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGAAGTAAATYWCCCTTWCACT
VK3-11	1	1	9	CCTGAAGATTTGCAGTTTATTACTGTSMWSWMAGAAGTAAATTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	SWMSWMAGAAGTAAATTGGCCCTMTCACT
VK3-11	10	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCNHCAAATTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCNHCAAATTGGCCCTMTCACT
VK3-11	11	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCAGTNHCTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTNHCTGGCCCTMTCACT
VK3-11	12	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGBHCAGTAAATYWCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTAAATYWCCCTMTCACT
VK3-11	13	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGANHCNHCCTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCNHCCTGGCCCTMTCACT
VK3-11	14	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGANHCAAATYWCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCAAATYWCCCTMTCACT
VK3-11	15	1	9	CCTGAAGATTTGCAGTTTATTACTGTCCAGCAGAGAAGTNHCYWCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGCAGAGAAGTNHCYWCCTMTCACT
VK3-11	2	1	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGBHCAGTAAATTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	SWMCAGBHCAGTAAATTGGCCCTMTCACT
VK3-11	3	1	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGAGANHCAAATTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	SWMCAGAGANHCAAATTGGCCCTMTCACT
VK3-11	4	1	9	CCTGAAGATTTGCAGTTTATTACTGTSMWCAGAGAAGTNHCTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	SWMCAGAGAAGTNHCTGGCCCTMTCACT

VK3-11	5	1	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGAGAAAGTAATYWCCTTMTCACTTTTGGCGGAGGGACCCAAG	SMMCAGAGAAAGTAATYWCCTTMTCACT
VK3-11	6	1	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMBHCAGTAATTTGGCCTTMTCACTTTTGGCGGAGGGACCCAAG	CAGSWMBHCAGTAATTTGGCCTTMTCACT
VK3-11	7	1	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGANHCAAATTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGANHCAAATTGGCCTTMTCACT
VK3-11	8	1	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTNHCTGGCCCTMTCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGAAAGTNHCTGGCCTTMTCACT
VK3-11	9	1	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAA TYWCCCTTMTCACTTTTGGCGGAGGGACCCAAG	CAGSWMAGAAAGTAA TYWCCCTTMTCACT
VK3-11	1	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMSWMAGAAAGTAAATTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	SWMSWMAGAAAGTAAATTGGCCTWGGACT
VK3-11	10	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCAAATTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCNHCAAATTGGCCTWGGACT
VK3-11	11	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTNHCTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTNHCTGGCCTWGGACT
VK3-11	12	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTAA TYWCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTAA TYWCCCTWGGACT
VK3-11	13	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGANHCNHCCTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCNHCCTGGCCTWGGACT
VK3-11	14	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGANHCAA TYWCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCAA TYWCCCTWGGACT
VK3-11	15	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGAAAGTNHCYWCCTTGGGAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGAAAGTNHCYWCCTWGGACT
VK3-11	2	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGBHCAGTAAATTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	SMMCAGBHCAGTAAATTGGCCTWGGACT
VK3-11	3	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGANHCAAATTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGANHCAAATTGGCCTWGGACT
VK3-11	4	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGAAAGTNHCTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGAAAGTNHCTGGCCTWGGACT
VK3-11	5	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGAAAGTAA TYWCCCTWGGAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGAAAGTAA TYWCCCTWGGACT
VK3-11	6	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMBHCAGTAA TTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGSWMBHCAGTAA TTGGCCTWGGACT
VK3-11	7	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGANHCAA TTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGSWMAGANHCAA TTGGCCTWGGACT
VK3-11	8	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTNHCTGGCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGSWMAGAAAGTNHCTGGCCTWGGACT
VK3-11	9	2	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAA TYWCCCTWGGAC TTTTGGCGGAGGGACCCAAG	CAGSWMAGAAAGTAA TYWCCCTWGGACT
VK3-11	1	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMSWMAGAAAGTAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	SWMSWMAGAAAGTAA TTGGCCTCCTACT
VK3-11	10	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCNHCAA TTGGCCTCCTACT
VK3-11	11	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTNHCTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTNHCTGGCCTCCTACT
VK3-11	12	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTAA TYWCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGBHCAGTAA TYWCCCTCCTACT
VK3-11	13	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGANHCNHCCTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCNHCCTGGCCTCCTACT
VK3-11	14	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGANHCAA TYWCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGANHCAA TYWCCCTCCTACT
VK3-11	15	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGAAAGTNHCYWCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGCAGAGAAAGTNHCYWCCCTCCTACT
VK3-11	2	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGBHCAGTAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	SMMCAGBHCAGTAA TTGGCCTCCTACT
VK3-11	3	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGANHCAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGANHCAA TTGGCCTCCTACT
VK3-11	4	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGAAAGTNHCTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGAAAGTNHCTGGCCTCCTACT
VK3-11	5	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGAGAAAGTAA TYWCCCTCCTAC TTTTGGCGGAGGGACCCAAG	SMMCAGAGAAAGTAA TYWCCCTCCTACT
VK3-11	6	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMBHCAGTAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGSWMBHCAGTAA TTGGCCTCCTACT
VK3-11	7	3	9	CCTGAAGATTTTGCAGTTTATTACTGTCTGTSWMCAGANHCAA TTGGCCCTCCTAC TTTTGGCGGAGGGACCCAAG	CAGSWMAGANHCAA TTGGCCTCCTACT

VK3 -11	8	3	9	CCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMAGAAAGTNHCTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCTGGCCTCCTACT
VK3 -11	9	3	9	CCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMAGAAAGTAATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCCCTCCTACT
VK3 -15	1	0	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMSWMTACAAATAATTGGCCTTWCACTTTTGGCGGAGGGACCAAG	SWMSWMTACAAATAATTGGCCTTWCACT
VK3 -15	10	0	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCN	

VK3-15	10	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHC	CAATTGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHC	CAATTGGCCTWGGACT										
VK3-15	11	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATNHC	TGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATNHC	TGGCCTWGGACT								
VK3-15	12	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATAA	TYWCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATAA	TYWCCCTWGGACT								
VK3-15	13	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	NHC	TGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	NHC	TGGCCTWGGACT								
VK3-15	14	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	CAATY	WCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	CAATY	WCCCTWGGACT								
VK3-15	15	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTAC	CAATNHC	Y	WCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGCAGTAC	CAATNHC	Y	WCCCTWGGACT						
VK3-15	2	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGBHC	CAATAA	TTGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	SWMCAGBHC	CAATAA	TTGGCCTWGGACT								
VK3-15	3	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTACNHC	CAATNHC	CAAT	TTGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	SWMCAGTACNHC	CAATNHC	CAAT	TTGGCCTWGGACT						
VK3-15	4	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	CAATY	WCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATNHC	CAATY	WCCCTWGGACT						
VK3-15	5	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATAA	TY	WCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATAA	TY	WCCCTWGGACT						
VK3-15	6	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMBHC	CAATAA	TTGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGSWMBHC	CAATAA	TTGGCCTWGGACT								
VK3-15	7	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHC	CAATA	TTGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGSWMTACNHC	CAATA	TTGGCCTWGGACT								
VK3-15	8	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATNHC	TGGCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATNHC	TGGCCTWGGACT								
VK3-15	9	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTAC	CAATAA	TY	WCCCTWGGAC	TTTTTGGCGGAGGGACCAAG	CAGSWMTAC	CAATAA	TY	WCCCTWGGACT						
VK3-15	1	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMSWMTAC	CAATAA	TTGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	SWMSWMTAC	CAATAA	TTGGCCTCCTACT								
VK3-15	10	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHC	CAATA	TTGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHC	CAATA	TTGGCCTCCTACT								
VK3-15	11	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATNHC	TGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATNHC	TGGCCTCCTACT								
VK3-15	12	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHC	CAATAA	TY	WCCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CAATAA	TY	WCCCTCCTACT						
VK3-15	13	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	NHC	TGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	NHC	TGGCCTCCTACT								
VK3-15	14	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHC	CAATY	WCCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGTACNHC	CAATY	WCCCTCCTACT								
VK3-15	15	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTAC	CAATNHC	Y	WCCCTCCTACT	TTTTTGGCGGAGGGACCAAG	CAGCAGTAC	CAATNHC	Y	WCCCTCCTACT						
VK3-15	2	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGBHC	CAATAA	TTGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	SWMCAGBHC	CAATAA	TTGGCCTCCTACT								
VK3-15	3	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTACNHC	CAATNHC	CAAT	TTGGCCTCCTACT	SWMCAGTACNHC	CAATNHC	CAAT	TTGGCCTCCTACT							
VK3-15	4	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	TGGCCTCCTACT	TTTTTGGCGGAGGGACCAAG	SWMCAGTAC	CAATNHC	TGGCCTCCTACT								
VK3-15	5	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATAA	TY	WCCCTCCTACT	SWMCAGTAC	CAATAA	TY	WCCCTCCTACT							
VK3-15	6	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT							
VK3-15	7	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT							
VK3-15	8	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT							
VK3-15	9	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT	SWMCAGTAC	CAATNHC	CAAT	TTGGCCTCCTACT							
VK3-20	1	0	9	CCTGAAGATTTTGCAGTGTATTACTGT	SWMSWMTAC	CGGAAGTAG	TCC	TTWC	ACTTTTGGCGGAGGGACCAAG	SWMSWMTAC	CGGAAGTAG	TCC	TTWC	ACT					
VK3-20	10	0	9	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CBHC	CAGTAG	TCC	TTWC	ACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CBHC	CAGTAG	TCC	TTWC	ACT			
VK3-20	11	0	9	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAVN	CAGTCC	TTWC	ACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CGGAVN	CAGTCC	TTWC	ACT	TTC	TWC	ACT		
VK3-20	12	0	9	CCTGAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAAG	TBHC	CC	TTWC	ACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	CGGAAG	TBHC	CC	TTWC	ACT	TTC	TWC	ACT

VK3-20	2	2	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGBHC	CGGAAGTAGTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHC	CGGAAGTAGTCCTWGGACT		
VK3-20	3	2	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	BHCAGTAGTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	BHCAGTAGTCCTWGGACT		
VK3-20	4	2	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	CGGAVNCAGTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	CGGAVNCAGTCCTWGGACT		
VK3-20	5	2	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	CGGAAGTBHCCCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	CGGAAGTBHCCCCTWGGACT		
VK3-20	6	2	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMBHC	CGGAAGTAGTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHC	CGGAAGTAGTCCTWGGACT		
VK3-20	7	2	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	BHCAGTAGTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTAC	BHCAGTAGTCCTWGGACT		
VK3-20	8	2	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	CGGAVNCAGTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTAC	CGGAVNCAGTCCTWGGACT		
VK3-20	9	2	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	CGGAAGTBHCCCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTAC	CGGAAGTBHCCCCTWGGACT		
VK3-20	1	3	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMSWMTAC	CGGAAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	SWMSWMTAC	CGGAAGTAGTCCTCCTACT		
VK3-20	10	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGBHC	BHCAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHC	BHCAGTAGTCCTCCTACT		
VK3-20	11	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGBHC	CGGAVNCAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAVNCAGTCCTCCTACT		
VK3-20	12	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGBHC	CGGAAGTBHCCCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAAGTBHCCCCTCCTACT		
VK3-20	13	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTAC	BHCVNCAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTAC	BHCVNCAGTCCTCCTACT		
VK3-20	14	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTAC	BHCAGTBHCCCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTAC	BHCAGTBHCCCCTCCTACT		
VK3-20	15	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTAC	CGGAVNCBHCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTAC	CGGAVNCBHCCTCCTACT		
VK3-20	2	3	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGBHC	CGGAAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGBHC	CGGAAGTAGTCCTCCTACT		
VK3-20	3	3	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	BHCAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTAC	BHCAGTAGTCCTCCTACT		
VK3-20	4	3	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	CGGAVNCAGTCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTAC	CGGAVNCAGTCCTCCTACT		
VK3-20	5	3	9	CCTGAAGATTTTGCAGTGTATTACTGTSWMCAGTAC	CGGAAGTBHCCCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTAC	CGGAAGTBHCCCCTCCTACT		
VK3-20	6	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMBHC	CGGAAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMBHC	CGGAAGTAGTCCTCCTACT		
VK3-20	7	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	BHCAGTAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTAC	BHCAGTAGTCCTCCTACT		
VK3-20	8	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	CGGAVNCAGTCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTAC	CGGAVNCAGTCCTCCTACT		
VK3-20	9	3	9	CCTGAAGATTTTGCAGTGTATTACTGTCAGSWMTAC	CGGAAGTBHCCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTAC	CGGAAGTBHCCCCTCCTACT		
Jumping Trimer									
VK1-05	t1	0	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTTWCACT
VK1-05	t1	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTMTCACT
VK1-05	t1	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTWGGACT
VK1-05	t1	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTCCTACT
VK1-05	t2	0	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	CAATMBC	YWCYCTTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CAATMBC	YWCYCTTWCACT
VK1-05	t2	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	CAATMBC	YWCYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CAATMBC	YWCYCTMTCACT
VK1-05	t2	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	CAATMBC	YWCYCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CAATMBC	YWCYCTWGGACT
VK1-05	t2	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	CAATMBC	YWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CAATMBC	YWCYCTCCTACT
VK1-05	t3	0	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMAGTYWC	YCTTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMAGTYWC	YCTTWCACT

VK1-05	t3	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTYWCYCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCYCTMTCACT
VK1-05	t3	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCYCTWGGACT
VK1-05	t3	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTYWCYCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCYCTCCTACT
VK1-05	t4	0	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWCYCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCYCTTWCACT
VK1-05	t4	1	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWCYCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCYCTMTCACT
VK1-05	t4	2	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGGACT
VK1-05	t4	3	9	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWCYCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCYCTCCTACT
VK1-12	t1	0	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCNHCTTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCNHCTTCCCTTWCACT
VK1-12	t1	1	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCNHCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCNHCTTCCCTMTCACT
VK1-12	t1	2	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCNHCTTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCNHCTTCCCTWGGACT
VK1-12	t1	3	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCNHCTTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCNHCTTCCCTCCTACT
VK1-12	t2	0	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCAGTYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTYWCCCTTWCACT
VK1-12	t2	1	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCAGTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTYWCCCTMTCACT
VK1-12	t2	2	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCAGTYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTYWCCCTWGGACT
VK1-12	t2	3	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRANANHCAGTYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTYWCCCTCCTACT
VK1-12	t3	0	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAA TNHCYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAA TNHCYWCCCTTWCACT
VK1-12	t3	1	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAA TNHCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAA TNHCYWCCCTMTCACT
VK1-12	t3	2	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAA TNHCYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAA TNHCYWCCCTWGGACT
VK1-12	t3	3	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNAAA TNHCYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAA TNHCYWCCCTCCTACT
VK1-12	t4	0	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCCTTWCACT
VK1-12	t4	1	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCCTMTCACT
VK1-12	t4	2	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCCTWGGACT
VK1-12	t4	3	9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNHCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCCTCCTACT
VK1-33	t1	0	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCCCTCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCCTCCCTTWCACT
VK1-33	t1	1	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCCCTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCCTCCCTMTCACT
VK1-33	t1	2	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCCTCCTWGGACT
VK1-33	t1	3	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCCCTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCCTCCCTCCTACT
VK1-33	t2	0	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCAATYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATYWCCCTTWCACT
VK1-33	t2	1	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCAATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATYWCCCTMTCACT
VK1-33	t2	2	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCAATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATYWCCCTWGGACT
VK1-33	t2	3	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCAATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATYWCCCTCCTACT
VK1-33	t3	0	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCAGTNHCYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCCCTTWCACT
VK1-33	t3	1	9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCAGTNHCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCCCTMTCACT

VK1-33	t3	2	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGBHC	GATNHCYWCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	GATNHCYWCCTWGGACT
VK1-33	t3	3	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGBHC	GATNHCYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	GATNHCYWCCTCCTACT
VK1-33	t4	0	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGTAC	NHCNHCYWCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCNHCYWCCTTWCACT
VK1-33	t4	1	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGTAC	NHCNHCYWCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCNHCYWCCTMTCACT
VK1-33	t4	2	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGTAC	NHCNHCYWCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCNHCYWCCTWGGACT
VK1-33	t4	3	9	CCTGAAGATATTGCAACATATTACTGTCTAGCAGTAC	NHCNHCYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTAC	NHCNHCYWCCTCCTACT
VK1-39	t1	0	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCNHCACCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCNHCACCTCCTTWCACT
VK1-39	t1	1	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCNHCACCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCNHCACCTCCTMTCACT
VK1-39	t1	2	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCNHCACCTCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCNHCACCTCCTWGGACT
VK1-39	t1	3	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCNHCACCTCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCNHCACCTCCTCCTACT
VK1-39	t2	0	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCAGTBHCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCAGTBHCCCTTWCACT
VK1-39	t2	1	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCAGTBHCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCAGTBHCCCTMTCACT
VK1-39	t2	2	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCAGTBHCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCAGTBHCCCTWGGACT
VK1-39	t2	3	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ABHNCAGTBHCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ABHNCAGTBHCCCTCCTACT
VK1-39	t3	0	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ATACNHCBHCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ATACNHCBHCCCTTWCACT
VK1-39	t3	1	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ATACNHCBHCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ATACNHCBHCCCTMTCACT
VK1-39	t3	2	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ATACNHCBHCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ATACNHCBHCCCTWGGACT
VK1-39	t3	3	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAVN	ATACNHCBHCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVN	ATACNHCBHCCCTCCTACT
VK1-39	t4	0	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAA	GCBHCNHCBHCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAAA	GCBHCNHCBHCCCTTWCACT
VK1-39	t4	1	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAA	GCBHCNHCBHCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAA	GCBHCNHCBHCCCTMTCACT
VK1-39	t4	2	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAA	GCBHCNHCBHCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAA	GCBHCNHCBHCCCTWGGACT
VK1-39	t4	3	9	CCTGAAGATTTTGCAACTTACTACTGTCTAGCAAA	GCBHCNHCBHCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAAA	GCBHCNHCBHCCCTCCTACT
VK2-28	t1	0	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNASRMACTCCTTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNASRMACTCCTTWCACT
VK2-28	t1	1	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNASRMACTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNASRMACTCCTMTCACT
VK2-28	t1	2	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNASRMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNASRMACTCCTWGGACT
VK2-28	t1	3	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNASRMACTCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNASRMACTCCTCCTACT
VK2-28	t2	0	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNACAGVBCCCTTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNACAGVBCCCTTWCACT
VK2-28	t2	1	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNACAGVBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNACAGVBCCCTMTCACT
VK2-28	t2	2	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNACAGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNACAGVBCCCTWGGACT
VK2-28	t2	3	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNAMNACAGVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVN	AMNACAGVBCCCTCCTACT
VK2-28	t3	0	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNACTCSRMBCCCTTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	ACTCSRMBCCCTTWCACT
VK2-28	t3	1	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNACTCSRMBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVN	ACTCSRMBCCCTMTCACT
VK2-28	t3	2	9	GCTGAGGATGTTGGGGTTTATTACTGCA	TGCAGVNACTCSRMBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVN	ACTCSRMBCCCTWGGACT

VK2-28	t3	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCSRMBVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMBVBCCCTCCTACT
VK2-28	t4	0	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMBVBCCCTTWCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMBVBCCCTTWCACT
VK2-28	t4	1	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMBVBCCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMBVBCCCTMTCACT
VK2-28	t4	2	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMBVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMBVBCCCTWGGACT
VK2-28	t4	3	9	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMBVBCCCTCCTACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMBVBCCCTCCTACT
VK3-11	t1	0	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTTWCACT
VK3-11	t1	1	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTMTCACT
VK3-11	t1	2	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTWGGACT
VK3-11	t1	3	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTCCTACT
VK3-11	t2	0	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTTWCACT
VK3-11	t2	1	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTMTCACT
VK3-11	t2	2	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTWGGACT
VK3-11	t2	3	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTCCTACT
VK3-11	t3	0	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCAGTNHCHYWCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCHYWCCTTWCACT
VK3-11	t3	1	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCAGTNHCHYWCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCHYWCCTMTCACT
VK3-11	t3	2	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCAGTNHCHYWCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCHYWCCTWGGACT
VK3-11	t3	3	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCAGTNHCHYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCHYWCCTCCTACT
VK3-11	t4	0	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGANHNHCHYWCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHNHCHYWCCTTWCACT
VK3-11	t4	1	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGANHNHCHYWCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHNHCHYWCCTMTCACT
VK3-11	t4	2	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGANHNHCHYWCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHNHCHYWCCTWGGACT
VK3-11	t4	3	9	CCTGAAGATTTTGCAGTTTATTACTGTACGAGANHNHCHYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHNHCHYWCCTCCTACT
VK3-15	t1	0	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTTWCACT
VK3-15	t1	1	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTMTCACT
VK3-15	t1	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTWGGACT
VK3-15	t1	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCTGGCCTCCTACT
VK3-15	t2	0	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTTWCACT
VK3-15	t2	1	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTMTCACT
VK3-15	t2	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTWGGACT
VK3-15	t2	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCNHNHCAATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNHCAATYWCCCTCCTACT
VK3-15	t3	0	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCATNHNCHYWCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCATNHNCHYWCCTTWCACT
VK3-15	t3	1	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCATNHNCHYWCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCATNHNCHYWCCTMTCACT
VK3-15	t3	2	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCATNHNCHYWCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCATNHNCHYWCCTWGGACT
VK3-15	t3	3	9	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGAGBHCATNHNCHYWCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCATNHNCHYWCCTCCTACT

VK3-15	t4 0 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGTACNHCNHCYWCCTTWCAC	TTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTTWCAC	T
VK3-15	t4 1 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGTACNHCNHCYWCCTTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTTMTCA	T
VK3-15	t4 2 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGTACNHCNHCYWCCTTWGGA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTTWGGA	T
VK3-15	t4 3 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGTACNHCNHCYWCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTCCTA	T
VK3-20	t1 0 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCVNCAGTCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCVNCAGTCCTTWCAC	T
VK3-20	t1 1 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCVNCAGTCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCVNCAGTCCTMTCA	T
VK3-20	t1 2 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCVNCAGTCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCVNCAGTCCTWGA	T
VK3-20	t1 3 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCVNCAGTCCTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCVNCAGTCCTCA	T
VK3-20	t2 0 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCAGTBHCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTBHCCTTWCAC	T
VK3-20	t2 1 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCAGTBHCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTBHCCTMTCA	T
VK3-20	t2 2 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCAGTBHCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTBHCCTWGA	T
VK3-20	t2 3 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCAGTBHCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTBHCCTCCTA	T
VK3-20	t3 0 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTTWCAC	T
VK3-20	t3 1 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTMTCA	T
VK3-20	t3 2 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTWGA	T
VK3-20	t3 3 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGBHCBCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTCCTA	T
VK3-20	t4 0 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGTACBHCVNCBHCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCVNCBHCCTTWCAC	T
VK3-20	t4 1 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGTACBHCVNCBHCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCVNCBHCCTMTCA	T
VK3-20	t4 2 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGTACBHCVNCBHCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCVNCBHCCTWGA	T
VK3-20	t4 3 9	CCTGAAGATTTTGGCAGTGTTACTGTCTCAGCAGTACBHCVNCBHCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCVNCBHCCTCCTA	T
VK4-01	t1 0 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCNHCAC	TCCTTWCAC	TCTTWCAC	T
VK4-01	t1 1 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCNHCAC	TCCTMTCA	CTTMTCA	T
VK4-01	t1 2 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCNHCAC	TCCTWGA	CTCCTWGA	T
VK4-01	t1 3 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCNHCAC	TCCTCCTA	CTCCTCCTA	T
VK4-01	t2 0 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTTWCAC	T
VK4-01	t2 1 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTMTCA	T
VK4-01	t2 2 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTWGA	T
VK4-01	t2 3 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCBCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCBCCTCCTA	T
VK4-01	t3 0 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTTWCAC	T
VK4-01	t3 1 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTMTCA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTMTCA	T
VK4-01	t3 2 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTWGA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTWGA	T
VK4-01	t3 3 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTCCTA	CTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTCCTA	T
VK4-01	t4 0 9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBCCTTWCAC	TTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBCCTTWCAC	T

VK4-01	t4_1_9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBBHCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBBHCCCTMTCACT
VK4-01	t4_2_9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBBHCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBBHCCCTWGGACT
VK4-01	t4_3_9	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBBHCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBBHCCCTCCTACT

Table 7. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDRL3 length 10.

Name	Sequence of Synthesized Oligonucleotide		Portion of Oligonucleotide Corresponding to CDRL3 Proper
Jumping Dimer			
VK1 - 05 1 0 10	CCTGATGATTTTGGCAACTTATTACTGCSWMSWMTACAATAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 10 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVRMAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 11 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATMBCTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 12 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATAGTYWC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 13 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMMBCTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 14 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMAGTYWC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 15 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACAATMBCYWC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 2 0 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGBHCATAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 3 0 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACVRMAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 4 0 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATMBCTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 5 0 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATAGTYWC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 6 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMBHCAATAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 7 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACVRMAGTTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 8 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATMBCTAC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 9 0 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATAGTYWC	YCTCCTTWCAC	TTTTGGCGGAGGGACCAAG
VK1 - 05 1 1 10	CCTGATGATTTTGGCAACTTATTACTGCSWMSWMTACAATAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 10 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVRMAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 11 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATMBCTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 12 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATAGTYWC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 13 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMMBCTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 14 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMAGTYWC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 15 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACAATMBCYWC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 2 1 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGBHCATAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 3 1 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACVRMAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 4 1 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATMBCTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 5 1 10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATAGTYWC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 6 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMBHCAATAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG
VK1 - 05 7 1 10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACVRMAGTTAC	YCTCCTMTCA	CTTTTGGCGGAGGGACCAAG

VK1 - 05	8	1	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMWTACAAATMBC	TAC	YCTCCTMT	CAC	TCTTTGGCGGAGGGACCAAAG
VK1 - 05	9	1	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMWTACAAATAGTYWCYCTCCTMT	C	ACTTTTGGCGGAGGGACCAAAG		
VK1 - 05	1	2	10	CCTGATGATTTTGCAACTTATTACTGCSWMSWMWTACAAATAGTTAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG		
VK1 - 05	10	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVRMAGTTAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG		
VK1 - 05	11	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVAATMBC	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	12	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCVAATAGTYWCYCTCCTWGG	A	CTTTTGGCGGAGGGACCAAAG		
VK1 - 05	13	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMMBC	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	14	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACVRMAGTYWCYCTCCTWGG	A	CTTTTGGCGGAGGGACCAAAG		
VK1 - 05	15	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACAAATMBCYWCYCTCCTWGG	A	CTTTTGGCGGAGGGACCAAAG		
VK1 - 05	2	2	10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGBHCAATAGTTAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG		
VK1 - 05	3	2	10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAGT	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	4	2	10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATMBC	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	5	2	10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAAATAGTYWCYCTCCTWGG	A	CTTTTGGCGGAGGGACCAAAG		
VK1 - 05	6	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBCAAATAGTTAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG		
VK1 - 05	7	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTCVMMAGT	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	8	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTCAAATMBC	TAC	YCTCCTWGG	ACTTTTGGCGGAGGGACCAAAG	
VK1 - 05	9	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTCAAATAGTYWCYCTCCTWGG	A	CTTTTGGCGGAGGGACCAAAG		
VK1 - 12	1	0	10	CCTGAAAGATTTTGCAACTTATTACTGTSWMSWMGCAAAATAGTTTCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	10	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGRNANHCAAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	11	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGRNAAATNHCTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	12	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGRNAAATAGTYWCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	13	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGCANHCHCTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	14	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGCANHCAGTYWCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	15	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGCAAAATNHCVWCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	2	0	10	CCTGAAAGATTTTGCAACTTATTACTGTSWMCAGRNAAAATAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	3	0	10	CCTGAAAGATTTTGCAACTTATTACTGTSWMCAGCANHCAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	4	0	10	CCTGAAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATNHCTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	5	0	10	CCTGAAAGATTTTGCAACTTATTACTGTSWMCAGGCAAAATAGTYWCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	6	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGSWMRMAAATAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	7	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGSWMCANHCAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	8	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGSWMGCAAAATNHCTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	9	0	10	CCTGAAAGATTTTGCAACTTATTACTGTACGSWMGCAAAATAGTYWCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		
VK1 - 12	1	1	10	CCTGAAAGATTTTGCAACTTATTACTGTSMWSWMGCAAAATAGTTTCCCCTCCTTWCA	C	TTTTCCCTCCTTWCACT		

[illegible]

[illegible]

VK1-33	2	2	10	CCTGAAGATATTGCAACATATTACTGTSMWCAGBHC	GATAAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHC	GATAAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	WO	
VK1-33	3	2	10	CCTGAAGATATTGCAACATATTACTGTSMWCAGTAC	NHCAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	NHCAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	2012/009568	
VK1-33	4	2	10	CCTGAAGATATTGCAACATATTACTGTSMWCAGTAC	GATNHCCCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	GATNHCCCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG		
VK1-33	5	2	10	CCTGAAGATATTGCAACATATTACTGTSMWCAGTAC	GATAATYWCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTAC	GATAATYWCCTCCTWGGACTTTTGGCGGAGGACCAAG		
VK1-33	6	2	10	CCTGAAGATATTGCAACATATTACTGTAGSWM	BHCGATAAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWM	BHCGATAAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG		
VK1-33	7	2	10	CCTGAAGATATTGCAACATATTACTGTAGSWM	TACNHCAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWM	TACNHCAATCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG		
VK1-33	8	2	10	CCTGAAGATATTGCAACATATTACTGTAGSWM	TACGATNHCCCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWM	TACGATNHCCCTCCCTCCTWGGACTTTTGGCGGAGGACT		
VK1-33	9	2	10	CCTGAAGATATTGCAACATATTACTGTAGSWM	TACGATAATYWCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWM	TACGATAATYWCCTCCTWGGACTTTTGGCGGAGGACT		
VK1-39	1	0	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	SWMAGCTACAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	SWMSW	MAGCTACAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	10	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AAVNABHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AAVNABHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	11	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AVNATACNHCACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AVNATACNHCACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	12	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AVNATACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AVNATACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	13	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AAAGCBHCNHC	ACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCBHCNHC	ACTCCTCCTTWCACT
VK1-39	14	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AAAGCBHCAGTBHCCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCBHCAGTBHCCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	15	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGCA	AAAGCTACNHCBHCCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCTACNHCBHCCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	2	0	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAVNATACAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	SWMCA	AVNATACAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	3	0	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAAAGCBHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	SWMCA	AAAGCBHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	4	0	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAAAGCTACNHCACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	SWMCA	AAAGCTACNHCACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	5	0	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAAAGCTACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACCAAG	SWMCA	AAAGCTACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	6	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	WMVNATACAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	VNATACAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	7	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	WMAGCBHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	MAGCBHCAGTACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	8	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	WMAGCTACNHCACTCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	MAGCTACNHCACTCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	9	0	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	WMAGCTACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGSWM	MAGCTACAGTBHCCCTCCTTWCACTTTTGGCGGAGGACT		
VK1-39	1	1	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	SWMAGCTACAGTACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMSW	MAGCTACAGTACTCCTCCTMTCACTTTTGGCGGAGGACT		
VK1-39	10	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAVNABHCAGTACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AVNABHCAGTACTCCTCCTMTCACTTTTGGCGGAGGACT		
VK1-39	11	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAVNATACNHCACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AVNATACNHCACTCCTCCTMTCACTTTTGGCGGAGGACT		
VK1-39	12	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAVNATACAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AVNATACAGTBHCCCTCCTMTCACTTTTGGCGGAGGACT	P	
VK1-39	13	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAAAGCBHCNHC	ACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCBHCNHC	ACTCCTCCTMTCACT
VK1-39	14	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAAAGCBHCAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCBHCAGTBHCCCTCCTMTCACTTTTGGCGGAGGACT	U	
VK1-39	15	1	10	CCTGAAGATTTTGGCAACTTACTACTGTAGS	CAAAAGCTACNHCBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCA	AAAGCTACNHCBHCCCTCCTMTCACTTTTGGCGGAGGACT	2011/044063	
VK1-39	2	1	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAVNATACAGTACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCA	AVNATACAGTACTCCTCCTMTCACTTTTGGCGGAGGACT		
VK1-39	3	1	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAAAGCBHCAGTACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCA	AAAGCBHCAGTACTCCTCCTMTCACTTTTGGCGGAGGACT		
VK1-39	4	1	10	CCTGAAGATTTTGGCAACTTACTACTGTSM	CAAAAGCTACNHCACTCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCA	AAAGCTACNHCACTCCTCCTMTCACTTTTGGCGGAGGACT		

VK2-28	8	0	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCSRMACTCCTCTTWCACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCCTCTTWCACTTTTGGCGGAGGGACCAAG	WO 2012/009568
VK2-28	9	0	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCCAGVBCCCCTCTTWCACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCCCTCTTWCACTTTTGGCGGAGGGACCAAG	
VK2-28	1	1	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	10	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNAMNACAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	11	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	12	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK2-28	13	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	14	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNACAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	15	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCACTCSRMBVCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGGCACTCSRMBVCCCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	2	1	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	3	1	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	4	1	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	5	1	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	DTSCAGGCACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	6	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMVMNACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSMVMNACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	7	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSMWGCACTCCAGACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	8	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	9	1	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCCCTCCTMTCACTTTTGGCGGAGGGACCACT	
VK2-28	1	2	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	10	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNAMNACAGACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	11	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	12	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGVNACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	13	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNASRMACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	14	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCAMNACAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	15	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGCAGGCACTCSRMBVCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCACTCSRMBVCCCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	2	2	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	3	2	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACAGACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	4	2	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	5	2	10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	6	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMVMNACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMVMNACTCCAGACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	7	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMWGCACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	8	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCCTCCTWGGACTTTTGGCGGAGGGACCACT	
VK2-28	9	2	10	GCTGAGGATGTTGGGGTTTATTACTGATGSMWGCACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	
VK3-11	1	0	10	CCTGAAGATTTGCAGTTTATTACTGTSWMSWMAGAAATAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMSWMAGAAATAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	

VK3-11	11	10	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCAAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	W
VK3-11	11	11	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCAGTAAATGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	O
VK3-11	11	12	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	2012/009568
VK3-11	11	13	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGANHCNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGANHCNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	14	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGANHCATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	15	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGAAAGTNHCYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	2	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGBHCAGTAATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	3	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGANHCATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	4	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGAGAAAGTNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	5	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGAGAAAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	6	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMBHCAGTAATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	7	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGANHCATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCATTTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	8	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	9	0	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCCTCCTTWCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	1	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAATGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMSWMAGAAAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	10	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	11	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	12	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	13	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGANHCNHCCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	14	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGANHCATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	15	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGAGAAAGTNHCYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	2	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGBHCAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	3	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGANHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	4	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAGAAAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	5	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAGAAAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	6	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMBHCAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	7	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGANHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	8	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	9	1	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	1	2	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGSWMAGAAAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	SWMSWMAGAAAGTAATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	10	2	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCNHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATTTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	11	2	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK3-11	11	12	2	10	CCTGAAAGATTTTGCAGTTTATTACTGTCTGTCAGCAGBHCAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCTCCTMTCACTTTTGGCGGAGGGACCAAG	

VK3-11	13	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGCAGAGANHCNHNCTGGCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHNCTGGCCTCCTTWGGA	WO 2012/009568
VK3-11	14	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGCAGAGANHCAATYWCCCTCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTCCTTWGGA	
VK3-11	15	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGCAGAGAAAGTNHCYWCCTCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCCTCCTTWGGA	
VK3-11	2	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	SWMCAGBHCAGTAATTGGCCTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTCCTTWGGA	
VK3-11	3	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	SWMCAGAGANHCAATTGGCCTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGCCTCCTTWGGA	
VK3-11	4	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	SWMCAGAGAAAGTNHCYWCCTCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCYWCCTCCTTWGGA	
VK3-11	5	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	SWMCAGAGAAAGTAATYWCCCTCCTWGGACTTTTGGCGGAGGGACCACT	SWMCAGAGAAAGTAATYWCCCTCCTTWGGA	
VK3-11	6	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGSWMBHCAGTAATTGGCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTGGCCTCCTTWGGA	
VK3-11	7	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGSWMAGANHCAATTGGCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGCCTCCTTWGGA	
VK3-11	8	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGSWMAGAAAGTNHCYWCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCYWCCTCCTTWGGA	
VK3-11	9	2	10	CCTGAAAGATTTTGCAGTTTATTACTGT	CAGSWMAGAAAGTAATYWCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCCCTCCTTWGGA	
VK3-15	1	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMSWMTACAATAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMSWMTACAATAATTGGCCTCCTTWCACT	
VK3-15	10	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNCAATTGGCCTCCTTWCACT	
VK3-15	11	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNCAATTGGCCTCCTTWCACT	
VK3-15	12	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGBHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHNCAATTGGCCTCCTTWCACT	
VK3-15	13	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHNCTGGCCTCCTTWCACT	
VK3-15	14	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHNCTGGCCTCCTTWCACT	
VK3-15	15	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGCAGTACNHCNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHNCTGGCCTCCTTWCACT	
VK3-15	2	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGBHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCNHNCAATTGGCCTCCTTWCACT	PCT/US2011/044063
VK3-15	3	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTACNHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCNHNCAATTGGCCTCCTTWCACT	
VK3-15	4	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTACNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHNCTGGCCTCCTTWCACT	
VK3-15	5	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	SWMCAGTACNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHNCTGGCCTCCTTWCACT	
VK3-15	6	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMBHCNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCNHNCAATTGGCCTCCTTWCACT	
VK3-15	7	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHNCAATTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHNCAATTGGCCTCCTTWCACT	
VK3-15	8	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHNCTGGCCTCCTTWCACT	
VK3-15	8	0	10	CAGTCTGAAGATTTTGCAGTTTATTACTGT	CAGSWMTACNHNCTGGCCTCCTTWCACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHNCTGGCCTCCTTWCACT	

VK3-15	9 0 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGSWMTACAATAATYWCCCTCCTTWCACTTTTGGCGGAGGGACCA AG	CAGSWMTACAATAATYWCCCTCCTTWCA	WO 2012/009568
VK3-15	1 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTSWMSWMTACAATAATTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMSWMTACAATAATTGGCCTCCTMTCA	
VK3-15	10 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGBHCNHCAATTGGCCTCCTMTCA	
VK3-15	11 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGBHCAAATNHCTGGCCTCCTMTCA	
VK3-15	12 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCAAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGBHCAAATAATYWCCCTCCTMTCACT	
VK3-15	13 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGTACNHCHNCTGGCCTCCTMTCACT	
VK3-15	14 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGTACNHCAATYWCCCTCCTMTCACT	
VK3-15	15 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATNHCTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGCAGTACNHCAATYWCCCTCCTMTCACT	
VK3-15	2 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMCAGBHCAAATAATTGGCCTCCTMTCACT	
VK3-15	3 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMCAGTACNHCAATAATTGGCCTCCTMTCACT	
VK3-15	4 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMCAGTACNHCAATNHCTGGCCTCCTMTCACT	
VK3-15	5 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMCAGTACNHCAATAATYWCCCTCCTMTCACT	
VK3-15	6 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATTGGCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGSWMBHCAAATAATTGGCCTCCTMTCACT	
VK3-15	7 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGSWMTACNHCAATAATTGGCCTCCTMTCACT	
VK3-15	8 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGSWMTACAATNHCTGGCCTCCTMTCACT	
VK3-15	9 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	CAGSWMTACAATAATYWCCCTCCTMTCACT	PCT/US2011/044063
VK3-15	1 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTMTCACTTTTGGCGGAGGGACCA AG	SWMSWMTACAATAATTGGCCTCCTWGGAC	
VK3-15	10 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATTGGCCTCCTWGGACTTTTGGCGGAGGGACCA AG	CAGCAGBHCNHCAATAATTGGCCTCCTWGGAC	
VK3-15	11 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATNHCTGGCCTCCTWGGACTTTTGGCGGAGGGACCA AG	CAGCAGBHCAAATNHCTGGCCTCCTWGGAC	
VK3-15	12 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATAATYWCCCTCCTWGGACTTTTGGCGGAGGGACCA AG	CAGCAGBHCAAATAATYWCCCTCCTWGGAC	
VK3-15	13 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTGAGBHCNHCAATNHCTGGCCTCCTWGGACTTTTGGCGGAGGGACCA AG	CAGCAGTACNHCHNCTGGCCTCCTWGGAC	

	AG				
VK3 -15 14 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGCAGTACNHC	CAATYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 15 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGCAGTACAA	TNHCYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 2 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTSWMCAGBHC	AAATAATTGGCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 3 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTSWMCAGTACNHC	AAATTGGCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 4 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTSWMCAGTACAA	TNHCYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 5 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTSWMCAGTACAA	TAATYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 6 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGSWM	BHCAATAATTGGCCTCCTWGGACTTTT	TGGCGGAGGGACCA	
VK3 -15 7 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGSWM	TACNHC	AAATTGGCCTCCTWGGACTTTT	TGGCGGAGGGACCA
VK3 -15 8 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGSWM	TACAA	TNHCYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA
VK3 -15 9 2 10	CAGTCTGAAGATTTTGCAGTATTATTA	CTGTCAGSWM	TACAA	TAATYWCCCTCCTWGGACTTTT	TGGCGGAGGGACCA
VK3 -20 1 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMS	WMTACGGAAGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 10 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGBHC	BHCAGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 11 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGBHC	GAVNCAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 12 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGBHC	GAAAGTBHCCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 13 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGTACBHC	VNCAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 14 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGTACBHC	AGTBHCCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 15 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGCAGTACGGA	VNCBHCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 2 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMCAGBHC	GAAAGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 3 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMCAGTACBHC	AGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 4 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMCAGTACGGA	VNCAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 5 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMCAGTACGGA	AGTBHCCCTCCT	TWCACTTTTGGCGGAGGGACCAAG	
VK3 -20 6 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGSWM	BHC	GAAAGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG
VK3 -20 7 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGSWM	TACBHC	AGTAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG
VK3 -20 8 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGSWM	TACGGA	VNCAGTCCTCCT	TWCACTTTTGGCGGAGGGACCAAG
VK3 -20 9 0 10	CCTGAAGATTTTGCAGTGATTA	CTGTCAGSWM	TACGGA	AGTBHCCCTCCT	TWCACTTTTGGCGGAGGGACCAAG
VK3 -20 1 1 10	CCTGAAGATTTTGCAGTGATTA	CTGTSWMS	WMTACGGAAGTAGTCCTCCT	MTCACTTTTGGCGGAGGGACCAAG	

VK3-20	10	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	BHCAGTAGTCTCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	BHCAGTAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	W
VK3-20	11	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAVNCAGTCTCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAVNCAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	O
VK3-20	12	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	2012/009568
VK3-20	13	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CBHCAGTACCTCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTA	CBHCAGTACCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	14	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CBHCAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTA	CBHCAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	15	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CGGAVNCBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTA	CGGAVNCBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	2	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGBHC	CGGAAAGTAGTCTCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGBHC	CGGAAAGTAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	3	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	CBHCAGTAGTCTCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTA	CBHCAGTAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	4	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	CGGAVNCAGTCTCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTA	CGGAVNCAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	5	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	CGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTA	CGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	6	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMBHC	CGGAAAGTAGTCTCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHC	CGGAAAGTAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	7	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	CBHCAGTAGTCTCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTA	CBHCAGTAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	8	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	ACGGAVNCAGTCTCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTA	ACGGAVNCAGTCTCTCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	9	1	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	ACGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTA	ACGGAAAGTBHCCCTCCTMTCACTTTTGGCGGAGGACCAAG	
VK3-20	1	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMSWMTA	CGGAAAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	SWMSWMTA	CGGAAAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	10	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CBHCAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CBHCAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	11	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAVNCAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAVNCAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	12	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGBHC	CGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	CGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	13	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CBHCAGTACCTCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTA	CBHCAGTACCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	14	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CBHCAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTA	CBHCAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	15	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGCAGTA	CGGAVNCBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTA	CGGAVNCBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	2	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGBHC	CGGAAAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHC	CGGAAAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	3	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	CBHCAGTACCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTA	CBHCAGTACCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	4	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	CGGAVNCAGTCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTA	CGGAVNCAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	5	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	SWMCAGTA	ACGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTA	ACGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	6	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMBHC	CGGAAAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHC	CGGAAAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	7	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	CBHCAGTAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTA	CBHCAGTAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	
VK3-20	8	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	CGGAVNCAGTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTA	CGGAVNCAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	PC
VK3-20	9	2	10	CCTGAAAGATTTTGCAGTGTATTACTGT	CAGSWMTA	ACGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTA	ACGGAAAGTBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	US
Jumping Trimer									
VK1-05	t1	0	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTCCTTWCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTCCTTWCACT
VK1-05	t1	1	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTCCTMTCACT
VK1-05	t1	2	10	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHC	VRMMBCTAC	YCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHC	VRMMBCTAC	YCTCCTWGGACT

VK3-11	t4	0	10	CCTGAAGATTTTGCAGTTATTACTGTACGACAGANHCNHCYWCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-11	t4	1	10	CCTGAAGATTTTGCAGTTATTACTGTACGACAGANHCNHCYWCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-11	t4	2	10	CCTGAAGATTTTGCAGTTATTACTGTACGACAGANHCNHCYWCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-15	t1	0	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCHTGGCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-15	t1	1	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCHTGGCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-15	t1	2	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCHTGGCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-15	t2	0	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCAATYWCCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-15	t2	1	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCAATYWCCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-15	t2	2	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCHCNHCAATYWCCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-15	t3	0	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCAATNHCYWCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-15	t3	1	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCAATNHCYWCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-15	t3	2	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGABHCAATNHCYWCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-15	t4	0	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGACGTAACNHCNHCYWCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-15	t4	1	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGACGTAACNHCNHCYWCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-15	t4	2	10	CAGTCGAAGATTTTGCAGTTATTACTGTACGACGTAACNHCNHCYWCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-20	t1	0	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBVCVNCAGTCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-20	t1	1	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBVCVNCAGTCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-20	t1	2	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBVCVNCAGTCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-20	t2	0	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBCHCAGTBHCCCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-20	t2	1	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBCHCAGTBHCCCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-20	t2	2	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHBCHCAGTBHCCCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-20	t3	0	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHCGGAVNCBHHCCCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-20	t3	1	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHCGGAVNCBHHCCCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-20	t3	2	10	CCTGAAGATTTTGCAGTGATACTGTACGAGBHCGGAVNCBHHCCCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK3-20	t4	0	10	CCTGAAGATTTTGCAGTGATACTGTACGACGTAACBHVNCBHHCCCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK3-20	t4	1	10	CCTGAAGATTTTGCAGTGATACTGTACGACGTAACBHVNCBHHCCCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK3-20	t4	2	10	CCTGAAGATTTTGCAGTGATACTGTACGACGTAACBHVNCBHHCCCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK4-01	t1	0	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHNHCACTCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK4-01	t1	1	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHNHCACTCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK4-01	t1	2	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHNHCACTCCTCCTWGGACTTTTGCGGAGGGACCAAAG
VK4-01	t2	0	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHCACTBHCCCCCTCCTTWCACTTTTGCGGAGGGACCAAAG
VK4-01	t2	1	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHCACTBHCCCCCTCCTMTCACTTTTGCGGAGGGACCAAAG
VK4-01	t2	2	10	GCTGAAGATGTGCCAGTTTTA TACTGTACGABHCBHCACTBHCCCCCTCCTWGGACTTTTGCGGAGGGACCAAAG

VK4-01	t3	0	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTACNHCBCCHCCCTCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTACNHCBCCHCCCTCCTTWCACCTTTTGGCGGAGGGACCAAG	WO 2012/009568
VK4-01	t3	1	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTACNHCBCCHCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTACNHCBCCHCCCTCCTTWCACCTTTTGGCGGAGGGACCAAG	
VK4-01	t3	2	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGBHCCTACNHCBCCHCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCCTACNHCBCCHCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	
VK4-01	t4	0	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBCCHCCCTCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBCCHCCCTCCTTWCACCTTTTGGCGGAGGGACCAAG	
VK4-01	t4	1	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBCCHCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBCCHCCCTCCTMTCACTTTTGGCGGAGGGACCAAG	
VK4-01	t4	2	10	GCTGAAGATGTGGCAGTTTATTACTGTCTCAGCAGTACBHCNHCBCCHCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCNHCBCCHCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	

Table 8. Number of unique CDRL3 amino acid sequences in exemplary jumping dimer ("JD") and jumping trimer ("JT") VK libraries and comparison to VK-v1.0.

	L = 8 aa		L = 9 aa		L = 10 aa	
Jumping Dimer						
Germline	JD	VK-v1.0	JD	VK-v1.0	JD	VK-v1.0
VK1-05	3549	3072	7098	6144	6084	13824
VK1-12	5250	2016	5250	2016	4500	5184
VK1-33	5502	4032	5502	4032	4716	10368
VK1-39	7224	3024	7224	3024	6192	7776
VK2-28	4396	2016	4396	2016	3768	5184
VK3-11	6048	2352	6048	2016	5184	6048
VK3-15	5789	2016	5789	2352	4962	5184
VK3-20	6405	2016	6671	2016	5490	5184
VK4-01	NPE*	2016	6405	2592	NPE*	5184
Total	4.42*10 ⁴	2.26*10 ⁴	5.44*10 ⁴	2.62*10 ⁴	4.09*10 ⁴	6.39*10 ⁴
Jumping Trimer						
Germline	JT	VK-v1.0	JT	VK-v1.0	JT	VK-v1.0
VK1-05	7872	3072	13776	6144	11808	13824
VK1-12	14469	2016	14469	2016	12402	5184
VK1-33	15960	4032	15960	4032	13680	10368
VK1-39	28980	3024	28980	3024	24840	7776
VK2-28	12306	2016	12306	2016	10548	5184
VK3-11	18900	2352	18900	2016	16200	6048
VK3-15	18256	2016	18256	2352	15648	5184
VK3-20	23688	2016	23688	2016	20304	5184
VK4-01	23688	2016	23688	2592	20304	5184
Total	1.64*10 ⁵	2.26*10 ⁴	1.70*10 ⁵	2.62*10 ⁴	1.46*10 ⁵	6.39*10 ⁴

* Not Presently Exemplified. However, given the teachings of the specification, a person of ordinary skill in the art could readily produce a library of such lengths, and these lengths are included within the scope of the invention.

Table 9. Matching output for exemplary CDRH3 sequences from the HPS and TSP1. Amino acid mismatches in the theoretical design are indicated in bold.

Test Case	CDRH3 Sequence from HPS	Mismatches	TN1	DH	N2	H3 -JH
1	RTAHHFDY	0	R	TA	H	HFDY
2.1	VGIVGAASY	0	V	GIVGA	AS	Y
2.2	VGIVGAASY	0	VG	IVGA	AS	Y
3.1	DRYSGHDLGY	1	DR	YSGYD	LG	Y
4.1	GIAAADSNWLDP	1	-	GIAAA	D	SNW F DP
4.2	GIAAADSNWLDP	1	-	IAAA	D	SNW F DP
5.1	ERTINWGWGGVYAFDI	3	EGTG	NWG	GGV	YAFDI
5.2	ERTINWGWGGVYAFDI	3	EGTG	NWG	WGT	YAFDI
5.3	ERTINWGWGGVYAFDI	3	ERGG	NWG	GGV	YAFDI

5.4	ERTINWGWGGVYAFDI	3	ERGG	NWG	WGT	YAFDI
-----	------------------	---	------	-----	-----	-------

Table 10. Theoretical segment pool of 212 TN1 sequences contained in Theoretical Segment Pool 1 (TSP1).

TN1 Segment Name	Amino Acid Sequence	SEQ ID NO	TN1 Segment Name	Amino Acid Sequence	SEQ ID NO
P000	-		P107	PT	
P001	E		P108	EPT	
P002	D		P109	DPT	
P003	G		P110	GPT	
P004	EG		P111	PV	
P005	DG		P112	EPV	
P006	GG		P113	DPV	
P007	R		P114	GPV	
P008	ER		P115	RP	
P009	DR		P116	ERP	
P010	GR		P117	DRP	
P011	S		P118	GRP	
P012	ES		P119	SP	
P013	DS		P120	ESP	
P014	GS		P121	DSP	
P015	P		P122	GSP	
P016	EP		P123	LP	
P017	DP		P124	ELP	
P018	GP		P125	DLP	
P019	L		P126	GLP	
P020	EL		P127	AP	
P021	DL		P128	EAP	
P022	GL		P129	DAP	
P023	A		P130	GAP	
P024	EA		P131	TP	
P025	DA		P132	ETP	
P026	GA		P133	DTP	
P027	T		P134	GTP	
P028	ET		P135	VP	
P029	DT		P136	EVP	
P030	GT		P137	DVP	
P031	V		P138	GVP	
P032	EV		P139	AGG	
P033	DV		P140	EAGG	
P034	GV		P141	DAGG	
P035	EGG		P142	GAGG	

P036	DGG		P143	EGAG	
P037	GGG		P144	DGAG	
P038	EGR		P145	GGAG	
P039	DGR		P146	EGGA	
P040	GGR		P147	DGGA	
P041	EGS		P148	GGGA	
P042	DGS		P149	EGGG	
P043	GGG		P150	DGGG	
P044	EGP		P151	GGGG	
P045	DGP		P152	EGGL	
P046	GGP		P153	DGGL	
P047	EGL		P154	GGGL	
P048	DGL		P155	EGGP	
P049	GGL		P156	DGGP	
P050	EGA		P157	GGGP	
P051	DGA		P158	EGGR	
P052	GGA		P159	DGGR	
P053	EGT		P160	GGGR	
P054	DGT		P161	EGGS	
P055	GGT		P162	DGGS	
P056	EGV		P163	GGGS	
P057	DGV		P164	EGGT	
P058	GGV		P165	DGGT	
P059	RG		P166	GGGT	
P060	ERG		P167	EGGV	
P061	DRG		P168	DGGV	
P062	GRG		P169	GGGV	
P063	SG		P170	EGLG	
P064	ESG		P171	DGLG	
P065	DSG		P172	GGLG	
P066	GSG		P173	EGPG	
P067	PG		P174	DGPg	
P068	EPG		P175	GGPG	
P069	DPG		P176	EGRG	
P070	GPG		P177	DGRG	
P071	LG		P178	GGRG	
P072	ELG		P179	EGSG	
P073	DLG		P180	DGSG	
P074	GLG		P181	GGSG	
P075	AG		P182	EGTG	
P076	EAG		P183	DGTG	

P077	DAG			P184	GGTG		
P078	GAG			P185	EGVG		
P079	TG			P186	DGVG		
P080	ETG			P187	GGVG		
P081	DTG			P188	LGG		
P082	GTG			P189	ELGG		
P083	VG			P190	DLGG		
P084	EVG			P191	GLGG		
P085	DVG			P192	PGG		
P086	GVG			P193	EPGG		
P087	PR			P194	DPGG		
P088	EPR			P195	GPGG		
P089	DPR			P196	RGG		
P090	GPR			P197	ERGG		
P091	PS			P198	DRGG		
P092	EPS			P199	GRGG		
P093	DPS			P200	SGG		
P094	GPS			P201	ESGG		
P095	PP			P202	DSGG		
P096	EPP			P203	GSGG		
P097	DPP			P204	TGG		
P098	GPP			P205	ETGG		
P099	PL			P206	DTGG		
P100	EPL			P207	GTGG		
P101	DPL			P208	VGG		
P102	GPL			P209	EVGG		
P103	PA			P210	DVGG		
P104	EPA			P211	GVGG		
P105	DPA						
P106	GPA						

Table 11. 1K DH Theoretical Segment Pool sequences (1,111 DH segments).

DH Segment Name	Amino Acid Sequence	SEQ ID NO	DH Segment Name	Amino Acid Sequence	SEQ ID NO
DHUNIV 001	GTTGTT		DHUNIV 557	YDYV	
DHUNIV 002	GTTGT		DHUNIV 558	DYVW	
DHUNIV 003	TTGTT		DHUNIV 559	YVWG	
DHUNIV 004	GTTG		DHUNIV 560	VWGS	
DHUNIV 005	TTGT		DHUNIV 561	WGSY	
DHUNIV 006	TGTT		DHUNIV 562	GSYA	
DHUNIV 007	GTT		DHUNIV 563	SYAY	
DHUNIV 008	TTG		DHUNIV 564	YAYT	
DHUNIV 009	TGT		DHUNIV 565	YYD	
DHUNIV 010	GT		DHUNIV 566	YDY	
DHUNIV 011	TT		DHUNIV 567	DYV	
DHUNIV 012	TG		DHUNIV 568	YVW	
DHUNIV 013	VQLER		DHUNIV 569	VWG	
DHUNIV 014	VQLE		DHUNIV 570	WGS	
DHUNIV 015	QLER		DHUNIV 571	SYA	
DHUNIV 016	VQL		DHUNIV 572	YAY	
DHUNIV 017	QLE		DHUNIV 573	AYT	
DHUNIV 018	LER		DHUNIV 574	YD	
DHUNIV 019	VQ		DHUNIV 575	DY	
DHUNIV 020	QL		DHUNIV 576	YV	
DHUNIV 021	LE		DHUNIV 577	VW	
DHUNIV 022	ER		DHUNIV 578	WG	
DHUNIV 023	YNWND		DHUNIV 579	IMITFGGVMLIP	
DHUNIV 024	YNWN		DHUNIV 580	IMITFGGVMLI	
DHUNIV 025	NWND		DHUNIV 581	MITFGGVMLIP	
DHUNIV 026	YNW		DHUNIV 582	IMITFGGVML	
DHUNIV 027	NWN		DHUNIV 583	MITFGGVMLI	
DHUNIV 028	WND		DHUNIV 584	ITFGGVMLIP	
DHUNIV 029	YN		DHUNIV 585	IMITFGGVM	
DHUNIV 030	NW		DHUNIV 586	MITFGGVML	
DHUNIV 031	WN		DHUNIV 587	ITFGGVMLI	
DHUNIV 032	ND		DHUNIV 588	TFGGVMLIP	
DHUNIV 033	GITGTT		DHUNIV 589	IMITFGGV	
DHUNIV 034	GITGT		DHUNIV 590	MITFGGVM	
DHUNIV 035	ITGTT		DHUNIV 591	ITFGGVML	
DHUNIV 036	GITG		DHUNIV 592	TFGGVMLI	
DHUNIV 037	ITGT		DHUNIV 593	FGGVMLIP	

DHUNIV 038	GIT			DHUNIV 594	IMITFGG	
DHUNIV 039	ITG			DHUNIV 595	MITFGGV	
DHUNIV 040	GI			DHUNIV 596	ITFGGVM	
DHUNIV 041	IT			DHUNIV 597	TFGGVML	
DHUNIV 042	GIVGATT			DHUNIV 598	FGGVMLI	
DHUNIV 043	GIVGAT			DHUNIV 599	GGVMLIP	
DHUNIV 044	IVGATT			DHUNIV 600	IMITFG	
DHUNIV 045	GIVGA			DHUNIV 601	MITFGG	
DHUNIV 046	IVGAT			DHUNIV 602	ITFGGV	
DHUNIV 047	VGATT			DHUNIV 603	TFGGVM	
DHUNIV 048	GIVG			DHUNIV 604	FGGVML	
DHUNIV 049	IVGA			DHUNIV 605	GGVMLI	
DHUNIV 050	VGAT			DHUNIV 606	GVMLIP	
DHUNIV 051	GATT			DHUNIV 607	IMITF	
DHUNIV 052	GIV			DHUNIV 608	MITFG	
DHUNIV 053	IVG			DHUNIV 609	ITFGG	
DHUNIV 054	VGA			DHUNIV 610	TFGGV	
DHUNIV 055	GAT			DHUNIV 611	FGGVM	
DHUNIV 056	ATT			DHUNIV 612	GGVML	
DHUNIV 057	IV			DHUNIV 613	GVMLI	
DHUNIV 058	VG			DHUNIV 614	VMLIP	
DHUNIV 059	GA			DHUNIV 615	IMIT	
DHUNIV 060	AT			DHUNIV 616	MITF	
DHUNIV 061	WELL			DHUNIV 617	ITFG	
DHUNIV 062	WEL			DHUNIV 618	TFGG	
DHUNIV 063	ELL			DHUNIV 619	FGGV	
DHUNIV 064	WE			DHUNIV 620	GGVM	
DHUNIV 065	EL			DHUNIV 621	GVML	
DHUNIV 066	LL			DHUNIV 622	VMLI	
DHUNIV 067	YSGSYY			DHUNIV 623	MLIP	
DHUNIV 068	YSGSY			DHUNIV 624	IMI	
DHUNIV 069	SGSYY			DHUNIV 625	MIT	
DHUNIV 070	YSGS			DHUNIV 626	ITF	
DHUNIV 071	SGSY			DHUNIV 627	TFG	
DHUNIV 072	GSYY			DHUNIV 628	FGG	
DHUNIV 073	YSG			DHUNIV 629	GGV	
DHUNIV 074	SGS			DHUNIV 630	GVM	
DHUNIV 075	GSY			DHUNIV 631	VML	
DHUNIV 076	SY Y			DHUNIV 632	MLI	
DHUNIV 077	YS			DHUNIV 633	LIP	
DHUNIV 078	SG			DHUNIV 634	IM	

DHUNIV 079	GS			DHUNIV 635	MI	
DHUNIV 080	SY			DHUNIV 636	TF	
DHUNIV 081	YY			DHUNIV 637	VM	
DHUNIV 082	LEL			DHUNIV 638	LI	
DHUNIV 083	YNWNY			DHUNIV 639	WLL	
DHUNIV 084	NWNY			DHUNIV 640	WLL	
DHUNIV 085	WNY			DHUNIV 641	WL	
DHUNIV 086	NY			DHUNIV 642	YYDSSGYY	
DHUNIV 087	RIL			DHUNIV 643	YYDSSGYY	
DHUNIV 088	LLL			DHUNIV 644	YYDSSGYY	
DHUNIV 089	RI			DHUNIV 645	YYDSSGY	
DHUNIV 090	IL			DHUNIV 646	YYDSSGYY	
DHUNIV 091	WW			DHUNIV 647	YDSSGYY	
DHUNIV 092	GYCSGGSCYS			DHUNIV 648	YYDSSG	
DHUNIV 093	GYCSGGSCY			DHUNIV 649	YYDSSGY	
DHUNIV 094	YCSGGSCYS			DHUNIV 650	YDSSGYY	
DHUNIV 095	GYCSGGSC			DHUNIV 651	DSSGYY	
DHUNIV 096	YCSGGSCY			DHUNIV 652	YYDSS	
DHUNIV 097	CSGGSCYS			DHUNIV 653	YYDSSG	
DHUNIV 098	YCSGGSC			DHUNIV 654	YDSSGY	
DHUNIV 099	CSGGSCY			DHUNIV 655	DSSGYY	
DHUNIV 100	CSGGSC			DHUNIV 656	SSGYY	
DHUNIV 101	SGGS			DHUNIV 657	YYDS	
DHUNIV 102	SGG			DHUNIV 658	YYDSS	
DHUNIV 103	GGs			DHUNIV 659	YDSSG	
DHUNIV 104	GY			DHUNIV 660	DSSGY	
DHUNIV 105	GG			DHUNIV 661	SSGYY	
DHUNIV 106	DIVVVVAATP			DHUNIV 662	SGYY	
DHUNIV 107	DIVVVVAAT			DHUNIV 663	YYD	
DHUNIV 108	IVVVVAATP			DHUNIV 664	YYDS	
DHUNIV 109	DIVVVVAA			DHUNIV 665	YDSS	
DHUNIV 110	IVVVVAAT			DHUNIV 666	DSSG	
DHUNIV 111	VVVVAATP			DHUNIV 667	SSGY	
DHUNIV 112	DIVVVVA			DHUNIV 668	SGY	
DHUNIV 113	IVVVVAA			DHUNIV 669	GY	
DHUNIV 114	VVVVAAT			DHUNIV 670	YDS	
DHUNIV 115	VVVVAATP			DHUNIV 671	DSS	
DHUNIV 116	DIVVVV			DHUNIV 672	SSG	
DHUNIV 117	IVVVVA			DHUNIV 673	SGY	
DHUNIV 118	VVVVAA			DHUNIV 674	GY	
DHUNIV 119	VVVAAAT			DHUNIV 675	DS	

DHUNIV 120	VVAATP			DHUNIV 676	ITMIVVITTT	
DHUNIV 121	DIVVV			DHUNIV 677	ITMIVVVIT	
DHUNIV 122	IVVVV			DHUNIV 678	TMIVVVITTT	
DHUNIV 123	VVVVA			DHUNIV 679	ITMIVVVI	
DHUNIV 124	VVVAA			DHUNIV 680	TMIVVVVIT	
DHUNIV 125	VVAAT			DHUNIV 681	MIVVVITTT	
DHUNIV 126	VAAATP			DHUNIV 682	ITMIVVV	
DHUNIV 127	DIVV			DHUNIV 683	TMIVVVI	
DHUNIV 128	IVVV			DHUNIV 684	MIVVVIT	
DHUNIV 129	VVVV			DHUNIV 685	IVVVITTT	
DHUNIV 130	VVVA			DHUNIV 686	ITMIVV	
DHUNIV 131	VVAA			DHUNIV 687	TMIVVV	
DHUNIV 132	VAAAT			DHUNIV 688	MIVVVI	
DHUNIV 133	AATP			DHUNIV 689	IVVVIT	
DHUNIV 134	DIV			DHUNIV 690	VVVITTT	
DHUNIV 135	IVV			DHUNIV 691	ITMIV	
DHUNIV 136	VVV			DHUNIV 692	TMIVV	
DHUNIV 137	VVA			DHUNIV 693	MIVVV	
DHUNIV 138	VAA			DHUNIV 694	VVVIT	
DHUNIV 139	AAT			DHUNIV 695	VVITTT	
DHUNIV 140	ATP			DHUNIV 696	ITMI	
DHUNIV 141	DI			DHUNIV 697	TMIV	
DHUNIV 142	VV			DHUNIV 698	MIVV	
DHUNIV 143	VA			DHUNIV 699	VVIT	
DHUNIV 144	AA			DHUNIV 700	VITTT	
DHUNIV 145	TP			DHUNIV 701	TMI	
DHUNIV 146	YQLL			DHUNIV 702	MIV	
DHUNIV 147	YQL			DHUNIV 703	VIT	
DHUNIV 148	QLL			DHUNIV 704	ITT	
DHUNIV 149	YQ			DHUNIV 705	VLRFEWLly	
DHUNIV 150	GYCSSTSCYA			DHUNIV 706	VLRFEWLL	
DHUNIV 151	GYCSSTSCY			DHUNIV 707	LRFEWLly	
DHUNIV 152	YCSSTSCYA			DHUNIV 708	VLRFEWL	
DHUNIV 153	GYCSSTSC			DHUNIV 709	LRFEWLL	
DHUNIV 154	YCSSTSCY			DHUNIV 710	RFLEWLly	
DHUNIV 155	CSSTSCYA			DHUNIV 711	VLRFEW	
DHUNIV 156	YCSSTSC			DHUNIV 712	LRFEWL	
DHUNIV 157	CSSTSCY			DHUNIV 713	RFLEWLL	
DHUNIV 158	CSSTSC			DHUNIV 714	FLEWLly	
DHUNIV 159	SSTS			DHUNIV 715	VLRFLE	
DHUNIV 160	SST			DHUNIV 716	LRFEW	

DHUNIV 161	STS			DHUNIV 717	RFLEWL	
DHUNIV 162	SS			DHUNIV 718	FLEWLL	
DHUNIV 163	ST			DHUNIV 719	LEWLLY	
DHUNIV 164	TS			DHUNIV 720	VLRFL	
DHUNIV 165	YA			DHUNIV 721	LRFLE	
DHUNIV 166	DIVVPPAAMP			DHUNIV 722	RFLEW	
DHUNIV 167	DIVVPPAAM			DHUNIV 723	FLEWL	
DHUNIV 168	IVVPPAAMP			DHUNIV 724	LEWLL	
DHUNIV 169	DIVVPPAA			DHUNIV 725	EWLLY	
DHUNIV 170	IVVPPAAM			DHUNIV 726	VLRF	
DHUNIV 171	VVPPAAMP			DHUNIV 727	LRFL	
DHUNIV 172	DIVVPPA			DHUNIV 728	RFLE	
DHUNIV 173	IVVPPAA			DHUNIV 729	FLEW	
DHUNIV 174	VVPPAAM			DHUNIV 730	LEWL	
DHUNIV 175	VVPPAAMP			DHUNIV 731	EWLL	
DHUNIV 176	DIVVVP			DHUNIV 732	WLLY	
DHUNIV 177	IVVPPA			DHUNIV 733	VLR	
DHUNIV 178	VVPPAA			DHUNIV 734	LRF	
DHUNIV 179	VVPPAAM			DHUNIV 735	RFL	
DHUNIV 180	VPAAMP			DHUNIV 736	FLE	
DHUNIV 181	IVVVP			DHUNIV 737	LEW	
DHUNIV 182	VVPPA			DHUNIV 738	EWL	
DHUNIV 183	VVPAA			DHUNIV 739	RF	
DHUNIV 184	VPAAM			DHUNIV 740	FL	
DHUNIV 185	PAAMP			DHUNIV 741	EW	
DHUNIV 186	VVVP			DHUNIV 742	YYDFWSGYT	
DHUNIV 187	VVPA			DHUNIV 743	YYDFWSGY	
DHUNIV 188	VPAA			DHUNIV 744	YDFWSGYT	
DHUNIV 189	PAAM			DHUNIV 745	YYDFWSGY	
DHUNIV 190	AAMP			DHUNIV 746	YDFWSGY	
DHUNIV 191	VVP			DHUNIV 747	DFWSGYT	
DHUNIV 192	VPA			DHUNIV 748	YYDFWSG	
DHUNIV 193	PAA			DHUNIV 749	YDFWSGY	
DHUNIV 194	AAM			DHUNIV 750	DFWSGY	
DHUNIV 195	AMP			DHUNIV 751	FWSGYT	
DHUNIV 196	VP			DHUNIV 752	YYDFWS	
DHUNIV 197	PA			DHUNIV 753	YDFWSG	
DHUNIV 198	AM			DHUNIV 754	DFWSGY	
DHUNIV 199	MP			DHUNIV 755	FWSGY	
DHUNIV 200	YQLLY			DHUNIV 756	WSGYT	
DHUNIV 201	QLLY			DHUNIV 757	YYDFW	

DHUNIV 202	LLY			DHUNIV 758	YDFWS	
DHUNIV 203	LY			DHUNIV 759	DFWSG	
DHUNIV 204	GYCSSTSCYT			DHUNIV 760	FWSGY	
DHUNIV 205	YCSSTSCYT			DHUNIV 761	WSGY	
DHUNIV 206	CSSTSCYT			DHUNIV 762	SGYYT	
DHUNIV 207	YT			DHUNIV 763	YYDF	
DHUNIV 208	DIVVVPAAIP			DHUNIV 764	YDFW	
DHUNIV 209	DIVVVPAAI			DHUNIV 765	DFWS	
DHUNIV 210	IVVVPAAIP			DHUNIV 766	FWSG	
DHUNIV 211	IVVVPAAI			DHUNIV 767	WSGY	
DHUNIV 212	VVVPAAIP			DHUNIV 768	GYT	
DHUNIV 213	VVVPAAI			DHUNIV 769	YDF	
DHUNIV 214	VVPAAIP			DHUNIV 770	DFW	
DHUNIV 215	VPAAI			DHUNIV 771	FWS	
DHUNIV 216	VPAAIP			DHUNIV 772	WSG	
DHUNIV 217	VPAAI			DHUNIV 773	YYT	
DHUNIV 218	PAAIP			DHUNIV 774	DF	
DHUNIV 219	PAAI			DHUNIV 775	FW	
DHUNIV 220	AAIP			DHUNIV 776	WS	
DHUNIV 221	AAI			DHUNIV 777	ITIFGVVIIP	
DHUNIV 222	AIP			DHUNIV 778	ITIFGVVII	
DHUNIV 223	AI			DHUNIV 779	TIFGVVIIP	
DHUNIV 224	IP			DHUNIV 780	ITIFGVVI	
DHUNIV 225	WIL			DHUNIV 781	TIFGVVII	
DHUNIV 226	WI			DHUNIV 782	IFGVVIIP	
DHUNIV 227	SILWW			DHUNIV 783	ITIFGVV	
DHUNIV 228	SILW			DHUNIV 784	TIFGVVI	
DHUNIV 229	ILWW			DHUNIV 785	IFGVVII	
DHUNIV 230	SIL			DHUNIV 786	FGVVIIP	
DHUNIV 231	ILW			DHUNIV 787	ITIFGV	
DHUNIV 232	LWW			DHUNIV 788	TIFGVV	
DHUNIV 233	LLF			DHUNIV 789	IFGVVI	
DHUNIV 234	SI			DHUNIV 790	FGVII	
DHUNIV 235	LW			DHUNIV 791	GVVIIP	
DHUNIV 236	LF			DHUNIV 792	ITIFG	
DHUNIV 237	AYCGGDCYS			DHUNIV 793	TIFGV	
DHUNIV 238	AYCGGDCY			DHUNIV 794	IFGVV	
DHUNIV 239	YCGGDCYS			DHUNIV 795	FGVI	
DHUNIV 240	AYCGGDC			DHUNIV 796	GVVII	
DHUNIV 241	YCGGDCY			DHUNIV 797	VVIIP	
DHUNIV 242	CGGDCYS			DHUNIV 798	ITIF	

DHUNIV 243	YCGGDC		DHUNIV 799	TIFG	
DHUNIV 244	CGGDCY		DHUNIV 800	IFGV	
DHUNIV 245	CGGDC		DHUNIV 801	FGVV	
DHUNIV 246	GGD		DHUNIV 802	GVVI	
DHUNIV 247	AY		DHUNIV 803	VVII	
DHUNIV 248	GD		DHUNIV 804	VIIP	
DHUNIV 249	HIVVIAIP		DHUNIV 805	ITI	
DHUNIV 250	HIVVIAI		DHUNIV 806	TIF	
DHUNIV 251	IVVIAIP		DHUNIV 807	IFG	
DHUNIV 252	HIVVIA		DHUNIV 808	FGV	
DHUNIV 253	IVVIAI		DHUNIV 809	GVV	
DHUNIV 254	VVIAIP		DHUNIV 810	IIP	
DHUNIV 255	HIVVI		DHUNIV 811	TI	
DHUNIV 256	IVVIA		DHUNIV 812	IF	
DHUNIV 257	VVIAI		DHUNIV 813	VLRYFDWLL	
DHUNIV 258	VVIAIP		DHUNIV 814	VLRYFDWL	
DHUNIV 259	HIVVV		DHUNIV 815	LRYFDWLL	
DHUNIV 260	IVVVI		DHUNIV 816	VLRYFDW	
DHUNIV 261	VVIA		DHUNIV 817	LRYFDWL	
DHUNIV 262	VVIAI		DHUNIV 818	RYFDWLL	
DHUNIV 263	VIAIP		DHUNIV 819	VLRYFD	
DHUNIV 264	HIVV		DHUNIV 820	LRYFDW	
DHUNIV 265	VVVI		DHUNIV 821	RYFDWL	
DHUNIV 266	VVIA		DHUNIV 822	YFDWLL	
DHUNIV 267	VIAI		DHUNIV 823	VLRYF	
DHUNIV 268	IAIP		DHUNIV 824	LRYFD	
DHUNIV 269	HIV		DHUNIV 825	RYFDW	
DHUNIV 270	VVI		DHUNIV 826	YFDWL	
DHUNIV 271	VIA		DHUNIV 827	FDWLL	
DHUNIV 272	IAI		DHUNIV 828	VLRY	
DHUNIV 273	HI		DHUNIV 829	LRYP	
DHUNIV 274	VI		DHUNIV 830	RYFD	
DHUNIV 275	IA		DHUNIV 831	YFDW	
DHUNIV 276	HIVVTAIP		DHUNIV 832	FDWL	
DHUNIV 277	HIVVTAI		DHUNIV 833	DWLL	
DHUNIV 278	IVVTAIP		DHUNIV 834	LRY	
DHUNIV 279	HIVVTA		DHUNIV 835	RYF	
DHUNIV 280	IVVTAI		DHUNIV 836	YFD	
DHUNIV 281	VVTAIP		DHUNIV 837	FDW	
DHUNIV 282	HIVVVT		DHUNIV 838	DWL	
DHUNIV 283	IVVTA		DHUNIV 839	RY	

DHUNIV 284	VVVTAI		DHUNIV 840	YF	
DHUNIV 285	VVTAIP		DHUNIV 841	FD	
DHUNIV 286	IVVVT		DHUNIV 842	DW	
DHUNIV 287	VVVTA		DHUNIV 843	YYDILTGYYN	
DHUNIV 288	VVTAI		DHUNIV 844	YYDILTGYY	
DHUNIV 289	VTAIP		DHUNIV 845	YDILTGYYN	
DHUNIV 290	VVVT		DHUNIV 846	YYDILTG	
DHUNIV 291	VVTA		DHUNIV 847	YDILTGYY	
DHUNIV 292	VTAI		DHUNIV 848	DILTGYYN	
DHUNIV 293	TAIP		DHUNIV 849	YYDILTG	
DHUNIV 294	VVT		DHUNIV 850	YDILTG	
DHUNIV 295	VTA		DHUNIV 851	DILTGYY	
DHUNIV 296	TAI		DHUNIV 852	ILTGYYN	
DHUNIV 297	VT		DHUNIV 853	YYDILT	
DHUNIV 298	TA		DHUNIV 854	YDILTG	
DHUNIV 299	RILY		DHUNIV 855	DILTG	
DHUNIV 300	ILY		DHUNIV 856	ILTGYY	
DHUNIV 301	MLY		DHUNIV 857	LTGYYN	
DHUNIV 302	ML		DHUNIV 858	YYDIL	
DHUNIV 303	GYCTNGVCYT		DHUNIV 859	YDILT	
DHUNIV 304	GYCTNGVCY		DHUNIV 860	DILTG	
DHUNIV 305	YCTNGVCYT		DHUNIV 861	ILTG	
DHUNIV 306	GYCTNGVC		DHUNIV 862	LTGYY	
DHUNIV 307	YCTNGVCY		DHUNIV 863	TGYYN	
DHUNIV 308	CTNGVCYT		DHUNIV 864	YYDI	
DHUNIV 309	YCTNGVC		DHUNIV 865	YDIL	
DHUNIV 310	CTNGVCY		DHUNIV 866	DILT	
DHUNIV 311	CTNGVC		DHUNIV 867	ILTG	
DHUNIV 312	TNGV		DHUNIV 868	LTG	
DHUNIV 313	TNG		DHUNIV 869	TGYY	
DHUNIV 314	NGV		DHUNIV 870	GYN	
DHUNIV 315	TN		DHUNIV 871	YDI	
DHUNIV 316	NG		DHUNIV 872	DIL	
DHUNIV 317	GV		DHUNIV 873	ILT	
DHUNIV 318	DIVLMVYAIP		DHUNIV 874	LTG	
DHUNIV 319	DIVLMVYAI		DHUNIV 875	TG	
DHUNIV 320	IVLMVYAIP		DHUNIV 876	LT	
DHUNIV 321	DIVLMVYA		DHUNIV 877	LVIIT	
DHUNIV 322	IVLMVYAI		DHUNIV 878	LVII	
DHUNIV 323	VLMVYAIP		DHUNIV 879	LVI	
DHUNIV 324	DIVLMVY		DHUNIV 880	LV	

DHUNIV 325	IVLMVYA		DHUNIV 881	DYGDY	
DHUNIV 326	VLMVYAI		DHUNIV 882	DYGD	
DHUNIV 327	LMVYAIIP		DHUNIV 883	YGDY	
DHUNIV 328	DIVLMV		DHUNIV 884	DYG	
DHUNIV 329	IVLMVY		DHUNIV 885	YGD	
DHUNIV 330	VLMVYA		DHUNIV 886	GDY	
DHUNIV 331	LMVYAI		DHUNIV 887	TTVTTT	
DHUNIV 332	MVYAIIP		DHUNIV 888	TTVT	
DHUNIV 333	DIVLM		DHUNIV 889	TVTT	
DHUNIV 334	IVLMV		DHUNIV 890	TTV	
DHUNIV 335	VLMVY		DHUNIV 891	TVT	
DHUNIV 336	LMVYA		DHUNIV 892	VTT	
DHUNIV 337	MVYAI		DHUNIV 893	TV	
DHUNIV 338	VYAIIP		DHUNIV 894	LRW	
DHUNIV 339	DIVL		DHUNIV 895	RW	
DHUNIV 340	IVLM		DHUNIV 896	DYGGNS	
DHUNIV 341	VLMV		DHUNIV 897	DYGGN	
DHUNIV 342	LMVY		DHUNIV 898	YGGNS	
DHUNIV 343	MVYA		DHUNIV 899	DYGG	
DHUNIV 344	VYAI		DHUNIV 900	YGN	
DHUNIV 345	YAIIP		DHUNIV 901	GGNS	
DHUNIV 346	IVL		DHUNIV 902	YGG	
DHUNIV 347	VLM		DHUNIV 903	GGN	
DHUNIV 348	LMV		DHUNIV 904	GNS	
DHUNIV 349	MVY		DHUNIV 905	GN	
DHUNIV 350	VYA		DHUNIV 906	NS	
DHUNIV 351	YAI		DHUNIV 907	TTVVTP	
DHUNIV 352	VL		DHUNIV 908	TTVVT	
DHUNIV 353	LM		DHUNIV 909	TVVTP	
DHUNIV 354	MV		DHUNIV 910	TTVV	
DHUNIV 355	VY		DHUNIV 911	TVVT	
DHUNIV 356	VLLWFGELL		DHUNIV 912	VVTP	
DHUNIV 357	VLLWFGEL		DHUNIV 913	TVV	
DHUNIV 358	LLWFGELL		DHUNIV 914	VTP	
DHUNIV 359	VLLWFGE		DHUNIV 915	LQ	
DHUNIV 360	LLWFGEL		DHUNIV 916	DYSNY	
DHUNIV 361	LWFGELL		DHUNIV 917	DYSN	
DHUNIV 362	VLLWFG		DHUNIV 918	YSNY	
DHUNIV 363	LLWFGE		DHUNIV 919	DYS	
DHUNIV 364	LWFGEL		DHUNIV 920	YSN	
DHUNIV 365	WFGELL		DHUNIV 921	SNY	

DHUNIV 366	VLLWF			DHUNIV 922	SN	
DHUNIV 367	LLWFG			DHUNIV 923	VDIVATIT	
DHUNIV 368	LWFG			DHUNIV 924	VDIVATI	
DHUNIV 369	WFGEL			DHUNIV 925	DIVATIT	
DHUNIV 370	FGELL			DHUNIV 926	VDIVAT	
DHUNIV 371	VLLW			DHUNIV 927	DIVATI	
DHUNIV 372	LLWF			DHUNIV 928	IVATIT	
DHUNIV 373	LWFG			DHUNIV 929	VDIVA	
DHUNIV 374	WFG			DHUNIV 930	DIVAT	
DHUNIV 375	FGEL			DHUNIV 931	IVATI	
DHUNIV 376	GELL			DHUNIV 932	VATIT	
DHUNIV 377	VLL			DHUNIV 933	VDIV	
DHUNIV 378	LLW			DHUNIV 934	DIVA	
DHUNIV 379	LWF			DHUNIV 935	IVAT	
DHUNIV 380	WFG			DHUNIV 936	VATI	
DHUNIV 381	FGE			DHUNIV 937	ATIT	
DHUNIV 382	GEL			DHUNIV 938	VDI	
DHUNIV 383	WF			DHUNIV 939	IVA	
DHUNIV 384	FG			DHUNIV 940	VAT	
DHUNIV 385	GE			DHUNIV 941	ATI	
DHUNIV 386	YYGSGSYN			DHUNIV 942	TIT	
DHUNIV 387	YYGSGSY			DHUNIV 943	VD	
DHUNIV 388	YGGSGSYN			DHUNIV 944	WLRL	
DHUNIV 389	YYGSGSY			DHUNIV 945	WLR	
DHUNIV 390	YGGSGSY			DHUNIV 946	GYSGYDY	
DHUNIV 391	YGGSGSYN			DHUNIV 947	GYSGYD	
DHUNIV 392	YYGSGS			DHUNIV 948	YSGYDY	
DHUNIV 393	YGGSGSY			DHUNIV 949	GYSGY	
DHUNIV 394	YGGSGSY			DHUNIV 950	YSGYD	
DHUNIV 395	GSGSYN			DHUNIV 951	SGYDY	
DHUNIV 396	YYGSG			DHUNIV 952	GYSG	
DHUNIV 397	YGGSGS			DHUNIV 953	YSGY	
DHUNIV 398	YGGSGY			DHUNIV 954	SGYD	
DHUNIV 399	GSGSY			DHUNIV 955	GYDY	
DHUNIV 400	SGSYN			DHUNIV 956	GYS	
DHUNIV 401	YYGGS			DHUNIV 957	GYD	
DHUNIV 402	YGGSG			DHUNIV 958	VEMATIT	
DHUNIV 403	YGGSGS			DHUNIV 959	VEMATI	
DHUNIV 404	GSGSY			DHUNIV 960	EMATIT	
DHUNIV 405	GSYYN			DHUNIV 961	VEMAT	
DHUNIV 406	YYYG			DHUNIV 962	EMATI	

DHUNIV 407	YYGS			DHUNIV 963	MATIT	
DHUNIV 408	YGS			DHUNIV 964	VEMA	
DHUNIV 409	SGS			DHUNIV 965	EMAT	
DHUNIV 410	SYN			DHUNIV 966	MATI	
DHUNIV 411	YY			DHUNIV 967	VEM	
DHUNIV 412	YG			DHUNIV 968	EMA	
DHUNIV 413	YGS			DHUNIV 969	MAT	
DHUNIV 414	SG			DHUNIV 970	VE	
DHUNIV 415	YN			DHUNIV 971	EM	
DHUNIV 416	YG			DHUNIV 972	MA	
DHUNIV 417	ITMVRGVIIT			DHUNIV 973	RWLQL	
DHUNIV 418	ITMVRGVII			DHUNIV 974	RWLQ	
DHUNIV 419	TMVRGVIIT			DHUNIV 975	WLQL	
DHUNIV 420	ITMVRGVI			DHUNIV 976	RWL	
DHUNIV 421	TMVRGVII			DHUNIV 977	WLQ	
DHUNIV 422	MVRGVIIT			DHUNIV 978	LQL	
DHUNIV 423	ITMVRGV			DHUNIV 979	RDGYN	
DHUNIV 424	TMVRGVI			DHUNIV 980	RDGYN	
DHUNIV 425	MVRGVII			DHUNIV 981	DGYN	
DHUNIV 426	VRGVIIT			DHUNIV 982	RDG	
DHUNIV 427	ITMVRG			DHUNIV 983	DGYN	
DHUNIV 428	TMVRGV			DHUNIV 984	GYN	
DHUNIV 429	MVRGVI			DHUNIV 985	RDG	
DHUNIV 430	VRGVII			DHUNIV 986	DG	
DHUNIV 431	RGVIIT			DHUNIV 987	GYN	
DHUNIV 432	ITMVR			DHUNIV 988	YN	
DHUNIV 433	TMVRG			DHUNIV 989	RD	
DHUNIV 434	MVRGV			DHUNIV 990	DG	
DHUNIV 435	VRGVI			DHUNIV 991	VDTAMVT	
DHUNIV 436	RGVI			DHUNIV 992	VDTAMV	
DHUNIV 437	GVIIT			DHUNIV 993	DTAMVT	
DHUNIV 438	ITMV			DHUNIV 994	VDTAM	
DHUNIV 439	TMVR			DHUNIV 995	DTAMV	
DHUNIV 440	MVRG			DHUNIV 996	TAMVT	
DHUNIV 441	VRGV			DHUNIV 997	VDTA	
DHUNIV 442	RGVI			DHUNIV 998	DTAM	
DHUNIV 443	GVI			DHUNIV 999	TAMV	
DHUNIV 444	VIIT			DHUNIV 1000	AMVT	
DHUNIV 445	ITM			DHUNIV 1001	VDT	
DHUNIV 446	TMV			DHUNIV 1002	DTA	
DHUNIV 447	MVR			DHUNIV 1003	TAM	

DHUNIV 448	VRG			DHUNIV 1004	AMV		
DHUNIV 449	RGV			DHUNIV 1005	MVT		
DHUNIV 450	GVI			DHUNIV 1006	DT		
DHUNIV 451	VII			DHUNIV 1007	WQLWL		
DHUNIV 452	IIT			DHUNIV 1008	WQLW		
DHUNIV 453	TM			DHUNIV 1009	IQLWL		
DHUNIV 454	VR			DHUNIV 1010	WQL		
DHUNIV 455	RG			DHUNIV 1011	IQLW		
DHUNIV 456	II			DHUNIV 1012	QLWL		
DHUNIV 457	VLLWFRELL			DHUNIV 1013	WIQ		
DHUNIV 458	VLLWFREL			DHUNIV 1014	IQL		
DHUNIV 459	LLWFRELL			DHUNIV 1015	QLW		
DHUNIV 460	VLLWFRE			DHUNIV 1016	LWL		
DHUNIV 461	LLWFREL			DHUNIV 1017	IQ		
DHUNIV 462	LWFRELL			DHUNIV 1018	GYSYG		
DHUNIV 463	VLLWFR			DHUNIV 1019	GYSYG		
DHUNIV 464	LLWFRE			DHUNIV 1020	YSYG		
DHUNIV 465	LWFREL			DHUNIV 1021	GYSY		
DHUNIV 466	WFRELL			DHUNIV 1022	YSYG		
DHUNIV 467	LLWFR			DHUNIV 1023	SYGY		
DHUNIV 468	LWFRE			DHUNIV 1024	YSY		
DHUNIV 469	WFREL			DHUNIV 1025	SYG		
DHUNIV 470	FRELL			DHUNIV 1026	YGY		
DHUNIV 471	LWFR			DHUNIV 1027	GYSSSWY		
DHUNIV 472	WFRE			DHUNIV 1028	GYSSSW		
DHUNIV 473	FREL			DHUNIV 1029	YSSSWY		
DHUNIV 474	RELL			DHUNIV 1030	GYSSS		
DHUNIV 475	WFR			DHUNIV 1031	YSSSW		
DHUNIV 476	FRE			DHUNIV 1032	SSSWY		
DHUNIV 477	REL			DHUNIV 1033	GYSS		
DHUNIV 478	FR			DHUNIV 1034	YSSS		
DHUNIV 479	RE			DHUNIV 1035	SSSW		
DHUNIV 480	ITMVQGVIIIT			DHUNIV 1036	SSWY		
DHUNIV 481	ITMVQGVII			DHUNIV 1037	YSS		
DHUNIV 482	TMVQGVIIIT			DHUNIV 1038	SSS		
DHUNIV 483	ITMVQGVI			DHUNIV 1039	SSW		
DHUNIV 484	TMVQGVII			DHUNIV 1040	SWY		
DHUNIV 485	MVQGVIIIT			DHUNIV 1041	SW		
DHUNIV 486	ITMVQGV			DHUNIV 1042	WY		
DHUNIV 487	TMVQGVI			DHUNIV 1043	GIAAAGT		
DHUNIV 488	MVQGVII			DHUNIV 1044	GIAAAG		

DHUNIV 489	VQGVIIIT		DHUNIV 1045	IAAAGT	
DHUNIV 490	ITMVQG		DHUNIV 1046	GIAAA	
DHUNIV 491	TMVQGV		DHUNIV 1047	IAAAG	
DHUNIV 492	MVQGVI		DHUNIV 1048	AAAGT	
DHUNIV 493	VQGVII		DHUNIV 1049	GIAA	
DHUNIV 494	QGVIIIT		DHUNIV 1050	IAAA	
DHUNIV 495	ITMVQ		DHUNIV 1051	AAAG	
DHUNIV 496	TMVQG		DHUNIV 1052	AAGT	
DHUNIV 497	MVQGV		DHUNIV 1053	GIA	
DHUNIV 498	VQGVI		DHUNIV 1054	IAA	
DHUNIV 499	QGVII		DHUNIV 1055	AAA	
DHUNIV 500	TMVQ		DHUNIV 1056	AAG	
DHUNIV 501	MVQG		DHUNIV 1057	AGT	
DHUNIV 502	VQGV		DHUNIV 1058	AG	
DHUNIV 503	QGVI		DHUNIV 1059	QQLV	
DHUNIV 504	MVQ		DHUNIV 1060	QQL	
DHUNIV 505	VQG		DHUNIV 1061	QLV	
DHUNIV 506	QGV		DHUNIV 1062	QQ	
DHUNIV 507	QG		DHUNIV 1063	GYSSGWY	
DHUNIV 508	LRLGEL		DHUNIV 1064	GYSSGW	
DHUNIV 509	LRLGE		DHUNIV 1065	YSSGWY	
DHUNIV 510	RLGEL		DHUNIV 1066	GYSSG	
DHUNIV 511	LRLG		DHUNIV 1067	YSSGW	
DHUNIV 512	RLGE		DHUNIV 1068	SSGWY	
DHUNIV 513	LGEL		DHUNIV 1069	YSSG	
DHUNIV 514	LRL		DHUNIV 1070	SSGW	
DHUNIV 515	RLG		DHUNIV 1071	SGWY	
DHUNIV 516	LGE		DHUNIV 1072	SGW	
DHUNIV 517	LR		DHUNIV 1073	GWY	
DHUNIV 518	RL		DHUNIV 1074	GW	
DHUNIV 519	LG		DHUNIV 1075	GIAVAGT	
DHUNIV 520	YYDYVWGSYAYT		DHUNIV 1076	GIAVAG	
DHUNIV 521	YYDYVWGSYAY		DHUNIV 1077	IAVAGT	
DHUNIV 522	YDYVWGSYAYT		DHUNIV 1078	GIAVA	
DHUNIV 523	YYDYVWGSYA		DHUNIV 1079	IAVAG	
DHUNIV 524	YDYVWGSYAY		DHUNIV 1080	AVAGT	
DHUNIV 525	DYVWGSYAYT		DHUNIV 1081	GIAV	
DHUNIV 526	YYDYVWGSY		DHUNIV 1082	IAVA	
DHUNIV 527	YDYVWGSYA		DHUNIV 1083	AVAG	
DHUNIV 528	DYVWGSYAY		DHUNIV 1084	VAGT	
DHUNIV 529	YVWGSYAYT		DHUNIV 1085	IAV	

DHUNIV 530	YYDYVWGS		DHUNIV 1086	AVA	
DHUNIV 531	YDYVWGSY		DHUNIV 1087	VAG	
DHUNIV 532	DYVWGSYA		DHUNIV 1088	AV	
DHUNIV 533	YVWGSYAY		DHUNIV 1089	QWL V	
DHUNIV 534	VWGSYAYT		DHUNIV 1090	QWL	
DHUNIV 535	YYDYVWG		DHUNIV 1091	WLV	
DHUNIV 536	YDYVWGS		DHUNIV 1092	QW	
DHUNIV 537	DYVWGSY		DHUNIV 1093	EYSSSS	
DHUNIV 538	YVWGSYA		DHUNIV 1094	EYSSS	
DHUNIV 539	VWGSYAY		DHUNIV 1095	YSSSS	
DHUNIV 540	WGSYAYT		DHUNIV 1096	EYSS	
DHUNIV 541	YYDYVW		DHUNIV 1097	SSSS	
DHUNIV 542	YDYVWG		DHUNIV 1098	EYS	
DHUNIV 543	DYVWGS		DHUNIV 1099	EY	
DHUNIV 544	YVWGSY		DHUNIV 1100	SIAARP	
DHUNIV 545	VWGSYA		DHUNIV 1101	SIAAR	
DHUNIV 546	WGSYAY		DHUNIV 1102	IAARP	
DHUNIV 547	GSYAYT		DHUNIV 1103	SIAA	
DHUNIV 548	YYDYV		DHUNIV 1104	IAAR	
DHUNIV 549	YDYVW		DHUNIV 1105	AARP	
DHUNIV 550	DYVWG		DHUNIV 1106	SIA	
DHUNIV 551	YVWGS		DHUNIV 1107	AAR	
DHUNIV 552	VWGSY		DHUNIV 1108	ARP	
DHUNIV 553	WGSYA		DHUNIV 1109	AR	
DHUNIV 554	GSYAY		DHUNIV 1110	RP	
DHUNIV 555	SYAYT		DHUNIV 1111	NWG	
DHUNIV 556	YYDY				

Table 12. Theoretical segment pool of 141 N2 segments in Theoretical Segment Pool 1 (TSP1).

Segment Type	Sequences	Number
"Zero"	(no addition) V segment joins directly to D segment	1
Monomers	G, P, R, A, S, L, T, V, D, E, F, H, I, K, M, Q, W, Y	18
Dimers	GG, GP, GR, GA, GS, GL, GT, GV, PG, RG, AG, SG, LG, TG, VG, PP, PR, PA, PS, PL, PT, PV, RP, AP, SP, LP, TP, VP, AR, AS, AT, AY, DL, DT, EA, EK, FH, FS, HL, HW, IS, KV, LD, LE, LR, LS, LT, NR, NT, QE, QL, QT, RA, RD, RE, RF, RH, RL, RR, RS, RV, SA, SD, SE, SF, SI, SK, SL, SQ, SR, SS, ST, SV, TA, TR, TS, TT, TW, VD, VS, WS, YS	82
Trimers	GGG, GPG, GRG, GAG, GSG, GLG, GTG, GVG, PGG, RGG, AGG, SGG, LGG, TGG, VGG, GGP, GGR, GGA, GGS, GGL, GGT, GGV, AAE, AYH, DTL, EKR, ISR, NTP, PKS, PRP, PTA, PTQ, REL, RPL, SAA, SAL, SGL, SSE, TGL, WGT	40

Table 13. Theoretical segment pool of 285 H3-JH segments.

H3-JH Segment Name	Amino Acid Sequence	SEQ ID NO	H3-JH Segment Name	Amino Acid Sequence	SEQ ID NO
JHUNIV 001	TEYFQH		JHUNIV 144	ADY	
JHUNIV 002	EYFQH		JHUNIV 145	YDY	
JHUNIV 003	YFQH		JHUNIV 146	NYFDY	
JHUNIV 004	FQH		JHUNIV 147	DNWFDP	
JHUNIV 005	QH		JHUNIV 148	NWFDP	
JHUNIV 006	H		JHUNIV 149	WFDP	
JHUNIV 007			JHUNIV 150	FDP	
JHUNIV 008	SEYFQH		JHUNIV 151	DP	
JHUNIV 009	PEYFQH		JHUNIV 152	P	
JHUNIV 010	FEYFQH		JHUNIV 153	YNWFDP	
JHUNIV 011	HEYFQH		JHUNIV 154	HNWFDP	
JHUNIV 012	REYFQH		JHUNIV 155	FNWFDP	
JHUNIV 013	LEYFQH		JHUNIV 156	SNWFDP	
JHUNIV 014	NEYFQH		JHUNIV 157	RNWFDP	
JHUNIV 015	IYFQH		JHUNIV 158	LNWFDP	
JHUNIV 016	DEYFQH		JHUNIV 159	PNWFDP	
JHUNIV 017	GEYFQH		JHUNIV 160	INWFDP	
JHUNIV 018	VEYFQH		JHUNIV 161	TNWFDP	
JHUNIV 019	YEYFQH		JHUNIV 162	GNWFDP	
JHUNIV 020	KYFQH		JHUNIV 163	VNWFDP	
JHUNIV 021	QYFQH		JHUNIV 164	ANWFDP	
JHUNIV 022	LYFQH		JHUNIV 165	DWFDP	
JHUNIV 023	SYFQH		JHUNIV 166	YWFDP	
JHUNIV 024	RYFQH		JHUNIV 167	HWFDP	
JHUNIV 025	PYFQH		JHUNIV 168	FWFDP	
JHUNIV 026	IYFQH		JHUNIV 169	SWFDP	
JHUNIV 027	TYFQH		JHUNIV 170	RWFDP	
JHUNIV 028	GYFQH		JHUNIV 171	LWFDP	
JHUNIV 029	VYFQH		JHUNIV 172	PWFDP	
JHUNIV 030	AYFQH		JHUNIV 173	IWFDP	
JHUNIV 031	NFQH		JHUNIV 174	TWFPD	
JHUNIV 032	DFQH		JHUNIV 175	GWFPD	

JHUNIV 033	HFQH			JHUNIV 176	VW FDP	
JHUNIV 034	FFQH			JHUNIV 177	AW FDP	
JHUNIV 035	SFQH			JHUNIV 178	RFDP	
JHUNIV 036	RFQH			JHUNIV 179	GFDP	
JHUNIV 037	LFQH			JHUNIV 180	LFDP	
JHUNIV 038	PFQH			JHUNIV 181	SFDP	
JHUNIV 039	IFQH			JHUNIV 182	QFDP	
JHUNIV 040	TFQH			JHUNIV 183	PFDP	
JHUNIV 041	GFQH			JHUNIV 184	KFDP	
JHUNIV 042	VFQH			JHUNIV 185	MFDP	
JHUNIV 043	AFQH			JHUNIV 186	TFDP	
JHUNIV 044	AEYFQH			JHUNIV 187	EFDP	
JHUNIV 045	YWYFDL			JHUNIV 188	VFDP	
JHUNIV 046	WYFDL			JHUNIV 189	AFDP	
JHUNIV 047	YFDL			JHUNIV 190	NNWFDP	
JHUNIV 048	FDL			JHUNIV 191	DYYYYYGMDV	
JHUNIV 049	DL			JHUNIV 192	YYYYYGMDV	
JHUNIV 050	L			JHUNIV 193	YYYYYGMDV	
JHUNIV 051	DWYFDL			JHUNIV 194	YYGMDV	
JHUNIV 052	HWYFDL			JHUNIV 195	YYGMDV	
JHUNIV 053	NWYFDL			JHUNIV 196	YGMDV	
JHUNIV 054	GYFDL			JHUNIV 197	GMDV	
JHUNIV 055	RYFDL			JHUNIV 198	MDV	
JHUNIV 056	HFDL			JHUNIV 199	DV	
JHUNIV 057	NFDL			JHUNIV 200	V	
JHUNIV 058	DFDL			JHUNIV 201	YYYYYYGMDV	
JHUNIV 059	DAFDI			JHUNIV 202	HYYYYYGMDV	
JHUNIV 060	AFDI			JHUNIV 203	FYYYYYGMDV	
JHUNIV 061	FDI			JHUNIV 204	SYYYYYGMDV	
JHUNIV 062	DI			JHUNIV 205	RYYYYYGMDV	
JHUNIV 063	I			JHUNIV 206	LYYYYYGMDV	
JHUNIV 064	YAFDI			JHUNIV 207	PYYYYYGMDV	
JHUNIV 065	HAFDI			JHUNIV 208	IYYYYYGMDV	
JHUNIV 066	FAFDI			JHUNIV 209	TYYYYYGMDV	
JHUNIV 067	SAFDI			JHUNIV 210	GYYYYYGMDV	
JHUNIV 068	RAFDI			JHUNIV 211	VYYYYYGMDV	
JHUNIV 069	LAFDI			JHUNIV 212	AYYYYYGMDV	

JHUNIV 070	PAFDI			JHUNIV 213	NYYYGMDV	
JHUNIV 071	IAFDI			JHUNIV 214	DYYYGMDV	
JHUNIV 072	TAFDI			JHUNIV 215	HYYYGMDV	
JHUNIV 073	GAFDI			JHUNIV 216	FYYYGMDV	
JHUNIV 074	VAFDI			JHUNIV 217	SYYYGMDV	
JHUNIV 075	AAFDI			JHUNIV 218	RYYYGMDV	
JHUNIV 076	TFDI			JHUNIV 219	LYYYGMDV	
JHUNIV 077	SFDI			JHUNIV 220	PYYYGMDV	
JHUNIV 078	PFDI			JHUNIV 221	IYYYGMDV	
JHUNIV 079	FFDI			JHUNIV 222	TYYYGMDV	
JHUNIV 080	HFDI			JHUNIV 223	GYYYGMDV	
JHUNIV 081	RFDI			JHUNIV 224	VYYYGMDV	
JHUNIV 082	LFDI			JHUNIV 225	AYYYGMDV	
JHUNIV 083	NFDI			JHUNIV 226	NYYYGMDV	
JHUNIV 084	IFDI			JHUNIV 227	DYYYGMDV	
JHUNIV 085	DFDI			JHUNIV 228	HYYYGMDV	
JHUNIV 086	GFDI			JHUNIV 229	FYYYGMDV	
JHUNIV 087	VFDI			JHUNIV 230	SYYYGMDV	
JHUNIV 088	YFDI			JHUNIV 231	RYYYGMDV	
JHUNIV 089	IDI			JHUNIV 232	LYYGMDV	
JHUNIV 090	VDI			JHUNIV 233	PYYGMDV	
JHUNIV 091	LDI			JHUNIV 234	IYYGMDV	
JHUNIV 092	SDI			JHUNIV 235	TYYYGMDV	
JHUNIV 093	HDI			JHUNIV 236	GYYGMDV	
JHUNIV 094	RDI			JHUNIV 237	VYYGMDV	
JHUNIV 095	PDI			JHUNIV 238	AYYGMDV	
JHUNIV 096	NDI			JHUNIV 239	NYYYYGMDV	
JHUNIV 097	TDI			JHUNIV 240	DYYYYMDV	
JHUNIV 098	DDI			JHUNIV 241	YYYYYMDV	
JHUNIV 099	GDI			JHUNIV 242	YYYYMDV	
JHUNIV 100	ADI			JHUNIV 243	YYYMDV	
JHUNIV 101	YDI			JHUNIV 244	YYMDV	
JHUNIV 102	NAFDI			JHUNIV 245	YMDV	
JHUNIV 103	DYFDY			JHUNIV 246	YMDV	
JHUNIV 104	YFDY			JHUNIV 247	YYYYYMDV	
JHUNIV 105	FDY			JHUNIV 248	HYYYYYMDV	
JHUNIV 106	DY			JHUNIV 249	FYYYYMDV	

JHUNIV 107	Y			JHUNIV 250	SYYYYYYMDV	
JHUNIV 108	YYFDY			JHUNIV 251	RYYYYYYMDV	
JHUNIV 109	HYFDY			JHUNIV 252	LYYYYYYMDV	
JHUNIV 110	FYFDY			JHUNIV 253	PYYYYYYMDV	
JHUNIV 111	SYFDY			JHUNIV 254	IYYYYYYMDV	
JHUNIV 112	RYFDY			JHUNIV 255	TYYYYYYMDV	
JHUNIV 113	LYFDY			JHUNIV 256	GYYYYYYMDV	
JHUNIV 114	PYFDY			JHUNIV 257	VYYYYYYMDV	
JHUNIV 115	IYFDY			JHUNIV 258	AYYYYYYMDV	
JHUNIV 116	TYFDY			JHUNIV 259	NYYYYYYMDV	
JHUNIV 117	GYFDY			JHUNIV 260	DYYYYYYMDV	
JHUNIV 118	VYFDY			JHUNIV 261	HYYYYYYMDV	
JHUNIV 119	AYFDY			JHUNIV 262	FYYYYYYMDV	
JHUNIV 120	NFDY			JHUNIV 263	SYYYYYYMDV	
JHUNIV 121	DFDY			JHUNIV 264	RYYYYYYMDV	
JHUNIV 122	HFDY			JHUNIV 265	LYYYYYYMDV	
JHUNIV 123	FFDY			JHUNIV 266	PYYYYYYMDV	
JHUNIV 124	SFDY			JHUNIV 267	IYYYYYYMDV	
JHUNIV 125	RFDY			JHUNIV 268	TYYYYYYMDV	
JHUNIV 126	LFDY			JHUNIV 269	GYYYYYYMDV	
JHUNIV 127	PFDY			JHUNIV 270	VYYYYYYMDV	
JHUNIV 128	IFDY			JHUNIV 271	AYYYYYYMDV	
JHUNIV 129	TFDY			JHUNIV 272	NYYYYYYMDV	
JHUNIV 130	GFDY			JHUNIV 273	DYYYYYYMDV	
JHUNIV 131	VFDY			JHUNIV 274	HYYYYYYMDV	
JHUNIV 132	AFDY			JHUNIV 275	FYYYYYYMDV	
JHUNIV 133	IDY			JHUNIV 276	SYYYYYYMDV	
JHUNIV 134	VDY			JHUNIV 277	RYYYYYYMDV	
JHUNIV 135	LDY			JHUNIV 278	LYYYYYYMDV	
JHUNIV 136	SDY			JHUNIV 279	PYYYYYYMDV	
JHUNIV 137	HDY			JHUNIV 280	IYYYYYYMDV	
JHUNIV 138	RDY			JHUNIV 281	TYYYYYYMDV	
JHUNIV 139	PDY			JHUNIV 282	GYYYYYYMDV	
JHUNIV 140	NDY			JHUNIV 283	VYYYYYYMDV	
JHUNIV 141	TDY			JHUNIV 284	AYYYYYYMDV	
JHUNIV 142	DDY			JHUNIV 285	NYYYYYYMDV	
JHUNIV 143	GDY					

Table 14. Twelve germline IGHJ genes and alleles.

IGHJ Gene	DNA Sequence	SEQ ID NO
IGHJ1 - 01	GCTGAATACTTCCAGCACTGGGGCCAGGGCACCCCTGGTCAACCGTCTCCTCAG	
IGHJ2 - 01	CTACTGGTACTTCGATCTCTGGGGCCGTGGCACCCCTGGTCACTGTCTCCTCAG	
IGHJ3 - 01	ATGCTTTTGATGTCTGGGGCCAAAGGGACAAATGGTCAACCGTCTCTTCAG	
IGHJ3 - 02	ATGCTTTTGATAATCTGGGGCCAAAGGGACAAATGGTCAACCGTCTCTTCAG	
IGHJ4 - 01	ACTACTTTTGACTACTGGGGCCAAAGGAACCCCTGGTCAACCGTCTCCTCAG	
IGHJ4 - 02	ACTACTTTTGACTACTGGGGCCAGGGAACCCCTGGTCAACCGTCTCCTCAG	
IGHJ4 - 03	GCTACTTTTGACTACTGGGGCCAAAGGGACCCCTGGTCAACCGTCTCCTCAG	
IGHJ5 - 01	ACAACTGGTTCGACTCCTGGGGCCAAAGGAACCCCTGGTCAACCGTCTCCTCAG	
IGHJ5 - 02	ACAACTGGTTCGACCCCTGGGGCCAGGGAACCCCTGGTCAACCGTCTCCTCAG	
	ATTACTACTACTACTACGGTATGGACGTCTGGGGGCAAGGGACCAACCGTCTCT	
IGHJ6 - 01	CCTCAG	
	ATTACTACTACTACTACGGTATGGACGTCTGGGGCCAAAGGGACCAACCGTCTCT	
IGHJ6 - 02	CCTCAG	
	ATTACTACTACTACTACTACTATGGACGTCTGGGGGCAAGGGACCAACCGTCTCT	
IGHJ6 - 03	CCTCAG	

Table 15. Theoretical segment pool of 248 parent H3-JH segments.

H3-JH Parent Segment Name	Amino Acid Sequence	SEQ ID NO	H3-JH Parent Segment Name	Amino Acid Sequence	SEQ ID NO
JHparent001	ADI		JHparent125	TAFDI	
JHparent002	ADY		JHparent126	TWFDP	
JHparent003	DDI		JHparent127	TYFDY	
JHparent004	DDY		JHparent128	TYFQH	
JHparent005	GDI		JHparent129	VAFDI	
JHparent006	GDY		JHparent130	VWFDP	
JHparent007	HDI		JHparent131	VYFDY	
JHparent008	HDY		JHparent132	VYFQH	
JHparent009	IDI		JHparent133	WYFDL	
JHparent010	IDY		JHparent134	YAFDI	
JHparent011	LDI		JHparent135	YWFDP	
JHparent012	LDY		JHparent136	YYFDY	
JHparent013	NDI		JHparent137	AEYFQH	
JHparent014	NDY		JHparent138	ANWFDP	
JHparent015	PDI		JHparent139	DEYFQH	
JHparent016	PDY		JHparent140	DNWFDP	
JHparent017	RDI		JHparent141	DWYFDL	
JHparent018	RDY		JHparent142	FEYFQH	
JHparent019	SDI		JHparent143	FNWFDP	
JHparent020	SDY		JHparent144	GEYFQH	
JHparent021	TDI		JHparent145	GNWFDP	
JHparent022	TDY		JHparent146	HEYFQH	
JHparent023	VDI		JHparent147	HNWFDP	
JHparent024	VDY		JHparent148	HWYFDL	
JHparent025	YDI		JHparent149	IYFQH	
JHparent026	YDY		JHparent150	INWFDP	
JHparent027	AFDP		JHparent151	LEYFQH	
JHparent028	AFDY		JHparent152	LNWFDP	
JHparent029	AFQH		JHparent153	NEYFQH	
JHparent030	DFDI		JHparent154	NNWFDP	
JHparent031	DFDL		JHparent155	NWYFDL	
JHparent032	DFDY		JHparent156	PEYFQH	
JHparent033	DFQH		JHparent157	PNWFDP	
JHparent034	EFDP		JHparent158	REYFQH	
JHparent035	FFDI		JHparent159	RNWFDP	

JHparent036	FFDY		JHparent160	SEYFQH	
JHparent037	FFQH		JHparent161	SNWFDP	
JHparent038	GFDI		JHparent162	TEYFQH	
JHparent039	GFDP		JHparent163	TNWFDP	
JHparent040	GFDY		JHparent164	VEYFQH	
JHparent041	GFQH		JHparent165	VNWFDP	
JHparent042	HFDI		JHparent166	YEYFQH	
JHparent043	HFDL		JHparent167	YNWFDP	
JHparent044	HFDY		JHparent168	YWYFDL	
JHparent045	HFQH		JHparent169	AYYYGMDV	
JHparent046	IFDI		JHparent170	AYYYYMDV	
JHparent047	IFDY		JHparent171	DYYYGMDV	
JHparent048	IFQH		JHparent172	DYYYYMDV	
JHparent049	KFDP		JHparent173	FYYYGMDV	
JHparent050	LFDI		JHparent174	FYYYYMDV	
JHparent051	LFDP		JHparent175	GYYYGMDV	
JHparent052	LFDY		JHparent176	GYYYYMDV	
JHparent053	LFQH		JHparent177	HYYYGMDV	
JHparent054	MFDP		JHparent178	HYYYYMDV	
JHparent055	NFDI		JHparent179	IYYYGMDV	
JHparent056	NFDL		JHparent180	IYYYYMDV	
JHparent057	NFDY		JHparent181	LYYYGMDV	
JHparent058	NFQH		JHparent182	LYYYYMDV	
JHparent059	PFDI		JHparent183	NYYYGMDV	
JHparent060	PFDP		JHparent184	NYYYYMDV	
JHparent061	PFDY		JHparent185	PYYYGMDV	
JHparent062	PFQH		JHparent186	PYYYYMDV	
JHparent063	QFDP		JHparent187	RYYYGMDV	
JHparent064	RFDI		JHparent188	RYYYYMDV	
JHparent065	RFDP		JHparent189	SYYYGMDV	
JHparent066	RFDY		JHparent190	SYYYYMDV	
JHparent067	RFQH		JHparent191	TYYYGMDV	
JHparent068	SFDI		JHparent192	TYYYYMDV	
JHparent069	SFDP		JHparent193	VYYYGMDV	
JHparent070	SFDY		JHparent194	VYYYYMDV	
JHparent071	SFQH		JHparent195	AYYYGMDV	
JHparent072	TFDI		JHparent196	AYYYYMDV	
JHparent073	TFDP		JHparent197	DYYYGMDV	
JHparent074	TFDY		JHparent198	DYYYYMDV	
JHparent075	TFQH		JHparent199	FYYYGMDV	
JHparent076	VFDI		JHparent200	FYYYYMDV	

JHparent077	VFDP		JHparent201	GYYYGMDV	
JHparent078	VFDY		JHparent202	GYYYGMDV	
JHparent079	VFQH		JHparent203	HYYYGMDV	
JHparent080	YFDI		JHparent204	HYYYGMDV	
JHparent081	YFDL		JHparent205	IYYYGMDV	
JHparent082	AAFDI		JHparent206	IYYYGMDV	
JHparent083	AWFDP		JHparent207	LYYYGMDV	
JHparent084	AYFDY		JHparent208	LYYYGMDV	
JHparent085	AYFQH		JHparent209	NYYYGMDV	
JHparent086	DAFDI		JHparent210	NYYYGMDV	
JHparent087	DWFDP		JHparent211	PYYYGMDV	
JHparent088	DYFDY		JHparent212	PYYYGMDV	
JHparent089	FAFDI		JHparent213	RYYYGMDV	
JHparent090	FWFDP		JHparent214	RYYYGMDV	
JHparent091	FYFDY		JHparent215	SYYYGMDV	
JHparent092	GAFDI		JHparent216	SYYYGMDV	
JHparent093	GWFDP		JHparent217	TYYYGMDV	
JHparent094	GYFDL		JHparent218	TYYYGMDV	
JHparent095	GYFDY		JHparent219	VYYYGMDV	
JHparent096	GYFQH		JHparent220	VYYYGMDV	
JHparent097	HAFDI		JHparent221	AYYYGMDV	
JHparent098	HWFDP		JHparent222	AYYYGMDV	
JHparent099	HYFDY		JHparent223	DYYYGMDV	
JHparent100	IAFDI		JHparent224	DYYYGMDV	
JHparent101	IWFDP		JHparent225	FYYYGMDV	
JHparent102	IYFDY		JHparent226	FYYYGMDV	
JHparent103	IYFQH		JHparent227	GYYYGMDV	
JHparent104	KYFQH		JHparent228	GYYYGMDV	
JHparent105	LAFDI		JHparent229	HYYYGMDV	
JHparent106	LWFDP		JHparent230	HYYYGMDV	
JHparent107	LYFDY		JHparent231		

					IYYYYYGM DV	
JHparent108	LYFQH			JHparent232	IYYYYYYMDV	
JHparent109	NAFDI			JHparent233	LYYYYYGM DV	
JHparent110	NYFDY			JHparent234	LYYYYYYYMDV	
JHparent111	PAFDI			JHparent235	NYYYYYGM DV	
JHparent112	PWFDP			JHparent236	NYYYYYYMDV	
JHparent113	PYFDY			JHparent237	PYYYYYGM DV	
JHparent114	PYFQH			JHparent238	PYYYYYYMDV	
JHparent115	QYFQH			JHparent239	RYYYYYGM DV	
JHparent116	RAFDI			JHparent240	RYYYYYYMDV	
JHparent117	RWFDP			JHparent241	SYYYYYGM DV	
JHparent118	RYFDL			JHparent242	SYYYYYYMDV	
JHparent119	RYFDY			JHparent243	TYYYYYGM DV	
JHparent120	RYFQH			JHparent244	TYYYYYYMDV	
JHparent121	SAFDI			JHparent245	VYYYYYGM DV	
JHparent122	SWFDP			JHparent246	VYYYYYYMDV	
JHparent123	SYFDY			JHparent247	YYYYYYGM DV	
JHparent124	SYFQH			JHparent248	YYYYYYYYMDV	

Table 16. Polynucleotide sequences of 27 human IGHD genes and alleles.

IGHD Gene	Polynucleotide Sequence	SEQ ID NO
IGHD1-(1)-01	GGTACAACTGGAACGAC	
IGHD1-20	GGTATAACTGGAACGAC	
IGHD1-26	GGTATAGTGGGAGCTACTAC	
IGHD1-7	GGTATAACTGGAACCTAC	
IGHD2-15-01	AGGATATTGTAGTGGTGGTAGCTGCTACTCC	
IGHD2-2-x	AGGATATTGTAGTAGTACCAGCTGCTATGCC	
IGHD2-2-y	AGGATATTGTAGTAGTACCAGCTGCTATACC	
IGHD2-2-z	TGGATATTGTAGTAGTACCAGCTGCTATGCC	
IGHD2-21-01	AGCATATTGTGGTGCTGATTGCTATTCC	
IGHD2-21-02	AGCATATTGTGGTGCTGACTGCTATTCC	
IGHD2-8-01	AGGATATTGTACTAATAAGGTGTATGCTATACC	
IGHD3-10-01	GTATTACTATGTTTCGGGAGTTATTATAAC	
IGHD3-10-03	GTATTACTATGTTTCAGGGAGTTATTATAAC	
IGHD3-16-02	GTATTATGATTACGTTTGGGGAGTTATGCTTATACC	
IGHD3-22-01	GTATTACTATGATAGTAGTGTTATTACTAC	
IGHD3-3-01	GTATTACGATTTTGGAGTGGTTATTATACC	
IGHD3-9-01	GTATTACGATAATTTGACTGGTTATTATAAC	
IGHD4-17	TGACTACGGTGACTAC	
IGHD4-23-01	TGACTACGGTGGTAACTCC	
IGHD4-4/11-01	TGACTACAGTAACTAC	
IGHD5-12-01	GTGGATATAGTGGCTACGATTAC	
IGHD5-24-01	GTAGAGATGGCTACAAATTAC	
IGHD5-5/18-01	GTGGATACAGCTATGGTTAC	
IGHD6-13-01	GGTATAGCAGCAGCTGGTAC	
IGHD6-19-01	GGTATAGCAGTGGCTGGTAC	
IGHD6-6-01	GAGTATAGCAGCTCGTCC	
IGHD7-27-01	CTAACTGGGGA	

Table 17. Theoretical segment pool of 73 DH parent Segments. "Z" represents a stop codon.

DH Parent Segment Name	Amino Acid Sequence	SEQ ID NO	DH Parent Segment Name	Amino Acid Sequence	SEQ ID NO
DHparent001	LTG		DHparent038	WIZWLRL	
DHparent002	NWG		DHparent039	VDIVATIT	
DHparent003	ZLG		DHparent040	AYCGDCYS	
DHparent004	DYGDY		DHparent041	HIVVZIAIP	
DHparent005	DYSNY		DHparent042	HIVVVTaip	
DHparent006	TTVTT		DHparent043	SILWZLLF	
DHparent007	VQLER		DHparent044	DIVLMVYAIP	
DHparent008	VZLEL		DHparent045	DIVVVPAAIP	
DHparent009	VZLER		DHparent046	DIVVVPAAAMP	
DHparent010	VZQLV		DHparent047	DIVVVVAATP	
DHparent011	YNWND		DHparent048	GYCSGGSCYS	
DHparent012	YNWNY		DHparent049	GYCSSTSCYA	
DHparent013	ZLQZL		DHparent050	GYCSSTSCYT	
DHparent014	ZLRZL		DHparent051	GYCTNGVCYT	
DHparent015	DYGGNS		DHparent052	ITIFGVVIIP	
DHparent016	EYSSSS		DHparent053	ITIFZLVIIT	
DHparent017	GITGTT		DHparent054	ITMIVVVIIT	
DHparent018	GTTGTT		DHparent055	ITMVQGVIIIT	
DHparent019	GYSYGY		DHparent056	ITMVRGVIIIT	
DHparent020	RDGYNY		DHparent057	RILYZWCMLY	
DHparent021	SIAARP		DHparent058	RILZWZLLL	
DHparent022	TTVVTP		DHparent059	RILZZYQLLC	
DHparent023	VZQQLV		DHparent060	RILZZYQLLY	
DHparent024	VZQWLv		DHparent061	VLLWFGEllZ	
DHparent025	VZWELL		DHparent062	VLLWFRellZ	
DHparent026	WlQLWL		DHparent063	VLLZZZWLLL	
DHparent027	YSGSYy		DHparent064	VLRFLewLLY	
DHparent028	ZLRWZL		DHparent065	VLRYFDWLLZ	
DHparent029	ZRWLQL		DHparent066	WILZZYQLLC	
DHparent030	GIAAAGT		DHparent067	YYDFWSGYT	
DHparent031	GIAVAGT		DHparent068	YYDILTGYN	
DHparent032	GIVGATT		DHparent069	YYDSSGYyY	
DHparent033	GYSGYDY		DHparent070	YYGSGSYYN	
DHparent034	GYSSGWY		DHparent071	IMITFGVMLIP	
DHparent035	GYSSSWY		DHparent072	VLZLRlGELCLY	

DHparent036	VDTAMVT		DHparent073	YYDYVWGSYAYT	
DHparent037	VEMATIT				

Table 18. Application of Equation 1 to Test Case 1.

Type	Segment	Weight
TN1	R	1.0
DH	TA	1.0
N2	H	1.0
H3 - JH	HFDY	1.0

Table 19. Application of Equation 1 to Test Cases 2.1 and 2.1.

Type	Segment	Weight
TN1	V and VG	0.5 each
DH	GIVGA and IVGA	0.5 each
N2	AS	1.0
H3 - JH	Y	1.0

Table 20. Application of Equation 1 to Test Case 3.1.

Type	Segment	Weight
TN1	DR	1.0
DH	YSGYD	0.8
N2	LG	1.0
H3 - JH	Y	1.0

Table 21. Application of Equation 1 to Test Cases 4.1 and 4.2.

Type	Segment	Weight
TN1	" - " and G	0.5 each
DH	GIAAA and IAAA	0.5 each
N2	D	1.0
H3 - JH	SNWFDP	0.83

Table 22. Application of Equation 1 to all Test Cases.

Type	Segments	Weight
TN1	DR, R, VG, V, G and "- "	0.25, 0.25, 0.125, 0.125, 0.125 and 0.125 respectively
DH	TA, YSGYD, IAAA, GIAAA, IVGA and GIVGA	0.25, 0.20, 0.125, 0.125, 0.125 and 0.125 respectively
N2	AS, H, D and LG	0.25 each
H3 -JH	Y, HFDY and SNWFDP	0.50, 0.25 and 0.209

Table 23. Segments used in Exemplary Library Design 1 (ELD-1). The sequences collectively form a theoretical segment pool that comprises individual theoretical segment pools of TN1, DH, N2, and H3-JH segments.

Segment No.	TN1	SEQ ID NO	DH	SEQ ID NO	N2	SEQ ID NO	H3-JH	SEQ ID NO
1	-		YYDSSGGY		-		DAFDI	
2	G		YGDY		Y		YYFDY	
3	D		DYGDY		G		Y	
4	A		YYDSSGY		D		FDY	
5	V		YCSSTSCY		S		DY	
6	DR		YDSSGY		P		YGMDV	
7	S		YCSGGSCY		F		-	
8	L		YCSSTSC		L		LDY	
9	DL		GG		A		GAFDI	
10	R		RG		E		YFDY	
11	GR		SGSY		V		YYGMDV	
12	T		YYDSSGGY		H		AFDI	
13	GG		SS		T		PFDY	
14	E		YDFWSGY		R		GMDV	
15	DS		GYCSSTSC		W		GW FDP	
16	VG		DY		SG		YYYYGMDV	
17	DG		YYDSSG		I		IDY	
18	AP		CSSTSCY		RG		GYFDY	
19	GL		YYDFWSGY		K		GFDY	
20	GS		SSGWY		LG		MDV	
21	DRG		AG		Q		VDY	
22	DLG		DSSGY		GP		YYYGMDV	
23	VP		SSSW		PG		NWFDP	
24	DP		VGAT		LP		PDY	
25	P		SY		AG		WFDP	
26	EG		DTAM		GS		NFDY	
27	GA		IAAAG		TS		YWYFDL	
28	AG		YSSSW		SS		NAFDI	
29	GV		GS		GG		HFDY	
30	GP		YYDSSG		YS		SFDY	
31	ER		VG		M		YYYYGMDV	
32	DV		YSSSWY		SL		DYYGMDV	
33	VGG		YCSGGSC		SP		DFDY	
34	SG		YDSSGGY		SD		YNWFDP	
35	GRG		GI		AP		DYYYGMDV	

36	DA			GYCSGGSCY		GR		YYYMDV
37	DRP			YSSS		TG		LFDY
38	DSG			SSGW		SR		DYFDY
39	GPR			TA		LD		NYYYGMDV
40	DT			DSSGY		LS		GDY
41	GGG			GYCSSTSCY		GA		YDY
42	DRGG			T'TV'T		VG		SYFDY
43	PL			YSSGWY		PP		YYMDV
44	DPS			GW		RR		TFDY
45	LP			LG		GSG		YYYYYYGMDV
46	RG			DYGD		GT		FFDY
47	GT			TV'TT		TP		SYYYGMDV
48	LG			AAA		RP		VFDY
49	DLP			YSSGW		RD		YAFDI
50	DGR			LV		QL		SDY
51	ERG			YYDFWSGY		TT		WYFDL
52	DGS			YYDSSGY		PL		DWFDP
53	ES			YYDFWSG		RS		AFDY
54	PS			YGD		WS		PYYYGMDV
55	GS			YG		RV		HYFDY
56	DPR			GT		RF		DV
57	EA			YSGSY		SF		RFDY
58	GGR			YYYSGSY		PT		NYFDY
59	DGG			LR		PS		IFDY
60	SP			SSS		AT		ADY
61	DPG			GD		RL		HYYGMDV
62	DSGG			CSGGSCY		SV		GYYYGMDV
63	DPL			GY		RA		DWYFDL
64	TP			YCGDCY		GRG		GYYGMDV
65	AGG			QWLV		VP		YYYMDV
66	PR			IAAA		SQ		V
67	DGT			QG		AS		FDP
68	GLG			YCSSTSCYT		PR		DDY
69	DSP			SG		VD		GYYYGMDV
70	GGV			T'TV'TT		GV		YMDV
71	GPS			QQL		LT		NWYFDL
72	GVG			IAVA		TR		PYFDY
73	GGA			GDY		VS		FDI
74	RP			YYDSSGY		SGL		NDY
75	DPP			DGYN		RGG		HYYYGMDV
76	DPT			CSSTSC		LR		VYYGMDV

77	EV		IVVVPAAI		SGG		VYFDY	
78	GAP		AVAG		SA		SNWFDP	
79	DGRG		AA		AY		NYYGMDV	
80	GPP		DSSG		LE		LYYYGMDV	
81	DLGG		VR		TGG		DYYYYYGM DV	
82	ET		YDFWSG		GL		AAFDI	
83	DGGP		GYSSSWY		GAG		SWFDP	
84	DGGR		SGW		TA		SYYYYYGM DV	
85	DGSG		WG		IS		SYYYGM DV	
86	GTG		SGSY		ST		P	
87	DPGG		RY		SSE		YYYYYMDV	
88	DGGS		DS		GLG		RAFDI	
89	GGT		TT		PRP		RYFDY	
90	GGRG		YCSGGSCYS		RPL		FYYYYGM DV	
91	EGR		PA		EA		GNWFDP	
92	DGA		IVVVPAA		LGG		RWFDP	
93	DGL		IAAAGT		FS		GFDP	
94	EL		AAAG		DL		DYYYYMDV	
95	EGG		GYCSGGSC		GGG		PYYGM DV	
96	LGG		SGWY		AR		DP	
97	DPA		DYGN		RE		HAFDI	
98	EGV		YYYS		PA		DNWFDP	
99	GSG		YSSG		PTQ		LNWFDP	
100	GGP		YDSSGY		AGG		AYYYGM DV	
101			VAG		GGV			
102			SSW		SI			
103			GSGSY		SAA			
104			SSSS		GGG			
105			NW		GGA			
106			DFWSGY		GPG			
107			QQLV		PGG			
108			YGGN		AAE			
109			YDSSG		GGR			
110			GYSYG		TW			
111			TV		GGP			
112			NG		GGL			
113			IVGAT		VGG			
114			IVGA		GTG			
115			YSGSY		NR			
116			SSWY		NTP			
117			ST		PV			

118				DFWSGYY		EK			
119				GSY		GVG			
120				YYDSS		KV			
121				VGA		EKR			
122				AT		QT			
123				RP		SE			
124				YYGSGS		SAL			
125				GIAAAG		FH			
126				SGY		RH			
127				TG		PTA			
128				LT		HL			
129				RD		WGT			
130				WEL		REL			
131				YSYG		NT			
132				TVT		HW			
133				GYCSGGSCYS		PKS			
134				AR		DT			
135				YYGSGSY		DTL			
136				RW		SK			
137				DIVVVPA		TGL			
138				YSGS		AYH			
139				GYSSSW		ISR			
140				YSSSS		GGT			
141				YYYDSS		QE			
142				QL					
143				GYSGYD					
144				GE					
145				MA					
146				DSS					
147				RF					
148				DTAMV					
149				YYGSGSY					
150				VDTAMV					
151				FGVV					
152				EYSSS					
153				TTV					
154				SWY					
155				IAARP					
156				VE					
157				SIAA					
158				YSGYD					

Page 229 of 420

Page 230 of 420

Table 24. Segments used in Exemplary Library Design 2 (ELD-2). The sequences collectively form a theoretical segment pool that comprises individual theoretical segment pools of TN1, DH, N2, and H3-JH segments.

Segment No.	TN1	SEQ ID NO	DH	SEQ ID NO	N2	SEQ ID NO	H3-JH	SEQ ID NO
1	-		YYDSSGGY		-		DAFDI	
2	G		YGDY		G		YYFDY	
3	D		DYGDY		D		Y	
4	A		YYDSSGY		A		FDY	
5	V		YCSSTSCY		V		DY	
6	S		YDSSGY		S		YGMDV	
7	DR		YCSGGSCY		DR		-	
8	L		YCSSTSC		L		LDY	
9	R		GG		R		GAFDI	
10	DL		RG		DL		YFDY	
11	T		SGSY		T		YYGMDV	
12	E		YYDSSGY		E		AFDI	
13	GR		SS		GR		PFDY	
14	GG		YDFWSGY		GG		GMDV	
15	DG		GYCSSTSC		DG		GW FDP	
16	DS		DY		DS		YYYYGMDV	
17	VG		YYDSSG		VG		IDY	
18	EG		CSSTSCY		EG		GYFDY	
19	P		YYDFWSGY		P		GFDY	
20	GL		SSGWY		GL		MDV	
21	GS		AG		GS		VDY	
22	DP		DSSGY		DP		YYYYGMDV	
23	GP		SSSW		GP		NWFDP	
24	GA		VGAT		GA		PDY	
25	GV		SY		GV		WFDP	
26	H		DTAM		H		NFDY	
27	DRG		IAAAG		DRG		YWYFDL	
28	DQ		YSSSW		DQ		NAFDI	
29	AG		GS		AG		HFDY	
30	DLG		YDSSG		DLG		SFDY	
31	DV		VG		DV		YYYYYGM DV	
32	Q		YSSSWY		Q		DYYYGMDV	
33	N		YCSGGSC		N		DFDY	
34	AP		YDSSGY		AP		YNWFDP	
35	GGG		GI		GGG		DYYYGMDV	

36	DH		GYCSGGSCY		DH		YYYMDV
37	VP		YSSS		VP		LFDY
38	SG		SSGW		SG		DYFDY
39	GRG		TA		GRG		NYYYGMDV
40	AR		DSSGY		AR		GDY
41	RG		GYCSSTSCY		RG		YDY
42	ER		TTVT		ER		SYFDY
43	DA		YSSGWY		DA		YYMDV
44	AS		GW		AS		TFDY
45	PL		LG		PL		YYYYYYGMDV
46	DQG		DYGD		DQG		FFDY
47	VL		TVTT		VL		SYYYGMDV
48	GT		AAA		GT		VFDY
49	DGG		YSSGW		DGG		YAFDI
50	DSG		LV		DSG		SDY
51	VGG		YDFWSGY		VGG		WYFDL
52	F		YYDSSGY		F		DWFDP
53	AL		YDFWSG		AL		AFDY
54	PS		YGD		PS		PYYYGMDV
55	ES		YG		ES		HYFDY
56	ERG		GT		ERG		DV
57	GGV		YSGY		GGV		RFDY
58	DRP		YYYSGSY		DRP		NYFDY
59	EA		LR		EA		IFDY
60	TP		SSS		TP		ADY
61	GPR		GD		GPR		HYYGMDV
62	LH		CSGGSCY		LH		GYYYGMDV
63	SR		GY		SR		DWYFDL
64	LP		YCGDCY		LP		GYYGMDV
65	LG		QWLV		LG		YYYMDV
66	DT		IAAA		DT		V
67	VA		QG		VA		FDP
68	SL		YCSSTSCYT		SL		DDY
69	EGG		SG		EGG		GYYYGMDV
70	DRS		TTVT		DRS		YMDV
71	K		QQL		K		NWYFDL
72	DPG		IAVA		DPG		PYFDY
73	I		GDY		I		FDI
74	GD		YYDSSGY		GD		NDY
75	DGT		DGYN		DGT		HYYYGMDV
76	GPP		CSSTSC		GPP		VYYGMDV

77	DPP		IVVPPAAI		DPP		VYFDY	
78	RR		AVAG		RR		SNWFDP	
79	EGV		AA		EGV		NYYYGMDV	
80	GF		DSSG		GF		LYYYGMDV	
81	GVG		VR		GVG		DYYYYYGMDV	
82	DPS		YDFWSG		DPS		AAFDI	
83	VD		GYSSSWY		VD		SWFDP	
84	GGT		SGW		GGT		SYYYYYGMDV	
85	DK		WG		DK		SYYYGMDV	
86	GTG		SGSYY		GTG		P	
87	DF		RY		DF		YYYYYMDV	
88	GQ		DS		GQ		RAFDI	
89	SP		TT		SP		RYFDY	
90	QG		YCSGGSCYS		QG		FYYYYGMDV	
91	DLT		PA		DLT		GNWFDP	
92	AK		IVVPPAA		AK		RWFDP	
93	GPS		IAAAGT		GPS		GFDP	
94	QR		AAAG		QR		DYYYYMDV	
95	VR		GYCSGGSC		VR		PYYGMDV	
96	DSP		SGWY		DSP		DP	
97	DPL		DYGGN		DPL		HAFDI	
98	EGR		YYYDS		EGR		DNWFDP	
99	GRRG		YSSG		GRRG		LNWFDP	
100	EV		YDSSGY		EV		AYYYGMDV	
101			VAG		RP			
102			SSW		GH			
103			GSGSY		DGR			
104			SSSS		AA			
105			NW		DD			
106			DFWSGY		W			
107			QQLV		GGG			
108			YGGN		DIS			
109			YDSSG		GGA			
110			GYSYG		GK			
111			TV		DGP			
112			NG		DLK			
113			IVGAT		ET			
114			IVGA		TT			
115			YSGGSY		VH			
116			SSWY		AE			
117			ST		VS			

118				DFWSGYY		LGG			
119				GSY		C			
120				YDSS		DKG			
121				VGA		HA			
122				AT		VI			
123				RP		HP			
124				YYGSGS		GGE			
125				GIAAAG		EP			
126				SGY		EF			
127				TG		DRN			
128				LT		DWG			
129				RD		GE			
130				WEL		DRA			
131				YSYG		VN			
132				TVT		DRE			
133				GYCSGGSCYS		DLA			
134				AR		EN			
135				YGSYSY		VT			
136				RW		HG			
137				DIVVPA		RA			
138				YSGS		M			
139				GYSSSW		DVP			
140				YSSSS		GAP			
141				YYDSS		GLG			
142				QL		GPG			
143				GYSGYD		PG			
144				GE		DSS			
145				MA		SS			
146				DSS		AGG			
147				RF		GGR			
148				DTAMV		GPN			
149				YYGSGSY		DRL			
150				VDTAMV		GRR			
151				FGVV		DSGG			
152				EYSSS		TR			
153				TTV		DLS			
154				SWY		RGG			
155				IAARP		Y			
156				VE		EVR			
157				SIAA		LI			
158				YSGYD		TF			

159				DIVVPA		LK			
160				CSGGSC		DLE			
161				DW		GY			
162				TS		DGS			
163				RL		GVR			
164				YSS		GQR			
165				GN		EGL			
166				SN		VLG			
167				GYSY		QP			
168				YYS		VM			
169				VDTAM		VE			
170				LE		DQGG			
171				AVAGT		PN			
172				YSY		DGL			
173				SW		PV			
174				SSG		HR			
175				FGV		AD			
176				VP		DLF			
177				VA		LD			
178				SY		GGD			
179				QWL		DRR			
180				GSG		DHH			
181				TIFGVV		DW			
182				AVA		DAS			
183				FWSGY		GW			
184				YSGSY		SV			
185				IAVAG		GLR			
186				YS		DGA			
187				YQL		LA			
188				SIAAR		EEG			
189				YCGGDC		AV			
190				NWNY		VQ			
191				SSSWY		AH			
192				GIABA		RS			
193				YSYGY		WA			
194				GIAAA		LR			
195				YYG		GSG			
196				AAG		GGSG			
197				AV		DLR			
198				AYCGDCY		VWG			
199				YGGSGS		HL			

Page 236 of 420

Table 25. Segments used in Exemplary Library Design 3 (ELD-3). The sequences collectively form a theoretical segment pool that comprises individual theoretical segment pools of TN1, DH, N2, and H3-JH segments.

Segm ent No.	TN1 (plus AR or AK) ¹	SEQ ID NO	TN1 Nucle otides (plus AR or AK) ¹	SEQ ID NO	DH	DH Nucleoti des	SEQ ID NO	N2	N2 Nucleotides	SEQ ID NO	H3-JH	H3-JH Nucleotides	SEQ ID NO
1	AR		GCCA GA		GT	GGTACT		-	-		-	-	
2	ARE		GCCA GAGA G		TT	ACTACT		A	GCT		Y	TAT	
3	ARD		GCCA GAGA C		TG	ACAGGC		D	GAT		DI	GATATT	
4	ARG		GCCA GAGG A		ER	GAGCGT		E	GAG		DL	GACTTG	
5	AREG		GCCA GAGA GGGA		QLE	CAATTAGA G		F	TTC		DP	GATCCT	
6	ARDG		GCCA GAGA CGGA		LER	TTAGAGCG T		G	GGC		DV	GATGTA	
7	ARGG		GCCA GAGG TGGA		VGAT	GTTGGCGC AACT		H	CAT		DY	GACTAT	
8	ARR		GCCA GGAG A		YSG	TATAGTGG T		I	ATC		QH	CAGCAC	
9	ARER		GCCA GAGA GAGA		YSGS Y	TACTCTGG CTCTTAT		K	AAA		ADY	GCTGATTAT	
10	ARDR		GCCA GAGA CAGA		VG	GTAGGC		L	CTG		DDY	GATGACTAT	
11	ARGR		GCCA		AT	GCCACT		M	ATG		FDI	TTTGACATT	

[illegible]

[illegible]

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 239

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 239

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

WHAT IS CLAIMED IS:

1. A library of synthetic polynucleotides comprising at least about 10^4 polynucleotides encoding polypeptides comprising CDRH3 sequences with the structure:

[TN1]-[DH]-[N2]-[H3-JH], wherein:

TN1 is a polypeptide corresponding to any of the TN1 polypeptides of Tables 9-10 and 18-26, or a polypeptide produced by translation of any of the TN1 polynucleotides of Tables 25-26;

DH is a polypeptide corresponding to any of the DH polypeptides of Tables 9, 11, 17-25 and 28, or a polypeptide produced by translation of any of the DH-encoding polynucleotides of Tables 16, 25 and 27;

N2 is a polypeptide corresponding to any of the N2 polypeptides of Tables 9, 12, 18-25 and 30, or a polypeptide produced by translation of any of the N2-encoding polynucleotides of Tables 25 and 29; and

H3-JH is a polypeptide corresponding to any of the H3-JH polypeptides of Tables 9, 13, 15, 18-25 and 32, or a polypeptide produced by translation of any of the H3-JH-encoding polynucleotides of Tables 14, 25 and 31.

2. The library of claim 1, wherein at least about 1%, 5%, or 10% of the sequences in the library have the provided structure.
3. The library of claim 1, wherein the polynucleotides encode CDRH3 polypeptides produced by the sets of TN1, DH, N2, and H3-JH polypeptides provided in any one of Tables 23-25.
4. The library of claim 1, wherein the polynucleotides encode CDRH3 polypeptides produced by the set of TN1 polypeptides provided in Table 26, the set of DH polypeptides provided in Table 28, the set of N2 polypeptides provided in Table 30 and the set of H3-JH polypeptides provided in Table 32.

5. A method of using the library of claim 1 to isolate an antibody binding an antigen, comprising contacting the polypeptide expression products of said library with an antigen and isolating polypeptide expression products that bind to the antigen.
6. The library of claim 1 wherein number of N-linked glycosylation sites, deamidation motifs, and/or Cys residues are reduced or eliminated in comparison to libraries produced by amplification of a repertoire from a biological source.
7. The library of claim 1, further comprising a polynucleotide encoding one or more light chain variable domain polypeptides.
8. The library of claim 7, wherein the polypeptides are expressed as full-length IgGs.
9. The polypeptide expression products of the library of claim 8.
10. An antibody isolated from the polypeptide expression products of claim 9.
11. A vector containing the library of claim 1.
12. A host cell containing the vector of claim 11.
13. The host cell of claim 12, wherein the host cell is a yeast cell.
14. The yeast cell of claim 13, wherein the yeast is *Saccharomyces cerevisiae*.
15. A kit containing the library of claim 1.
16. A representation of the library of claim 1 in a computer readable format.
17. A library of synthetic polynucleotides comprising at least about 10^4 polynucleotides encoding polypeptides comprising CDRH3 sequences with the structure:

[TN1]-[DH]-[N2]-[H3-JH], wherein:

TN1 is a polypeptide at least about 80%, 90%, or 95% identical to any of the TN1 polypeptides of Tables 9-10 and 18-26, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the TN1 polynucleotides of Tables 25-26;

DH is a polypeptide at least about 80%, 90%, or 95% identical to any of the DH polypeptides of Tables 9, 11, 17-25 and 28, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the DH-encoding polynucleotides of Tables 16, 25 and 27;

N2 is a polypeptide at least about 80%, 90%, or 95% identical to any of the N2 polypeptides of Tables 9, 12, 18-25 and 30, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the N2-encoding polynucleotides of Tables 25 and 29; and

H3-JH is a polypeptide at least about 80%, 90%, or 95% identical to any of the H3-JH polypeptides of Tables 9, 13, 15, 18-25 and 32, or a polypeptide at least about 80%, 90%, or 95% identical to a polypeptide produced by translation of any of the H3-JH-encoding polynucleotides of Tables 14, 25 and 31.

18. A library of synthetic polynucleotides encoding light chain variable regions, wherein the light chain variable regions are selected from the group consisting of:

- (a) a VK1-05 sequence varied at one or more of positions 4, 49, and 46;
- (b) a VK1-12 sequence varied at one or more of positions 4, 49, 46, and 66;
- (c) a VK1-33 sequence varied at one or more of positions 4, 49, and 66;
- (d) a VK1-39 sequence varied at one or more of positions 4, 49, and 46;
- (e) a VK2-28 sequence varied at one or more of positions 2, 4, 46, and 49;
- (f) a VK3-11 sequence varied at one or more of positions 2, 4, 36, and 49;

- (g) a VK3-15 sequence varied at one or more of positions 2, 4, 48, and 49;
 - (h) a VK3-20 sequence varied at one or more of positions 2, 4, 48, and 49; and
 - (i) a VK4-1 sequence varied at one or more of positions 4, 46, 49, and 66.
19. The library of claim 18, wherein the library comprises polynucleotides encoding light chain variable regions that comprise polypeptide sequences at least about 80%, 90%, or 95% identical to two or more of the light chain polypeptide sequences provided in Table 3.
 20. The library of claim 18, wherein the light chain variable regions comprise the polypeptide sequences provided in Table 3.
 21. A library of synthetic polynucleotides encoding light chain variable regions, wherein the polypeptide sequences of the light chain variable regions are varied at two or three residues between positions 89 to 94, inclusive, of the variable light chain polypeptide sequence.
 22. The library of claim 21, wherein the library comprises polynucleotides encoding light chain variable regions comprising polypeptide sequences at least about 80%, 90%, or 95% identical to polypeptides produced by translation of two or more of the polynucleotide sequences provided in Tables 5-7.
 23. The library of claim 21, wherein the light chain variable regions comprise the polypeptides produced by translation of the polynucleotide sequences provided in Tables 5-7.
 24. A representation of any of the following, in a computer readable format:
 - the TN1 polypeptides of Tables 10, 23-25 and 26;
 - the DH polypeptides of Tables 11, 23-25 and 28;
 - the N2 polypeptides of Tables 12, 23-25 and 30;
 - the H3-JH polypeptides of Tables 13, 15, 17, 23-25 and 32;
 - the TN1 polynucleotides of Tables 25-26;

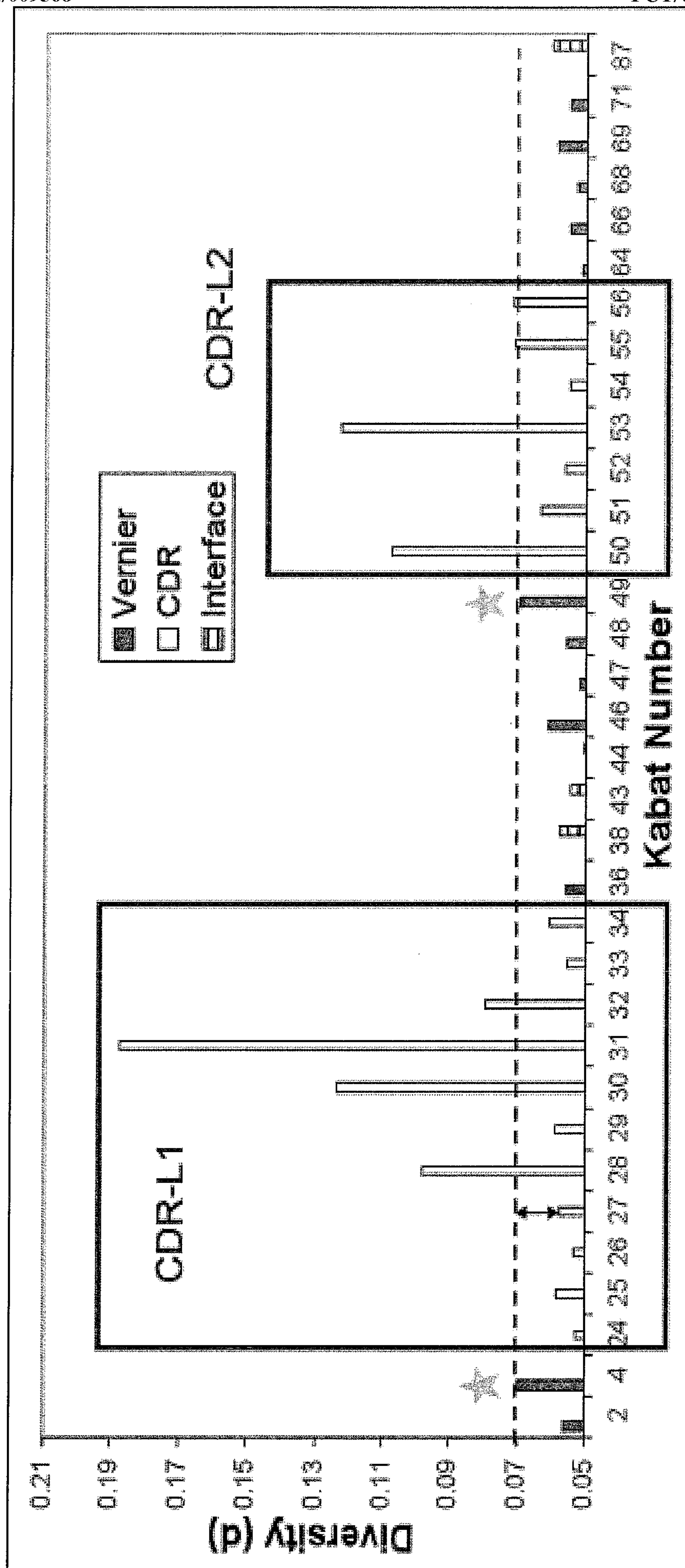
- the DH polynucleotides of Tables 25 and 27;
- the N2 polynucleotides of Tables 25 and 29; and
- the H3-JH polynucleotides of Tables 25 and 31.
25. A representation of the polynucleotide sequences of the Human Preimmune Set (Appendix A), or the polypeptide expression products thereof, in a computer readable format.
26. A method of making a library of synthetic polynucleotides encoding polypeptides comprising CDRH3 sequences, the method comprising:
- (a) providing a theoretical segment pool containing TN1, DH, N2, and H3-JH segments;
 - (b) providing a reference set of CDRH3 sequences;
 - (c) utilizing the theoretical segment pool of (a) to identify the closest match(es) to each CDRH3 sequence in the reference set of (b);
 - (d) selecting segments from the theoretical segment pool for inclusion in a synthetic library; and
 - (e) synthesizing the synthetic CDRH3 library.
27. A library of polynucleotides made according to the method of claim 26.
28. The method of claim 26, wherein the segments selected for inclusion in the synthetic library are selected according to their segment usage weight in the reference set of CDRH3 sequences.
29. The method of claim 26, wherein the segments selected for inclusion in the synthetic library are selected according to one or more physicochemical properties.
30. The method of claim 26, wherein the reference set of CDRH3 sequences is a reference set of preimmune CDRH3 sequences.
31. The method of claim 26, further comprising selecting additional TN1 and N2 segments occurring in the reference set but not in the theoretical segment pool.

32. The method of claim 26, wherein stop codons are reduced or eliminated from the library.
33. The method of claim 26, wherein the unpaired Cys residues, N-linked glycosylation motifs, and deamidation motifs are reduced or eliminated in the translation products of the library.
34. The method of claim 26, wherein the DH segments and H3-JH segments are progressively truncated prior to matching with the CDRH3 sequences in the reference set.
35. The method of claim 26, further comprising introducing one or two degenerate codons in a segment selected from the group consisting of DH and N2, or combinations thereof.
36. The method of claim 26, further comprising introducing one degenerate codon in the H3-JH segment.
37. The polypeptide expression products of the library of claim 26.
38. An antibody isolated from the polypeptide expression products of claim 37.
39. A method of making synthetic polynucleotides encoding a CDRL3 library, comprising:
 - (i) obtaining a reference set of light chain sequences, wherein the reference set contains light chain sequences with VL segments originating from the same IGVL germline gene and/or its allelic variants;
 - (ii) determining which amino acids occur at each of the CDRL3 positions in the reference set that are encoded by the IGVL gene;
 - (iii) synthesizing light chain variable domain encoding sequences wherein two or three positions between positions 89 and 94, inclusive, contain degenerate codons encoding two or more of the five most frequently occurring amino acid residues at the corresponding positions in the reference set; and
 - (iv) synthesizing the polynucleotides encoding the CDRL3 library.

40. A library of polynucleotides made according to the method of claim 39.
41. A library of synthetic polynucleotides encoding polypeptides comprising:
- one or more VH chassis comprising Kabat residues 1 to 94 of a human IGHV sequence;
 - one or more TN1 segments selected from a reference set of human CDRH3 sequences;
 - one or more DH segments selected from a theoretical segment pool that has been matched to a reference set of human CDRH3 sequences;
 - one or more N2 segments selected from a reference set of human CDRH3 sequences; and
 - one or more H3-JH segments selected from a theoretical segment pool that has been matched to a reference set of human CDRH3 sequences.

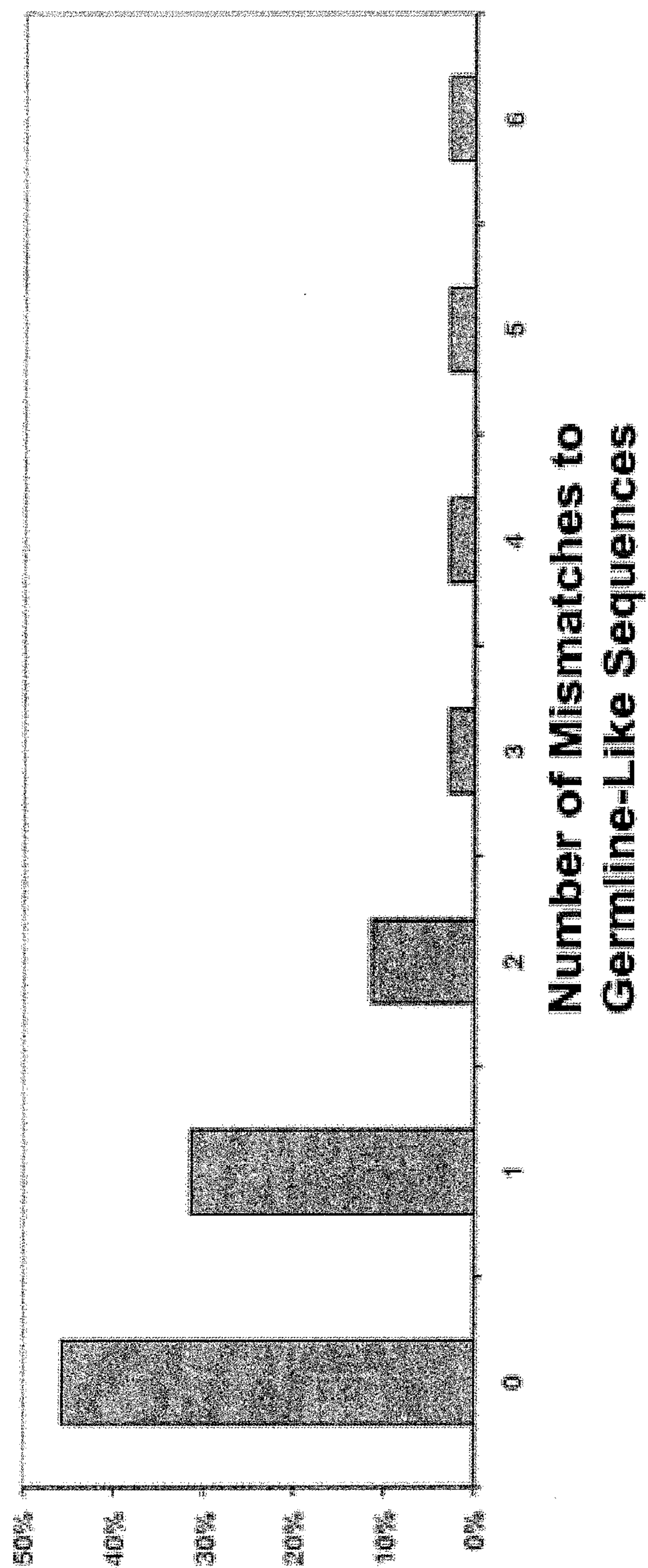
1/15

FIGURE 1



2/15

FIGURE 2



3/15

FIGURE 3

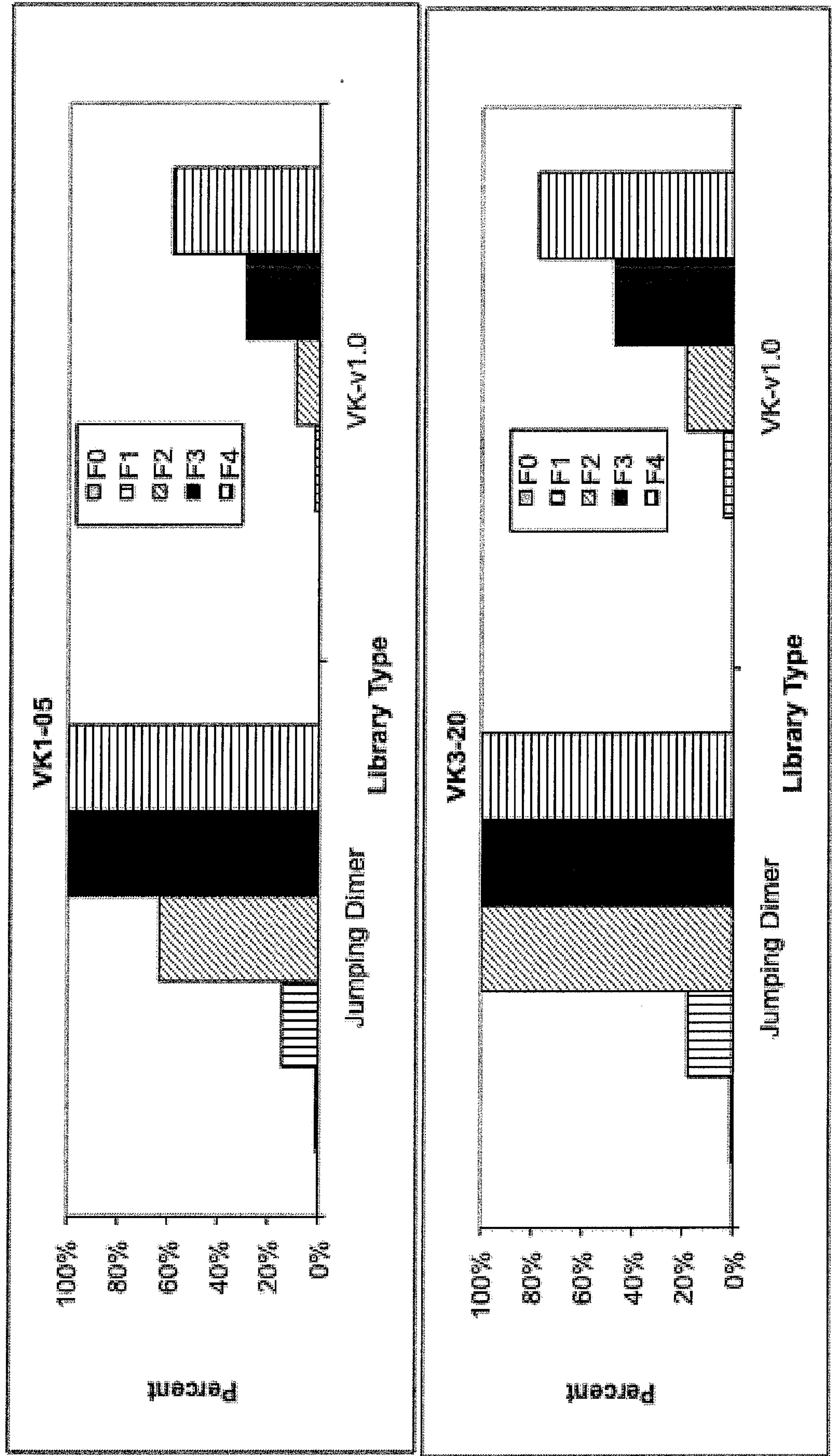


FIGURE 4

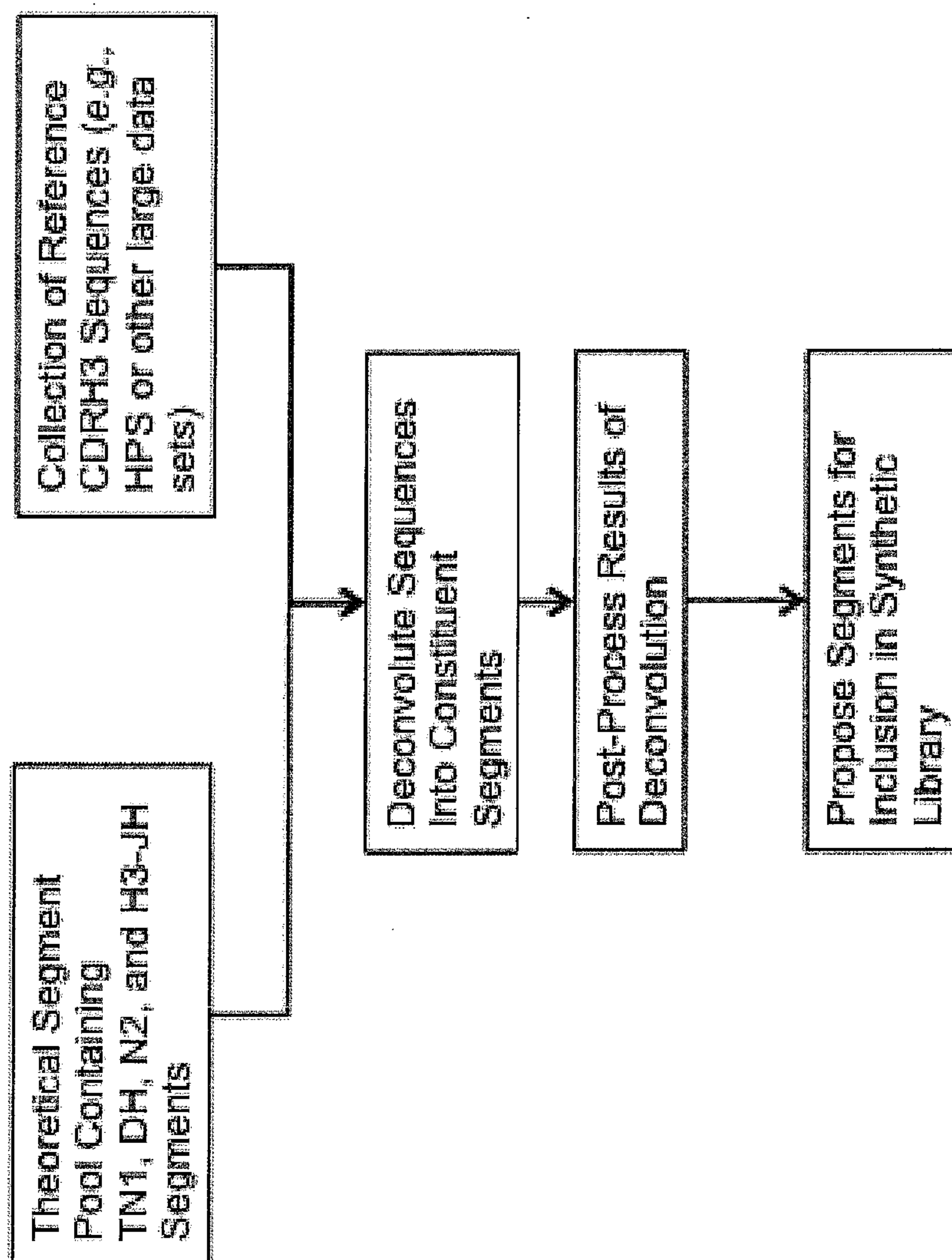
IGHJ3-02	"	ATGCTTTTGATATCT...TCCTTCAG
IGHJ3-02-1	Mutate 5' doublet	NNTGCTTTTGATATCT...TCCTTCAG
IGHJ3-02-2	Delete 5' codon	GCTTTTGATATCT...TCCTTCAG
IGHJ3-02-3	Mutate 5' doublet	NNTTTTGATATCT...TCCTTCAG
IGHJ3-02-4	Delete 5' codon	TTTGATATCT...TCCTTCAG
IGHJ3-02-5	Mutate 5' doublet	NNTGATATCT...TCCTTCAG

IGHJ1-01	"	GCTGAATACTTCCAG...TCCTTCAG
IGHJ1-01-1	Mutate 5' doublet	NNTGAATACTTCCAG...TCCTTCAG
IGHJ1-01-2	Delete 5' codon	GAATACTTCCAG...TCCTTCAG
IGHJ1-01-3	Mutate 5' doublet	NNATACTTCCAG...TCCTTCAG
IGHJ1-01-4	Delete 5' codon	TACTTCCAG...TCCTTCAG
IGHJ1-01-5	Mutate 5' doublet	NNCTTCCAG...TCCTTCAG

IGHJ2-01	"	CTACTGGTACTTCGA...TCCTTCAG
IGHJ2-01-1	Mutate 5' doublet	NNACTGGTACTTCGA...TCCTTCAG
IGHJ2-01-2	Delete 5' codon	CTGGTACTTCGA...TCCTTCAG
IGHJ2-01-3	Mutate 5' doublet	NNGGTACTTCGA...TCCTTCAG
IGHJ2-01-4	Delete 5' codon	GTA CTTCGA...TCCTTCAG
IGHJ2-01-5	Mutate 5' doublet	NNACTTCGA...TCCTTCAG

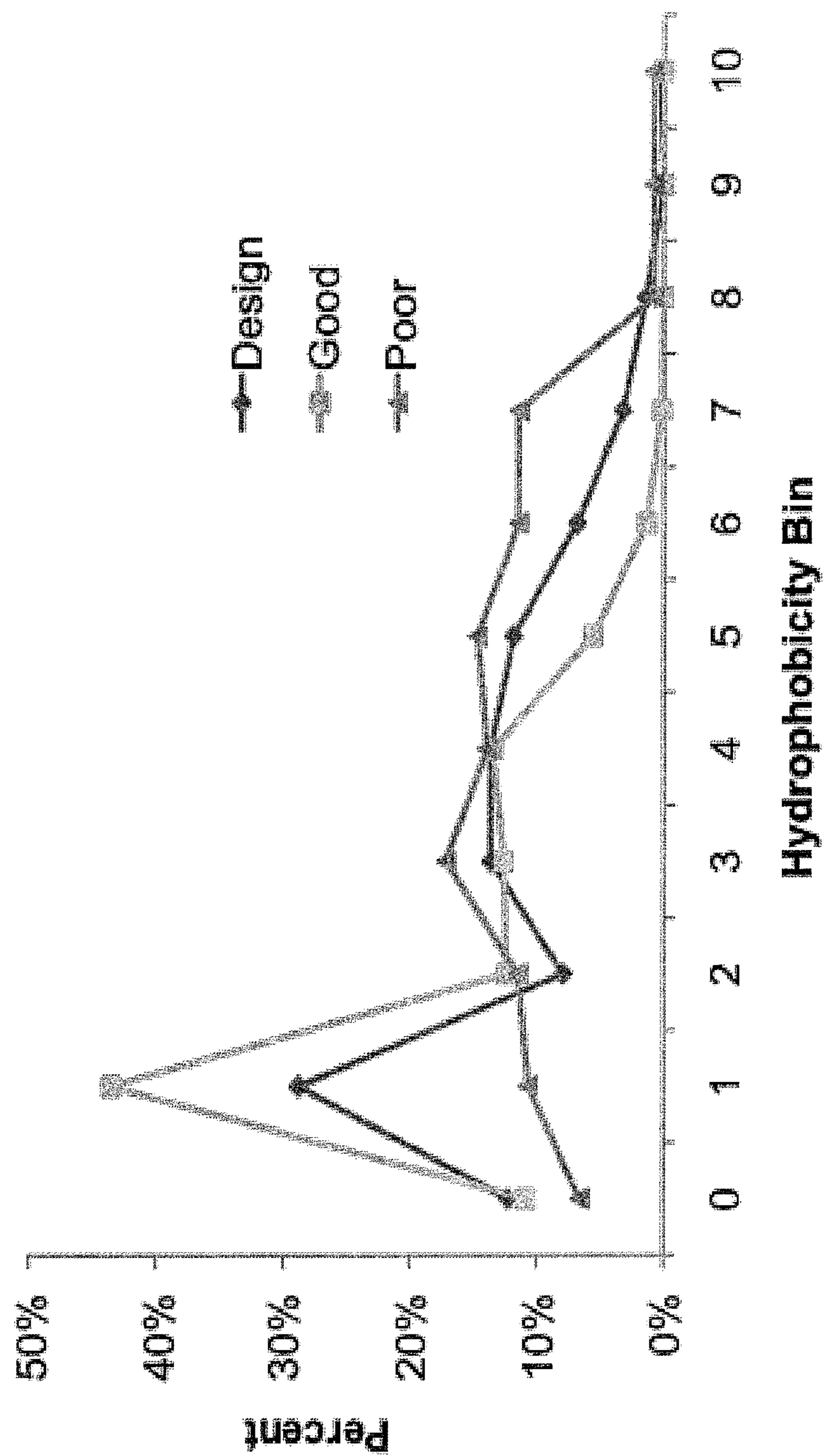
5/15

FIGURE 5



6/15

FIGURE 6



7/15

FIGURE 7

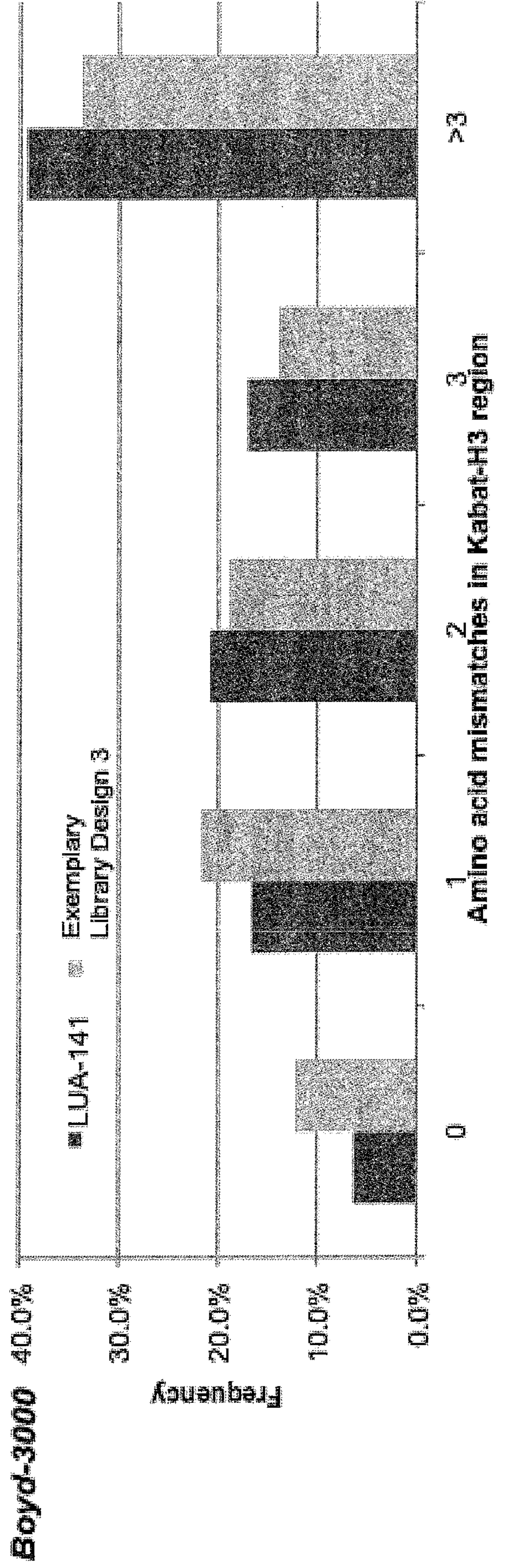
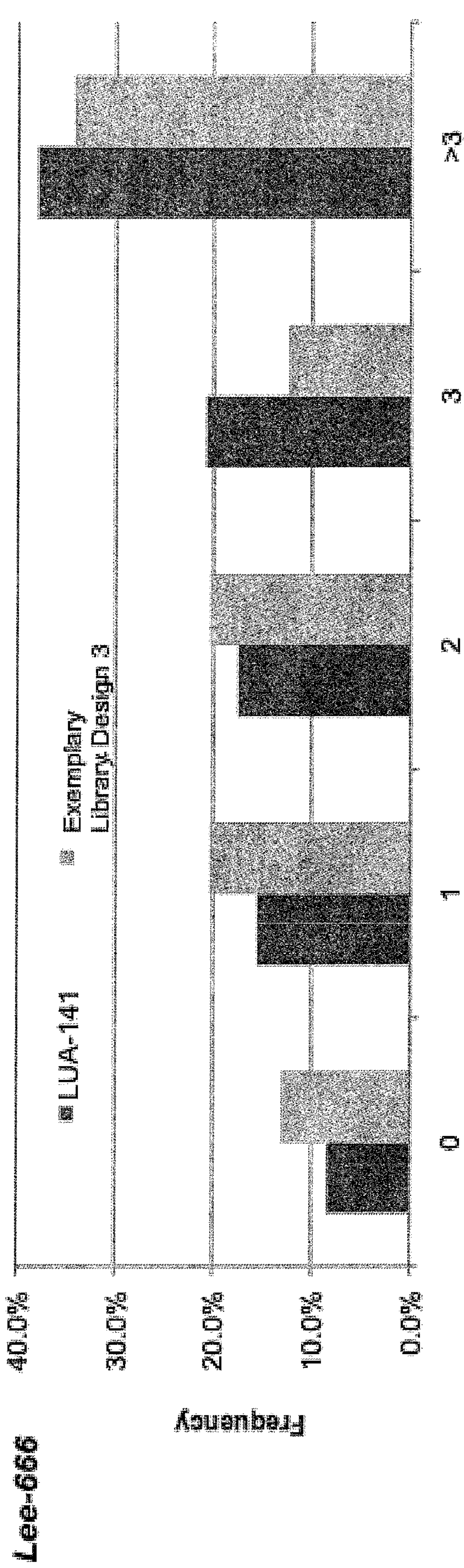


FIGURE 8

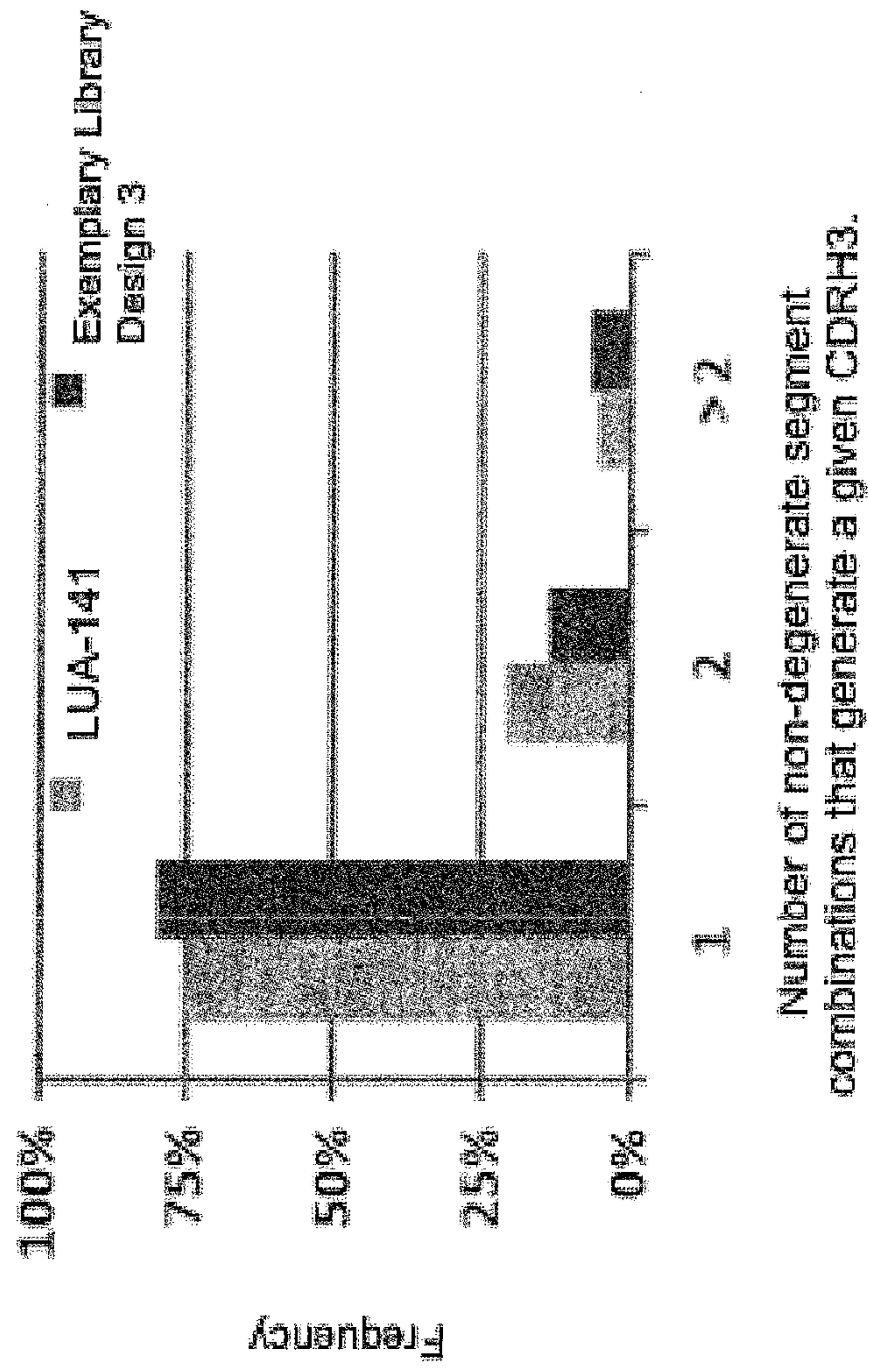
Based on a set of CDRH3 sequences derived from 55 clinically relevant human mAbs

Amino acid mismatches in Kabat-H3 region	LUA-141	Exemplary Library Design 3 (EDLD-3)	Extended Diversity Library Design (EDLD)
0	4	7	25
1	16	27	15
2	19	5	9
3-or-more	16	16	6
Total	55	55	55

9/15

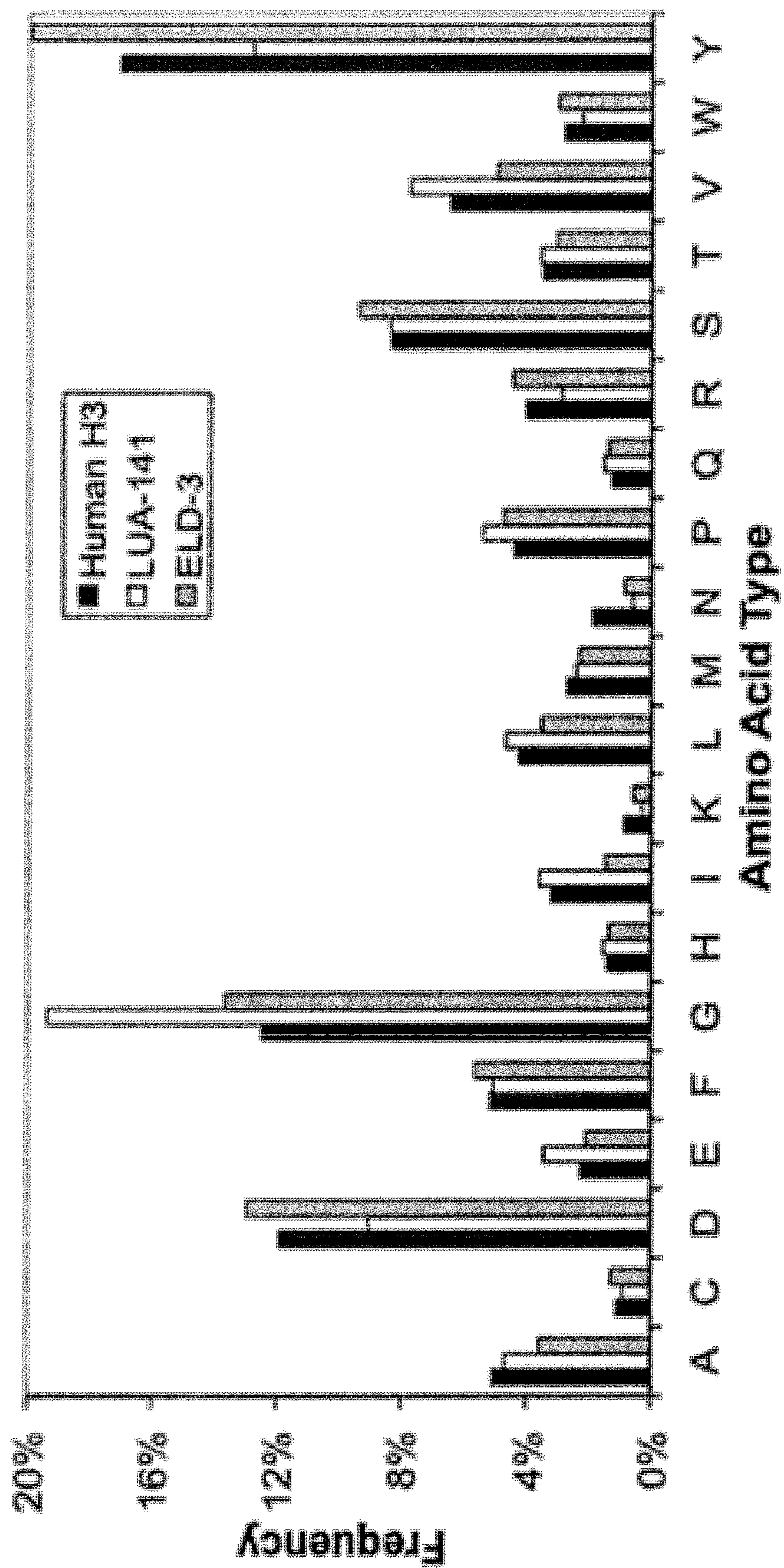
FIGURE 9

Based on "self-matching" using a sample of 2,000 CDRH3 sequences from each library



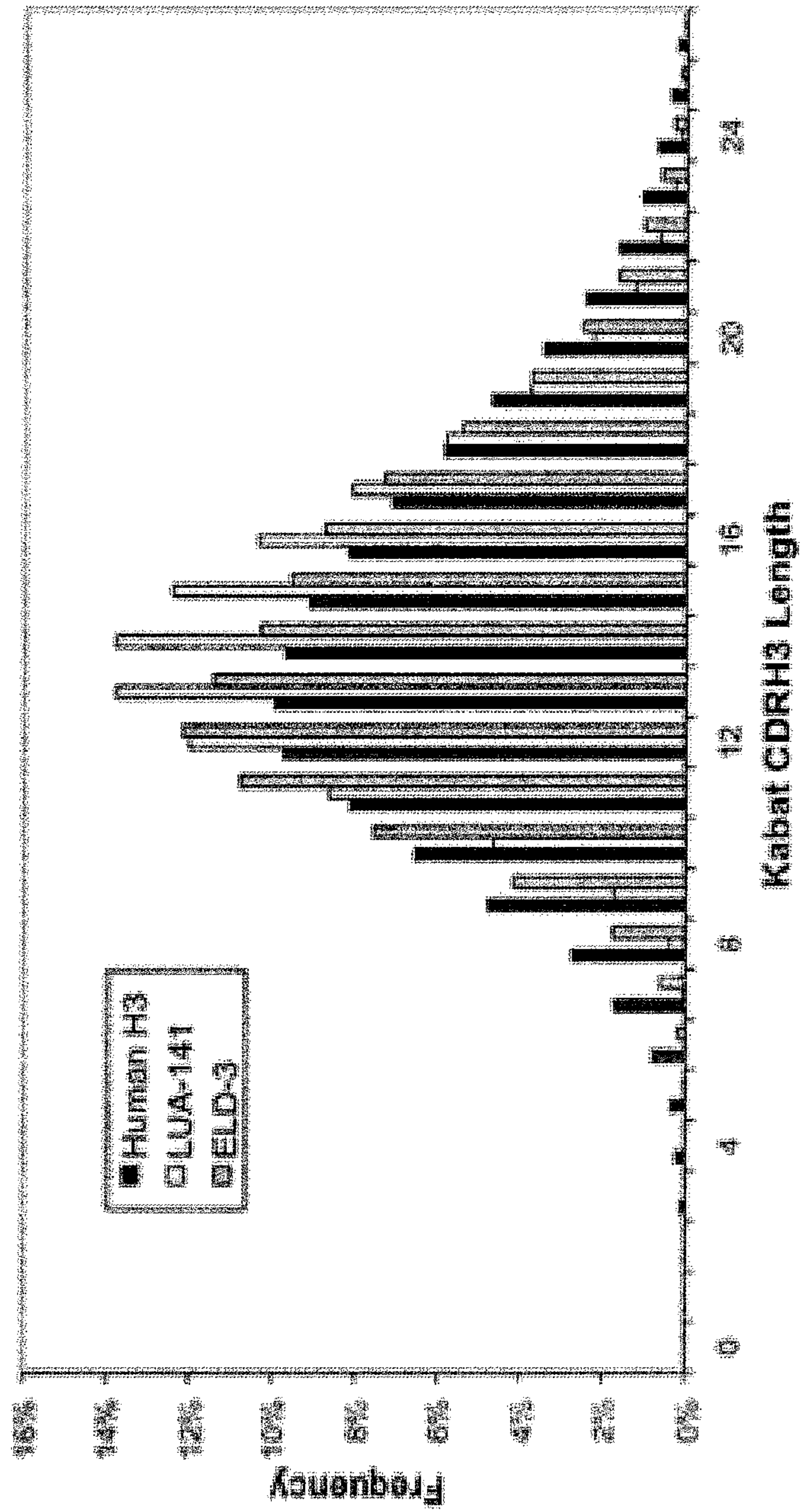
10/15

FIGURE 10



11/15

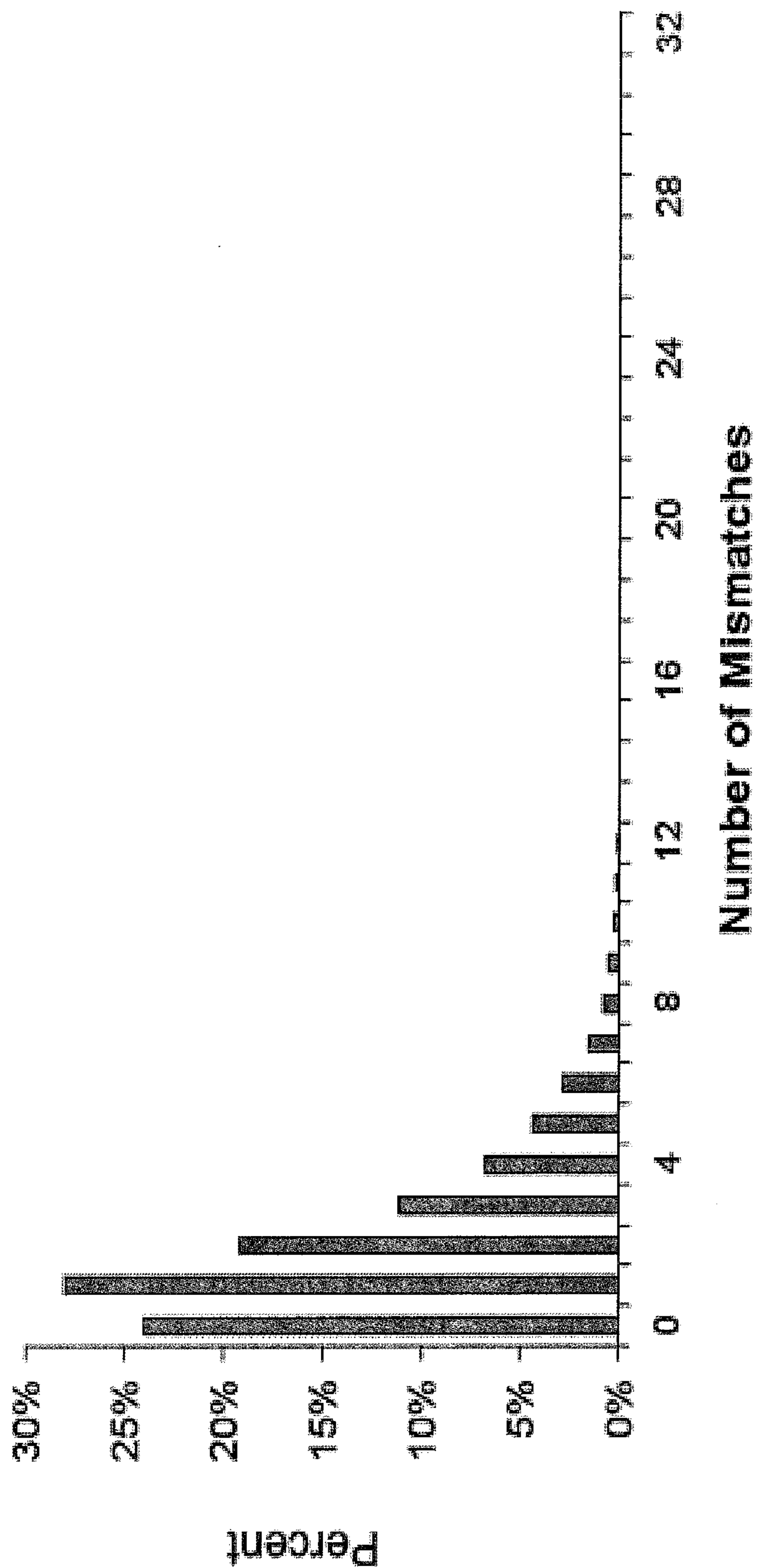
FIGURE 11



Library		LUA-141		Exemplary Library Design 3		Human H3	
Average		14.37		13.94		13.99	
Stdev		2.82		3.35		3.55	
Shortest		5		2		2	
Longest		26		28		35	

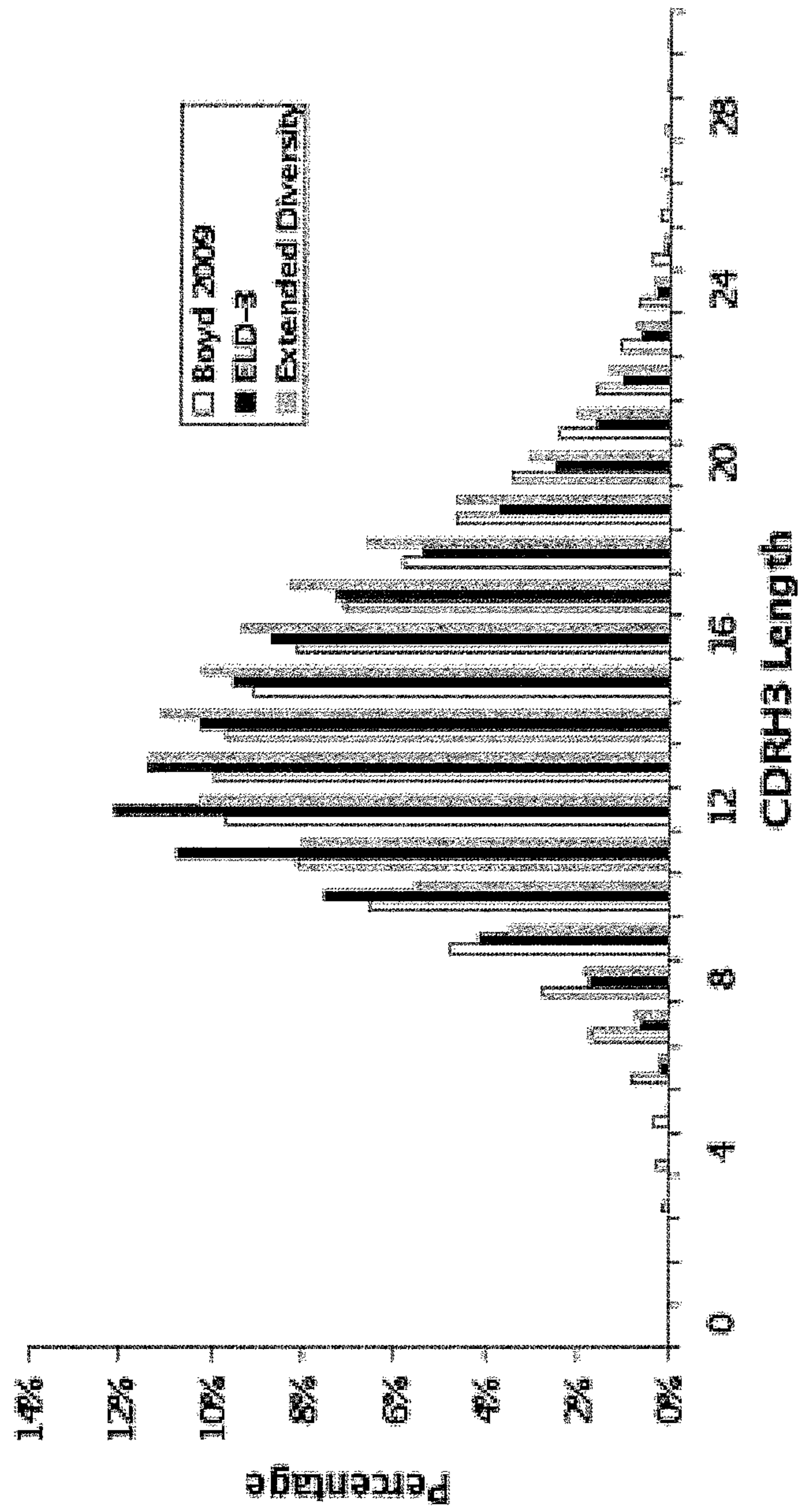
12/15

FIGURE 12



13/15

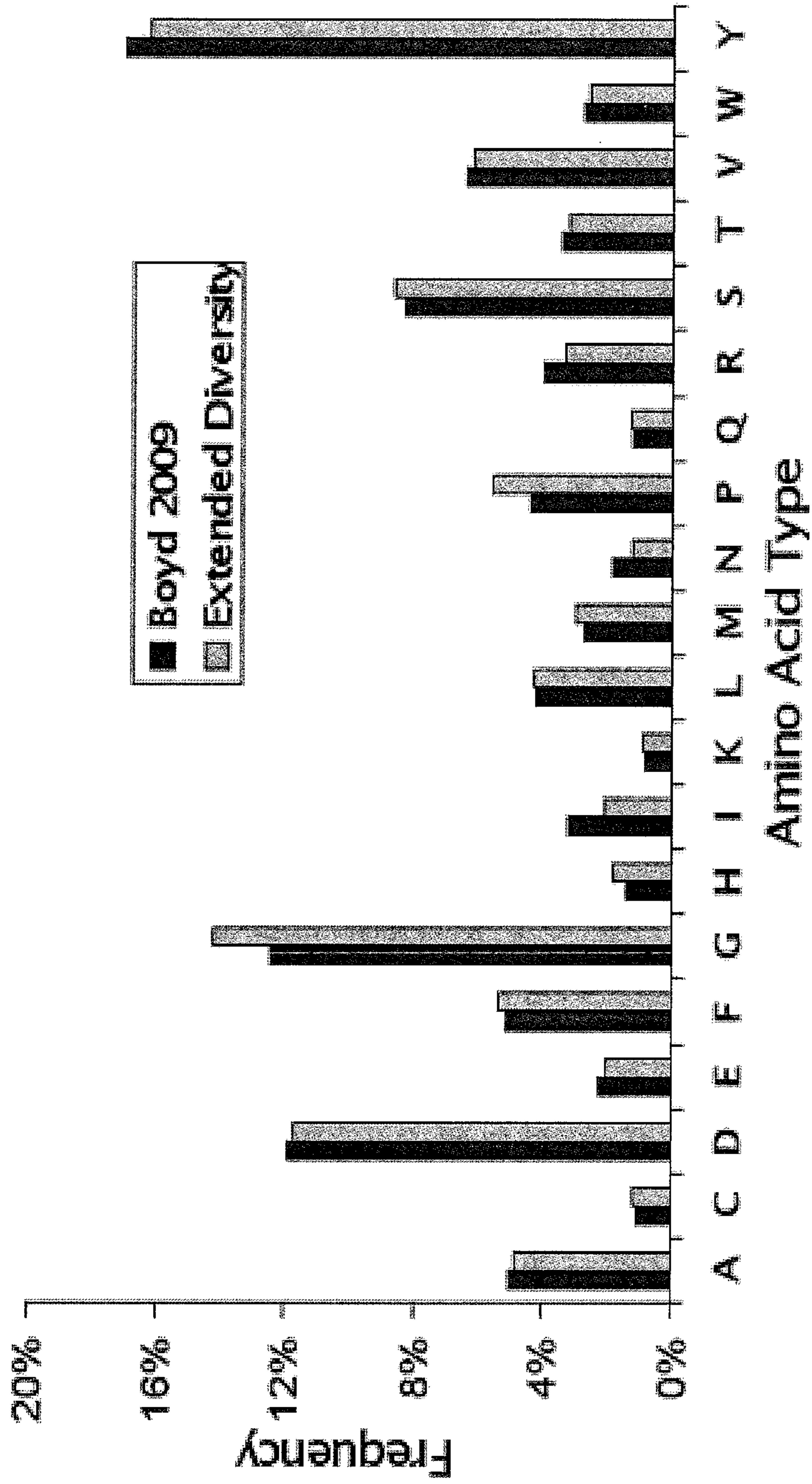
FIGURE 13



Sat:	Boyd	ELD-3	Extended Diversity
r2-to-Boyd	1.00	0.986	0.990
Mean	14.27	13.99	14.47
Standard Deviation	4.02	3.35	3.47
Shortest	1	2	2
Longest	30	28	30

14/15

FIGURE 14



15/15

FIGURE 15

