



(86) Date de dépôt PCT/PCT Filing Date: 2011/07/14
(87) Date publication PCT/PCT Publication Date: 2012/01/19
(45) Date de délivrance/Issue Date: 2021/08/24
(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 30/04* (2006.01),
C40B 40/02 (2006.01), *C40B 40/10* (2006.01),
C40B 50/00 (2006.01), *G16B 35/00* (2019.01)

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

(73) Propriétaire/Owner:
ADIMAB, LLC, US

(74) Agent: TORYS LLP

(54) Titre : LIBRAIRIES RENFERMANT DES GROUPEMENTS DE SEGMENTATION ET LEURS METHODES DE PREPARATION ET UTILISATION

(54) Title: LIBRARIES COMPRISING SEGMENTAL POOLS, AND METHODS FOR THEIR PREPARATION AND USE

(57) Abrégé/Abstract:

Methods are provided for generating polypeptide libraries by 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), as well as libraries that are generated thereby, methods for using such libraries (e.g., an antibody library), kits comprising such libraries, and expression products that are encoded by such libraries.

ABSTRACT

Methods are provided for generating polypeptide libraries by generating theoretical segment pools of TN1, DH, N2, and H3-JH segments to consider for inclusion in a physical manifestation
5 of a library (*e.g.*, polynucleotide or polypeptide), as well as libraries that are generated thereby, methods for using such libraries (*e.g.*, an antibody library), kits comprising such libraries, and expression products that are encoded by such libraries.

DEMANDES OU BREVETS VOLUMINEUX

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

CECI EST LE TOME __1__ DE __2__

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__

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

LIBRARIES COMPRISING SEGMENTAL POOLS, AND METHODS FOR THEIR PREPARATION AND USE

REFERENCE TO RELATED APPLICATIONS

This application is a national phase application under 35 U.S.C. § 371 of PCT International Application No. PCT/US2011/044063 filed July 14, 2011, which claims priority to U.S. Provisional Application No. 61/365,194 filed July 16, 2010.

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,

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 (c) 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 (SEQ ID NOS 8748-8759, respectively, in order of appearance) 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, 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).

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.

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:

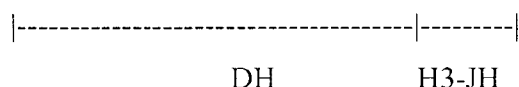
13 Amino Acid CDR-H3 with TN1 and N2

(95)(96)(96A)(97)(97A)(97B)(97C)(97D) (98) (99)(100)(101)(102)



10 Amino Acid CDR-H3 without TN1 and N2

(97)(97A)(97B)(97C)(97D)(97E)(97F)(97G)(101)(102)



“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);

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 IGVL 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 κ) would be combined with the sequences largely, but not exclusively, derived from the J κ 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 κ -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),

Miltenyi (US Patent No. 7,166,423),

Fandl

(U.S. Patent No. 6,919,183, US Pub No. 20060234311),

Clausell-Tormos *et al.* (Chem. Biol., 2008, 15: 427),

Love *et al.* (Nat. Biotechnol., 2006, 24: 703),

and Kelly *et al.* (Chem. Commun., 2007, 14:

1773).

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

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).

"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) 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).

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 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).

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.

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,

and sequences compiled from rearranged VK and V λ sequences (see WO/2009/036379).

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.

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).

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),

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

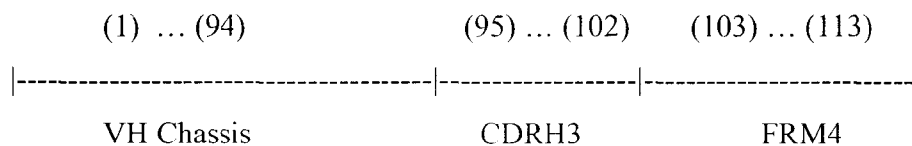
library. Therefore, chassis may be selected to produce libraries that represent the largest and most structurally diverse group of functional human antibodies.

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.

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),

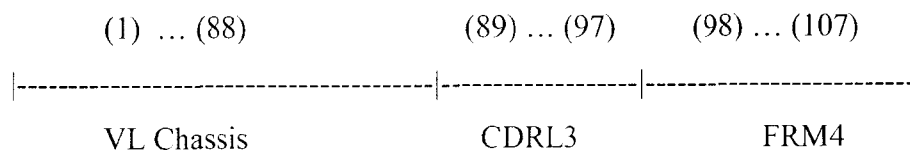
these chassis can also be varied, particularly by altering the amino acid residues in the CDRH1 and CDRH2 regions, further increasing the diversity of the library.

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.

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 IG $\square$ V 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 IG $\square$ V germline sequences: IG $\square$ V3-1, IG $\square$ V3-21, IG $\square$ 44, IG $\square$ V1-40, IG $\square$ V3-19, IG $\square$ V1-51, IG $\square$ V1-44, IG $\square$ V6-57, IG $\square$ 11, IG $\square$ V3-25, IG $\square$ 53, IG $\square$ V3-10, IG $\square$ V4-69, IG $\square$ V1-47, IG $\square$ 41, IG $\square$ V7-43, IG $\square$ V7-46, IG $\square$ V5-45, IG $\square$ V4-60, IG $\square$ V10-54, IG $\square$ V8-61, IG $\square$ V3-9, IG $\square$ V1-36, IG $\square$ 48, IG $\square$ V3-16, IG $\square$ V3-27, IG $\square$ V4-3, IG $\square$ V5-39, IG $\square$ V9-49, and/or IG $\square$ 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 $\square$ 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. Physicochemical 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); BROC820101 Retention coefficient in TFA (Browne et al., 1982); BROC820102 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 3JH_{alpha}-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 pi (Fauchere-Pliska, 1983); FAUJ880101 Graph shape index (Fauchere et al., 1988); FAUJ880102 Smoothed upilon 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 σ
 $\times 1.0E4$ (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 noside (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/McCN/H2O (Wilce et al. 1995); WILM950102 Hydrophobicity coefficient in RP-HPLC, C8 with 0.1%TFA/McCN/H2O (Wilce et al. 1995); WILM950103

Hydrophobicity coefficient in RP-HPLC, C4 with 0.1%TFA/McCN/H₂O (Wilce et al. 1995); WILM950104 Hydrophobicity coefficient in RP-HPLC, C18 with 0.1%TFA/2-PrOH/McCN/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.

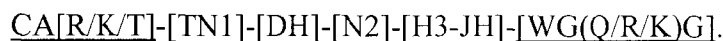
Design of CDRH3 Libraries Using the TN1, DH, N2, and H3-JH Segments

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:

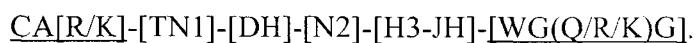
[TN1]-[DH]-[N2]-[H3-JH].

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.

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.

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 IGV λ 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.

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).

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).

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).

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). 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).

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).

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).

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).

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).

In 2000, Uetz *et al.* conducted a comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae* (Uetz *et al.*, *Nature*, 2000, 403: 623).

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), and the genomic protein linkage map of Vaccinia virus was studied using this system (McCraith *et al.*, PNAS, 2000, 97: 4879).

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.

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).

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).

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.

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).

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).

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 US2003/0036093.

Higher eukaryotic cells, such as mammalian cells, for example myeloma cells (*e.g.*, NS/0 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).

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

or by other anchoring methods as described, for example, in Mazor *et al.*, Nature Biotechnology, 2007, 25: 563.

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

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.

Determining the ability of candidate antibodies to bind therapeutic targets can be assayed *in vitro* using, *e.g.*, a BIAcore™ 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

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).

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 κ -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), κ 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,

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.

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.

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.

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.

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

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).

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).

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).

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).

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).

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) 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 V λ 1-5 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,

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).

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 = {
    'AIA':'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')
```

```
oul = 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].progdcl = set()

# Define here minimum length for DNA (4) and for protein (2)
minlen = 4
minrp = 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].progdcl.add(nseq)
    alln += len(seqs[i].progdcl)

# Collect unique DNA sequences across all DH genes of origin
and ignore redundant ones
progdcl = [ table() for i in range(alln) ]
n = 0
for i in range(nseg):
    for kk in seqs[i].progdcl:
        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(nr):
    out.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 (SEQ ID NO: 8760) (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 (SEQ ID NO: 8761) (*i.e.* [TN1]-[DH]-[N2]-[H3-JH], respectively). The DH and H3-JH segments from the reference CDRH3, NWG and YAFDI (SEQ ID NO: 4540) respectively, are identically present in TSP1. However, the TN1 (ERTI) (SEQ ID NO: 8718) 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:

Equation 2

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

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_i , 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 (SEQ ID NO: 3660) 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 (SEQ ID NO: 3661) 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 (SEQ ID NO: 3662) 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-YS**G**HD-LG-Y (SEQ ID NO: 3662)

Nearest Neighbor in TSP1: DR-YS**G****Y**D-LG-Y (SEQ ID NO: 8719)

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 GIAAADSNWLDP (SEQ ID NO: 3663) 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-SNW**L**DP (SEQ ID NO: 3663)

Nearest Neighbor in TSP1: (-) -GIAAA-D-SNW**F**DP (SEQ ID NO: 8720)

Sequence from HPS: G-IAAA-D-SNW**L**DP (SEQ ID NO: 3663)

Nearest Neighbor in TSP1: G-IAAA-D-SNW**F**DP (SEQ ID NO: 8720)

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 (SEQ ID NO: 8760) 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 (SEQ ID NO: 8718) and WGV respectively), and the DH and H3-JH segments from TSP1 (NWG and YAFDI (SEQ ID NO: 4540) 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)

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	QQYNSYST	1	VK2-28	1	MQALQTPT	101
VK1-05	2	QQYNSYFT	2	VK2-28	2	MQALQIFT	102
VK1-05	3	QQYNSYLT	3	VK2-28	3	MQALQILT	103
VK1-05	4	QQYNSYIT	4	VK2-28	4	MQALQITIT	104
VK1-05	5	QQYNSYRT	5	VK2-28	5	MQALQITRI	105
VK1-05	6	QQYNSYWT	6	VK2-28	6	MQALQITWT	106
VK1-05	7	QQYNSYIT	7	VK2-28	7	MQALQIYT	107
VK1-05	8	QQYNSYSPT	8	VK2-28	8	MQALQITPPT	108
VK1-05	9	QQYNSYSFT	9	VK2-28	9	MQALQITPFT	109
VK1-05	10	QQYNSYSLT	10	VK2-28	10	MQALQITPLT	110
VK1-05	11	QQYNSYSIT	11	VK2-28	11	MQALQITPII	111
VK1-05	12	QQYNSYSRT	12	VK2-28	12	MQALQITPRT	112
VK1-05	13	QQYNSYSWT	13	VK2-28	13	MQALQITPWT	113
VK1-05	14	QQYNSYSYT	14	VK2-28	14	MQALQITPYT	114
VK1-05	15	QQYNSYSPFT	15	VK2-28	15	MQALQITPPFT	115
VK1-05	16	QQYNSYSPLT	16	VK2-28	16	MQALQITPPLI	116
VK1-05	17	QQYNSYSPIT	17	VK2-28	17	MQALQITPPII	117
VK1-05	18	QQYNSYSPRT	18	VK2-28	18	MQALQITPPRT	118
VK1-05	19	QQYNSYSPWT	19	VK2-28	19	MQALQITPPWT	119
VK1-05	20	QQYNSYSPYT	20	VK2-28	20	MQALQITPPYT	120
VK1-12	1	QQANSFPT	21	VK3-11	1	QQRSNWPT	121
VK1-12	2	QQANSFFT	22	VK3-11	2	QQRSNWFT	122
VK1-12	3	QQANSFLT	23	VK3-11	3	QQRSNWLT	123
VK1-12	4	QQANSFIT	24	VK3-11	4	QQRSNWIT	124
VK1-12	5	QQANSFRT	25	VK3-11	5	QQRSNWRT	125
VK1-12	6	QQANSFWT	26	VK3-11	6	QQRSNNWT	126
VK1-12	7	QQANSFYT	27	VK3-11	7	QQRSNWYT	127
VK1-12	8	QQANSFPPT	28	VK3-11	8	QQRSNWPPT	128
VK1-12	9	QQANSFPFT	29	VK3-11	9	QQRSNWPFT	129
VK1-12	10	QQANSFPILT	30	VK3-11	10	QQRSNWPLT	130

VK1-12	11	QANSFPIT	31	VK3-11	11	QRSNWPIIT	131
VK1-12	12	QANSFPRT	32	VK3-11	12	QRSNWPIIT	132
VK1-12	13	QANSFPWT	33	VK3-11	13	QRSNWPIIT	133
VK1-12	14	QANSFPYT	34	VK3-11	14	QRSNWPIIT	134
VK1-12	15	QANSFPFPT	35	VK3-11	15	QRSNWPPFT	135
VK1-12	16	QANSFPPLT	36	VK3-11	16	QRSNWPPLT	136
VK1-12	17	QANSFPPIIT	37	VK3-11	17	QRSNWPPIT	137
VK1-12	18	QANSFPPIIT	38	VK3-11	18	QRSNWPPRI	138
VK1-12	19	QANSFPPIIT	39	VK3-11	19	QRSNWPPWI	139
VK1-12	20	QANSFPPIIT	40	VK3-11	20	QRSNWPPYT	140
VK1-33	1	QYDNLPT	41	VK3-15	1	QYNNWPI	141
VK1-33	2	QYDNLPT	42	VK3-15	2	QYNNWPI	142
VK1-33	3	QYDNLPT	43	VK3-15	3	QYNNWPI	143
VK1-33	4	QYDNLPT	44	VK3-15	4	QYNNWPI	144
VK1-33	5	QYDNLPT	45	VK3-15	5	QYNNWPI	145
VK1-33	6	QYDNLPT	46	VK3-15	6	QYNNWPI	146
VK1-33	7	QYDNLPT	47	VK3-15	7	QYNNWPI	147
VK1-33	8	QYDNLPT	48	VK3-15	8	QYNNWPPPT	148
VK1-33	9	QYDNLPT	49	VK3-15	9	QYNNWPPFT	149
VK1-33	10	QYDNLPT	50	VK3-15	10	QYNNWPLT	150
VK1-33	11	QYDNLPT	51	VK3-15	11	QYNNWPIIT	151
VK1-33	12	QYDNLPT	52	VK3-15	12	QYNNWPIIT	152
VK1-33	13	QYDNLPT	53	VK3-15	13	QYNNWPIIT	153
VK1-33	14	QYDNLPT	54	VK3-15	14	QYNNWPIIT	154
VK1-33	15	QYDNLPPFT	55	VK3-15	15	QYNNWPPFT	155
VK1-33	16	QYDNLPPFT	56	VK3-15	16	QYNNWPPLT	156
VK1-33	17	QYDNLPPIT	57	VK3-15	17	QYNNWPPIT	157
VK1-33	18	QYDNLPPRT	58	VK3-15	18	QYNNWPPRT	158
VK1-33	19	QYDNLPPWT	59	VK3-15	19	QYNNWPPWT	159
VK1-33	20	QYDNLPPYT	60	VK3-15	20	QYNNWPPYT	160
VK1-39	1	QSYSTPT	61	VK3-20	1	QYGSST	161
VK1-39	2	QSYSTFT	62	VK3-20	2	QYGSST	162
VK1-39	3	QSYSTLT	63	VK3-20	3	QYGSST	163
VK1-39	4	QSYSTIT	64	VK3-20	4	QYGSST	164

VK1-39	5	QSYSTRT	65	VK3-20	5	QYSSRT	165
VK1-39	6	QSYSTWT	66	VK3-20	6	QYSSWT	166
VK1-39	7	QSYSTYT	67	VK3-20	7	QYSSYT	167
VK1-39	8	QSYSTPPT	68	VK3-20	8	QYSSPPT	168
VK1-39	9	QSYSTPFT	69	VK3-20	9	QYSSPFT	169
VK1-39	10	QSYSTPLT	70	VK3-20	10	QYSSPLT	170
VK1-39	11	QSYSTPIT	71	VK3-20	11	QYSSPIT	171
VK1-39	12	QSYSTPRT	72	VK3-20	12	QYSSPRT	172
VK1-39	13	QSYSTPWT	73	VK3-20	13	QYSSPWT	173
VK1-39	14	QSYSTPYT	74	VK3-20	14	QYSSPYT	174
VK1-39	15	QSYSTPPFT	75	VK3-20	15	QYSSPPFT	175
VK1-39	16	QSYSTPPLT	76	VK3-20	16	QYSSPPLT	176
VK1-39	17	QSYSTPPIT	77	VK3-20	17	QYSSPPIT	177
VK1-39	18	QSYSTPPRT	78	VK3-20	18	QYSSPPRT	178
VK1-39	19	QSYSTPPWT	79	VK3-20	19	QYSSPPWT	179
VK1-39	20	QSYSTPPYT	80	VK3-20	20	QYSSPPYT	180
VK4-01	1	QYYSTPT	81				
VK4-01	2	QYYSTFT	82				
VK4-01	3	QYYSTLT	83				
VK4-01	4	QYYSTIT	84				
VK4-01	5	QYYSTRT	85				
VK4-01	6	QYYSTWT	86				
VK4-01	7	QYYSTYT	87				
VK4-01	8	QYYSTPFT	88				
VK4-01	9	QYYSTPFT	89				
VK4-01	10	QYYSTPLT	90				
VK4-01	11	QYYSTPIT	91				
VK4-01	12	QYYSTPRT	92				
VK4-01	13	QYYSTPWT	93				
VK4-01	14	QYYSTPYT	94				
VK4-01	15	QYYSTPPFT	95				
VK4-01	16	QYYSTPPLT	96				
VK4-01	17	QYYSTPPIT	97				
VK4-01	18	QYYSTPPRT	98				

VK4-01	19	QYYSTPPWT	99						
VK4-01	20	QYYSTPPYT	100						

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 CDR1, CDR2, and frameworks. The Kabat numbers for segment boundaries are indicated. Here, L1 and L2 (in the "Category" column) indicate variability in CDR1 and CDR2, 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.

Name	Chassis	Category	FRM1: 1-23	CDR1: 24-34	FRM2: 35-49	CDR2: 50-56	FRM3: 57-88	SEQ ID Nos:
VK1-39	VK1-39	Germline	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	AASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	181
VK1-39.1	VK1-39	L1	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	AASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	182
VK1-39.2	VK1-39	L1	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	AASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	183
VK1-39.3	VK1-39	L1	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	AASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	184
VK1-39.6	VK1-39	L2	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	GASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	185
VK1-39.7	VK1-39	L2	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	SASSLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	186
VK1-39.8	VK1-39	L2	DIQMTQSPSSLSASVGDRVTITC	RASQSISSYL N	WYQQKPGKAPK LLIY	AASNLQS	GVPSRFGSGG SGTDFLTITIS SLQPEDFATY YC	187

VK1-39.10	VK1-39	F	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	188
VK1-39.11	VK1-39	FL1	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	189
VK1-39.12	VK1-39	FL1	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	190
VK1-39.15	VK1-39	FL1	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	191
VK1-39.17	VK1-39	FL2	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	SASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	192
VK1-39.18	VK1-39	FL2	DIQLTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIY	AASNLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	193
VK1-39.20	VK1-39	F	DIQMTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIS	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	194
VK1-39.23	VK1-39	FL1	DIQMTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIS	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	195
VK1-39.24	VK1-39	FL1	DIQMTQSPSSLSAS VGDRVITIC	RASQSISSYL N	WYQKPGKAPK LLIS	AASSLOS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY	196

VK1-39.25	VK1-39	FL1	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL N	WYQKPGKAPK LLIS	AASSLQS	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	197
VK1-05	VK1-05	Germlin e	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	198
VK1-05.1	VK1-05	L1	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	199
VK1-05.5	VK1-05	L2	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	EASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	200
VK1-05.6	VK1-05	L2	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	201
VK1-05.7	VK1-05	L12	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	202
VK1-05.8	VK1-05	L12	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	203
VK1-05.9	VK1-05	L12	DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC	204

								YC		
VK1-05.10	VK1-05	F		DIQLTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		205
VK1-05.11	VK1-05	FL1		DIQLTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		206
VK1-05.12	VK1-05	FL1		DIQLTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		207
VK1-05.14	VK1-05	FL1		DIQLTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		208
VK1-05.20	VK1-05	F		DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIS	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		209
VK1-05.21	VK1-05	FL1		DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIS	DASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		210
VK1-05.25	VK1-05	FL2		DIQLTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIY	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		211
VK1-05.26	VK1-05	FL2		DIQMTQSPSTLSAS VGDRVITIC	RASQISSWL A	WYQKPGKAPK LLIS	KASSLES	GVPSRFSGSG SGTEFTLTIS SLQPDDEFATY YC		212

VK1-12	VK1-12	Germlin e	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	213
VK1-12.2	VK1-12	L1	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	214
VK1-12.3	VK1-12	L1	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	215
VK1-12.4	VK1-12	L1	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	216
VK1-12.5	VK1-12	L2	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	GASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	217
VK1-12.6	VK1-12	L2	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	SASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	218
VK1-12.7	VK1-12	L2	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASNLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	219
VK1-12.10	VK1-12	F	DIQMTQSPSSVSAS VGDRVITIC	RASQGISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFTLTIS SLQPEDFATY YC	220

VK1-12.11	VK1-12	FL1	DIQMTQSPSSVSAS VGDRVITIC	RASQDISSWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	221
VK1-12.14	VK1-12	FL1	DIQMTQSPSSVSAS VGDRVITIC	RASGGISRWL A	WYQKPGKAPK LLIY	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	222
VK1-12.15	VK1-12	FL2	DIQMTQSPSSVSAS VGDRVITIC	RASGGISSWL A	WYQKPGKAPK LLIY	GASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	223
VK1-12.16	VK1-12	FL2	DIQMTQSPSSVSAS VGDRVITIC	RASGGISSWL A	WYQKPGKAPK LLIY	SASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	224
VK1-12.17	VK1-12	FL2	DIQMTQSPSSVSAS VGDRVITIC	RASGGISSWL A	WYQKPGKAPK LLIY	AASNLOS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	225
VK1-12.20	VK1-12	F	DIQMTQSPSSVSAS VGDRVITIC	RASGGISSWL A	WYQKPGKAPK LLIS	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	226
VK1-12.21	VK1-12	FL1	DIQMTQSPSSVSAS VGDRVITIC	RASQDISSWL A	WYQKPGKAPK LLIS	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	227
VK1-12.23	VK1-12	FL1	DIQMTQSPSSVSAS VGDRVITIC	RASGGIDSWL A	WYQKPGKAPK LLIS	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY YC	228
VK1-12.24	VK1-12	FL1	DIQMTQSPSSVSAS VGDRVITIC	RASGGISRWL A	WYQKPGKAPK LLIS	AASSLQS	GVPSRFSGSG SGTDFLTIS SLQPEDFATY	229

									YC		
VK1-33	VK1-33	Germlin e	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	230			
VK1-33.1	VK1-33	L1	DIQMTQSPSSLSAS VGDRVITIC	QASQDITNYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	231			
VK1-33.2	VK1-33	L1	DIQMTQSPSSLSAS VGDRVITIC	QASQDIANYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	232			
VK1-33.8	VK1-33	L2	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	233			
VK1-33.10	VK1-33	F	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	234			
VK1-33.13	VK1-33	FL1	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNSL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	235			
VK1-33.14	VK1-33	FL1	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNFL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	236			
VK1-33.17	VK1-33	FL2	DIQMTQSPSSLSAS VGDRVITIC	QASQDISNYL N	WYQQKPGKAPK LLIY	DASNLET	GVPSRFSGSG SGTDFTFITIS SLQPEDIAIY YC	237			

VK1-33.20	VK1-33	F		DIQMTQSPSSLSAS VGDRVITIC	QASQDISNYL N	WYQKPGKAPK LLIY	DASNLET	YC		
VK1-33.21	VK1-33	FL1		DIQMTQSPSSLSAS VGDRVITIC	QASQDITNYL N	WYQKPGKAPK LLIY	DASNLET	GVPSRFSGSR SGTDFTFITIS SLQPEDIATY YC	238	
VK1-33.22	VK1-33	FL1		DIQMTQSPSSLSAS VGDRVITIC	QASQDIANYL N	WYQKPGKAPK LLIY	DASNLET	GVPSRFSGSR SGTDFTFITIS SLQPEDIATY YC	239	
VK1-33.23	VK1-33	FL1		DIQMTQSPSSLSAS VGDRVITIC	QASQDISNSL N	WYQKPGKAPK LLIY	DASNLET	GVPSRFSGSR SGTDFTFITIS SLQPEDIATY YC	240	
VK1-33.24	VK1-33	FL1		DIQMTQSPSSLSAS VGDRVITIC	QASQDISNFL N	WYQKPGKAPK LLIY	DASNLET	GVPSRFSGSR SGTDFTFITIS SLQPEDIATY YC	241	
									242	
VK2-28	VK2-28	Germlin e		DIVMTQSPSLPVT PGEPAISIC	RSSQSLHNS GNYLD	WYQKPGQSPQ LLIY	LGSNRAS	GVPDFSGSG SGTDFTLKIS RVEAEDVGVI YC	243	
VK2-28.1	VK2-28	L1		DIVMTQSPSLPVT PGEPAISIC	RSSQSLYSN GNYLD	WYQKPGQSPQ LLIY	LGSNRAS	GVPDFSGSG SGTDFTLKIS RVEAEDVGVI YC	244	
VK2-28.2	VK2-28	L1		DIVMTQSPSLPVT PGEPAISIC	RSSQSLHRN GNYLD	WYQKPGQSPQ LLIY	LGSNRAS	GVPDFSGSG SGTDFTLKIS RVEAEDVGVI	245	

VK2-28.3	VK2-28	L1		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHTN GNYLD	WYLQKPGQSPQ LLIY	LGSNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	246
VK2-28.4	VK2-28	L1		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ LLIY	LGSNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	247
VK2-28.5	VK2-28	L2		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ LLIY	LASNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	248
VK2-28.6	VK2-28	L2		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ LLIY	LGSHRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	249
VK2-28.10	VK2-28	F		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ VLIY	LGSNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	250
VK2-28.11	VK2-28	L1		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLYSN GNYLD	WYLQKPGQSPQ VLIY	LGSNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	251
VK2-28.15	VK2-28	FL2		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ VLIY	LASNRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	252
VK2-28.17	VK2-28	FL2		DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLLHSN GNYLD	WYLQKPGQSPQ VLIY	LGSSRAS	YC GVPDRFSGSG SGTDFTLKIS RVEAEDVGVY YC	253

VK2-28.20	VK2-28	F	DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLHNS GNYLD	WYQKPGQSPQ LLIF	LGSNRAS	GVPDRFSGSG SGTDFTLTKIS RVEAEDVGVI YC	254
VK2-28.24	VK2-28	FL1	DIVMTQSPPLSLPVT PGEPAISIC	RSSQSLHNS GNYLD	WYQKPGQSPQ LLIF	LGSNRAS	GVPDRFSGSG SGTDFTLTKIS RVEAEDVGVI YC	255
VK3-11	VK3-11	Germlin e	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	256
VK3-11.2	VK3-11	L1	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	257
VK3-11.3	VK3-11	L1	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	258
VK3-11.4	VK3-11	L2	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DSSNRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	259
VK3-11.5	VK3-11	L2	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DTSNRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	260
VK3-11.6	VK3-11	L2	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQKPGQAPR LLIY	DASKRAT	GIPARFSGSG SGTDFTLTKIS SLEPEDFAVI YC	261

VK3-11.10	VK3-11	F	EIVMTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQOKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	262
VK3-11.13	VK3-11	FL1	EIVMTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQOKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	263
VK3-11.14	VK3-11	FL2	EIVMTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQOKPGQAPR LLIY	DSSNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	264
VK3-11.15	VK3-11	FL2	EIVMTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQOKPGQAPR LLIY	DTSNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	265
VK3-11.16	VK3-11	FL2	EIVMTQSPATLSLS PGERATLSC	RASQSVSSYL A	WYQOKPGQAPR LLIY	DASKRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	266
VK3-11.20	VK3-11	F	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WFQOKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	267
VK3-11.21	VK3-11	FL1	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WFQOKPGQAPR LLIY	DASNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	268
VK3-11.24	VK3-11	FL2	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WFQOKPGQAPR LLIY	DSSNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY YC	269
VK3-11.25	VK3-11	FL2	EIVLTQSPATLSLS PGERATLSC	RASQSVSSYL A	WFQOKPGQAPR LLIY	DTSNRAT	GIPARFSGSG SGTDFILITIS SLEPEDFAVY	270

									YC		
VK3-15	VK3-15	Germlin e	EIVMTQSPATLSVS PGERATLSC	RASQSVSSNL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	271			
VK3-15.1	VK3-15	L1	EIVMTQSPATLSVS PGERATLSC	RASQSVGSNL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	272			
VK3-15.6	VK3-15	L1	EIVMTQSPATLSVS PGERATLSC	RASQSVSSSL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	273			
VK3-15.7	VK3-15	L2	EIVMTQSPATLSVS PGERATLSC	RASQSVSSNL A	WYQKPGQAPR LLIY	DASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	274			
VK3-15.8	VK3-15	L2	EIVMTQSPATLSVS PGERATLSC	RASQSVSSNL A	WYQKPGQAPR LLIY	SASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	275			
VK3-15.10	VK3-15	F	EIVLTQSPATLSVS PGERATLSC	RASQSVSSNL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	276			
VK3-15.11	VK3-15	FL1	EIVLTQSPATLSVS PGERATLSC	RASQSVGSNL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	277			
VK3-15.14	VK3-15	FL1	EIVLTQSPATLSVS PGERATLSC	RASQSVSTNL A	WYQKPGQAPR LLIY	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY	278			

VK3-15.16	VK3-15	FL1		EIVLTQSPATLSVS PGERATLSC	RASQSVSSDL A	WYQKPGQAPR LLIY	GASTRAT	YC		
VK3-15.20	VK3-15	F		EIVMTQSPATLSVS PGERATLSC	RASQSVSSNL A	WYQKPGQAPR LLIF	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	279	
VK3-15.21	VK3-15	FL1		EIVMTQSPATLSVS PGERATLSC	RASQSVGSNL A	WYQKPGQAPR LLIF	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	280	
VK3-15.25	VK3-15	FL1		EIVMTQSPATLSVS PGERATLSC	RASQSVSSDL A	WYQKPGQAPR LLIF	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	281	
VK3-15.26	VK3-15	FL1		EIVMTQSPATLSVS PGERATLSC	RASQSVSSSL A	WYQKPGQAPR LLIF	GASTRAT	GIPARFSGSG SGTEFTLTIS SLQSEDFAVY YC	282	
									283	
VK3-20	VK3-20	Germlin e		EIVLTQSPGTLSLS PGERATLSC	RASQSVSSSY LA	WYQKPGQAPR LLIY	GASSRAT	GIPDRFSGSG SGTDFTLTIS RLEPEDFAVY YC	284	
VK3-20.1	VK3-20	L1		EIVLTQSPGTLSLS PGERATLSC	RASQSVRSSY LA	WYQKPGQAPR LLIY	GASSRAT	GIPDRFSGSG SGTDFTLTIS RLEPEDFAVY YC	285	
VK3-20.4	VK3-20	L1		EIVLTQSPGTLSLS PGERATLSC	RASQSVSSDY LA	WYQKPGQAPR LLIY	GASSRAT	GIPDRFSGSG SGTDFTLTIS RLEPEDFAVY	286	

VK3-20.7	VK3-20	L2		EIVLTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASNRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	287
VK3-20.8	VK3-20	L2		EIVLTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASRRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	288
VK3-20.10	VK3-20	F		EIVMTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASSRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	289
VK3-20.17	VK3-20	FL2		EIVMTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASNRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	290
VK3-20.20	VK3-20	F		EIVLTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASSRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	291
VK3-20.22	VK3-20	FL1		EIVLTQSPGILSLSPGERATLSC	RASQSVSSSYLA	WYQKPGQAPRLLIY	GASSRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	292
VK4-01	VK4-01	Germline		DIVMTQSPDLSAVSLGERATINC	KSSQSVLYSSNNKNYLA	WYQKPGQAPRLLIY	WASTRES	GVPRFSGSGSGTDFTLTISRLEPEDFAVYYC	293
VK4-01.1	VK4-01	L1		DIVMTQSPDLSAVSLGERATINC	KSSQSVLYSSNNKNYLA	WYQKPGQAPRLLIY	WASTRES	GVPRFSGSGSGTDFTLTISRLEPEDFAVYYC	294

VK4-01.2	VK4-01	L1		DIVMTQSPDSLAVS LGERATINC	KSSQSILYSS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	295
VK4-01.3	VK4-01	L1		DIVMTQSPDSLAVS LGERATINC	KSSQSVLHSS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	296
VK4-01.4	VK4-01	L1		DIVMTQSPDSLAVS LGERATINC	KSSQSVLFSS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	297
VK4-01.5	VK4-01	L1		DIVMTQSPDSLAVS LGERATINC	KSSQSVLYTS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	298
VK4-01.7	VK4-01	L2		DIVMTQSPDSLAVS LGERATINC	KSSQSVLYSS NNKNYLA	WYQOKPGQPPK LLIY	WASSRES	YC	299
VK4-01.10	VK4-01	F		DIVLTQSPDSLAVS LGERATINC	KSSQSVLYSS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	300
VK4-01.13	VK4-01	FL1		DIVLTQSPDSLAVS LGERATINC	KSSQSVLHSS NNKNYLA	WYQOKPGQPPK LLIY	WASTRES	YC	301
VK4-01.17	VK4-01	FL2		DIVLTQSPDSLAVS LGERATINC	KSSQSVLYSS NNKNYLA	WYQOKPGQPPK LLIY	WASSRES	YC	302

VK4-01.20	VK4-01	F	DIVMTQSPDSLAVS LGERATINC	KSSQSVLYSS NNKNYLA	WYQQKPGQPPK LLIS	WASTRES	GVPDFRSGSG SGTDFTLTIS SLQAEDVAVY YC	303
VK4-01.23	VK4-01	FL1	DIVMTQSPDSLAVS LGERATINC	KSSQSVLHSS NNKNYLA	WYQQKPGQPPK LLIS	WASTRES	GVPDFRSGSG SGTDFTLTIS SLQAEDVAVY YC	304

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 EGG, to yield: ...YYC-[89-97]-EGG... . The sequences in this table encompassed by positions "[89-97]" are disclosed as SEQ ID NOS 8721-8745, respectively, in order of appearance.

Name	Oligo	SEQ ID NO	89	90	91	92	93	94	95	96	97
Jumping Dimer											
VK1-39_1	SWMSWMAGC TACAGTACT CCTTWCAC	305	DEHLQV	DEHLQV	S	Y	S	T	P	FY	T
VK1-39_2	SWMCAAVNA TACAGTACT CCTTWCAC	306	DEHLQV	Q	AEGIKLPQR TV	Y	S	T	P	FY	T
VK1-39_3	SWMCAAAAGC BHCAGTACT CCTTWCAC	307	DEHLQV	Q	S	ADFHLPS VY	S	T	P	FY	T
VK1-39_4	SWMCAAAAGC TACNHCAC CCTTWCAC	308	DEHLQV	Q	S	Y	ADFHLNP STVY	T	P	FY	T
VK1-39_5	SWMCAAAAGC TACAGTBHC CCTTWCAC	309	DEHLQV	Q	S	Y	S	ADFHL PSVY	P	FY	T
VK1-39_6	CAGSWMVNA TACAGTACT CCTTWCAC	310	Q	DEHLQV	AEGIKLPQR TV	Y	S	T	P	FY	T
VK1-39_7	CAGSWMAGC BHCAGTACT CCTTWCAC	311	Q	DEHLQV	S	ADFHLPS VY	S	T	P	FY	T

VK1-39_8	CAGSWMAGC TACNHCAC CCTTWCAC	312	Q		DEHLQV	S		Y	ADFHLNP STVY	T	P	FY	T
VK1-39_9	CAGSWMAGC TACAGTBHC CCTTWCAC	313	Q		DEHLQV	S		Y	S	ADFHL PSVY	P	FY	T
VK1-39_10	CAGCAAVNA BHCAGTACT CCTTWCAC	314	Q		Q	AEGIKLPQR TV		ADFHLPS VY	S	T	P	FY	T
VK1-39_11	CAGCAAVNA TACNHCAC CCTTWCAC	315	Q		Q	AEGIKLPQR TV		Y	ADFHLNP STVY	T	P	FY	T
VK1-39_12	CAGCAAVNA TACAGTBHC CCTTWCAC	316	Q		Q	AEGIKLPQR TV		Y	S	ADFHL PSVY	P	FY	T
VK1-39_13	CAGCAAAAGC BHCNHCAC CCTTWCAC	317	Q		Q	S		ADFHLPS VY	ADFHLNP STVY	T	P	FY	T
VK1-39_14	CAGCAAAAGC BHCAGTBHC CCTTWCAC	318	Q		Q	S		ADFHLPS VY	S	ADFHL PSVY	P	FY	T
VK1-39_15	CAGCAAAAGC TACNHCAC CCTTWCAC	319	Q		Q	S		Y	ADFHLNP STVY	ADFHL PSVY	P	FY	T
Jumping Trimer													
VK1-39_10_0_9	CAGCAAVNA BHCAGTACT CCTTWCAC	314	Q		Q	AEGIKLPQR TV		ADFHLPS VY	S	T	P	FY	T

VK1- 39_11_0_9	CAGCAAVNA TACNHCAC CCTTWCAC	315	Q	Q	Q	AEGIKLPQR TV	Y	ADFHLNP STVY	T	P	FY	T
VK1- 39_12_0_9	CAGCAAVNA TACAGTBHC CCTTWCAC	316	Q	Q	Q	AEGIKLPQR TV	Y	S	ADFHL PSVY	P	FY	T
VK1- 39_13_0_9	CAGCAAAAGC BHCNHCAC CCTTWCAC	317	Q	Q	Q	S	ADFHLPS VY	ADFHLNP STVY	T	P	FY	T
VK1- 39_14_0_9	CAGCAAAAGC BHCAGTBHC CCTTWCAC	318	Q	Q	Q	S	ADFHLPS VY	S	ADFHL PSVY	P	FY	T
VK1- 39_15_0_9	CAGCAAAAGC TACNHCAC CCTTWCAC	319	Q	Q	Q	S	Y	ADFHLNP STVY	ADFHL PSVY	P	FY	T
VK1- 39_t1_0_9	CAGCAAVNA BHCNHCAC CCTTWCAC	320	Q	Q	Q	AEGIKLPQR TV	ADFHLPS VY	ADFHLNP STVY	T	P	FY	T
VK1- 39_t2_0_9	CAGCAAVNA BHCAGTBHC CCTTWCAC		Q	Q	Q	AEGIKLPQR TV	ADFHLPS VY	S	ADFHL PSVY	P	FY	T
VK1- 39_t3_0_9	CAGCAAVNA TACNHCAC CCTTWCAC		Q	Q	Q	AEGIKLPQR TV	Y	ADFHLNP STVY	ADFHL PSVY	P	FY	T
VK1- 39_t4_0_9	CAGCAAAAGC BHCNHCAC CCTTWCAC		Q	Q	Q	S	ADFHLPS VY	ADFHLNP STVY	ADFHL PSVY	P	FY	T

Table 5. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDRL3 length 8.

Name	Sequence of Synthesized Oligonucleotide	Portion of Oligonucleotide Corresponding to CDRL3 Proper	SEQ ID NO (Oligo)	SEQ ID NO (CDRL3 Portion)
Jumping Dimer				
VK1-05_1_0_8	CCGATGATTTTGCAACTTATTACTGCSWMSWMTACA ATAGTTACTWCACITTTGGCGGAGGGACCAAG	SWMSWMTACAAATAGTTACTWCAC	324	948
VK1-05_10_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGBHC RMAGTTACTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTTACTWCAC	325	949
VK1-05_11_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGBHC ATMBCTACTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCATMBCTACTWCAC	326	950
VK1-05_12_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGBHC ATAGTYWCTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCATAGTYWCTWCAC	327	951
VK1-05_13_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMMBCTACTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCTACTWCAC	328	952
VK1-05_14_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMAGTYWCTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACVRMAGTYWCTWCAC	329	953
VK1-05_15_0_8	CCGATGATTTTGCAACTTATTACTGCCAGCAGTACA ATMBCYWCTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACAAATMBCYWCTWCAC	330	954
VK1-05_2_0_8	CCGATGATTTTGCAACTTATTACTGCSWMCAGBHC ATAGTTACTWCACITTTGGCGGAGGGACCAAG	SWMCAGBHCATAGTTACTWCAC	331	955
VK1-05_3_0_8	CCGATGATTTTGCAACTTATTACTGCSWMCAGTACV RMAGTTACTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACVRMAGTTACTWCAC	332	956
VK1-05_4_0_8	CCGATGATTTTGCAACTTATTACTGCSWMCAGTACA ATMBCTACTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACAAATMBCTACTWCAC	333	957
VK1-05_5_0_8	CCGATGATTTTGCAACTTATTACTGCSWMCAGTACA ATAGTYWCTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACAAATAGTYWCTWCAC	334	958
VK1-05_6_0_8	CCGATGATTTTGCAACTTATTACTGCCAGSWMBHCA ATAGTTACTWCACITTTGGCGGAGGGACCAAG	CAGSWMBHCAATAGTTACTWCAC	335	959
VK1-05_7_0_8	CCGATGATTTTGCAACTTATTACTGCCAGSWMTACV RMAGTTACTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACVRMAGTTACTWCAC	336	960

VK1-05_8_0_8	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACA	CAGSWMTACAATMBCTACTWCCTACT	337	961
VK1-05_9_0_8	ATMBCTACTWCCTTTTGGCGGAGGGACCAAG	CAGSWMTACAATAGTYWCTWCCTACT	338	962
VK1-05_1_1_8	CCTGATGATTTTGGCAACTTATTACTGCSWMSWMTACA	SWMSWMTACAATAGTTACMTCTACT	339	963
VK1-05_10_1_8	ATAGTTACMTCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTTACMTCTACT	340	964
VK1-05_11_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHC	CAGCAGBHC	341	965
VK1-05_12_1_8	ATMBCTACTWCCTTTTGGCGGAGGGACCAAG	CAGCAGBHC	342	966
VK1-05_13_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHC	CAGCAGTACVRMMBCTACTWCCTACT	343	967
VK1-05_14_1_8	ATAGTYWCTCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMAGTYWCTWCCTACT	344	968
VK1-05_15_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACV	CAGCAGTACVRMAGTYWCTWCCTACT	345	969
VK1-05_2_1_8	ATMBCTACTWCCTTTTGGCGGAGGGACCAAG	SWMCAGBHC	346	970
VK1-05_3_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACA	SWMCAGTACVRMAGTTACMTCTACT	347	971
VK1-05_4_1_8	ATMBCTACTWCCTTTTGGCGGAGGGACCAAG	SWMCAGTACA	348	972
VK1-05_5_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACA	SWMCAGTACA	349	973
VK1-05_6_1_8	ATAGTYWCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMBHC	350	974
VK1-05_7_1_8	ATAGTTACMTCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACVRMAGTTACMTCTACT	351	975
VK1-05_8_1_8	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACA	CAGSWMTACAATMBCTACTWCCTACT	352	976
VK1-05_9_1_8	ATAGTYWCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMTACAATAGTYWCTWCCTACT	353	977

VK1- 12_13_0_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCNHCITTCWCACCTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCITTCWCACCT	388	1012
VK1- 12_14_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCAGTYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCTWCACCT	389	1013
VK1- 12_15_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATNHCTTCWCACCTTTTGGCGGAGGGACCAAG	CAGCAGGCANATNHCTTCWCACCT	390	1014
VK1-12_2_0_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTTCWCACCTTTTGGCGGAGGGACCAAG	SWMCAGRANAAATAGTTTCWCACCT	391	1015
VK1-12_3_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCAGTTCWCACCTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWCACCT	392	1016
VK1-12_4_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATNHCTTCWCACCTTTTGGCGGAGGGACCAAG	SWMCAGGCANATNHCTTCWCACCT	393	1017
VK1-12_5_0_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTYWCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGGCANAAATAGTYWCTWCACCT	394	1018
VK1-12_6_0_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTTCWCACCTTTTGGCGGAGGGACCAAG	CAGSWMRNAAATAGTTTCWCACCT	395	1019
VK1-12_7_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCAGTTCWCACCTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCWCACCT	396	1020
VK1-12_8_0_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATNHCTTCWCACCTTTTGGCGGAGGGACCAAG	CAGSWMGCANATNHCTTCWCACCT	397	1021
VK1-12_9_0_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMGCANAAATAGTYWCTWCACCT	398	1022
VK1-12_1_1_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTTCWCACCTTTTGGCGGAGGGACCAAG	SWMSWMGCANAAATAGTTTCWCACCT	399	1023
VK1- 12_10_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCAGTTCWCACCTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTTTCWCACCT	400	1024
VK1- 12_11_1_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATNHCTTCWCACCTTTTGGCGGAGGGACCAAG	CAGCAGRANAAATNHCTTCWCACCT	401	1025
VK1- 12_12_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN ATAGTYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGRANAAATAGTYWCTWCACCT	402	1026
VK1- 12_13_1_8	CCITGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCNHCITTCWCACCTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCITTCWCACCT	403	1027
VK1- 12_14_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGAGGCAN HCAGTYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCTWCACCT	404	1028

VK1- 12_15_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCYWCMTCACT	405	1029
VK1-12_2_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTTTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGRNAATAAGTTTCMTCACT	406	1030
VK1-12_3_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA HCAGTTTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCMTCACT	407	1031
VK1-12_4_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATNHCCTTCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATNHCCTTCMTCACT	408	1032
VK1-12_5_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTYWCMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYWCMTCACT	409	1033
VK1-12_6_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTTTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMRNAATAAGTTTCMTCACT	410	1034
VK1-12_7_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA HCAGTTTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCMTCACT	411	1035
VK1-12_8_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATNHCCTTCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCMTCACT	412	1036
VK1-12_9_1_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTYWCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCMTCACT	413	1037
VK1-12_1_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMSWMGCAAAATAGTTTCWGGACT	414	1038
VK1- 12_10_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRANANHCAGTTTCWGGACT	415	1039
VK1- 12_11_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATNHCCTTCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNAATAATNHCCTTCWGGACT	416	1040
VK1- 12_12_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGRNAATAATAGTYWCWGGACT	417	1041
VK1- 12_13_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA HCNHCCTTCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCCTTCWGGACT	418	1042
VK1- 12_14_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA HCAGTYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCWGGACT	419	1043
VK1- 12_15_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATNHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCYWCWGGACT	420	1044
VK1-12_2_2_8	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAA ATAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGRNAATAAGTTTCWGGACT	421	1045

VK1-12_3_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	422	1046
VK1-12_4_2_8	HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	423	1047
VK1-12_5_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	424	1048
VK1-12_6_2_8	ATAGTYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	425	1049
VK1-12_7_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	426	1050
VK1-12_8_2_8	HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	427	1051
VK1-12_9_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	428	1052
VK1-12_10_2_8	ATAGTYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	429	1053
VK1-12_11_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	430	1054
VK1-12_12_2_8	HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	431	1055
VK1-12_13_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	432	1056
VK1-12_14_2_8	ATAGTYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	433	1057
VK1-12_15_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	434	1058
VK1-12_16_2_8	HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	435	1059
VK1-12_17_2_8	ATAGTYWCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	436	1060
VK1-12_18_2_8	CCTGAAGATTTTGGCACTTATTACTGTGTSWMCAGGCAN	SWMCAGGCANHCAGTTTCWGGACT	437	1061
VK1-12_19_2_8	HCAGTTTCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCWGGACT	438	1062

VK1-12_5_3_8	CCTGAAGATTTTGCAACTTATTACIGTSMWCAGGCAA ATAGTYWCCCTACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATAGIYWCCCTACT	439	1063
VK1-12_6_3_8	CCIGAAAGATTTTGCAACTTATTACTGTGAGSWMRNAA ATAGTTTCCCTACIIITGGCGGAGGGACCAAG	CAGSWMRNAAAIAGIITCCCTACT	440	1064
VK1-12_7_3_8	CCIGAAAGATTTTGCAACTTATTACTGTGAGSWMGCAN HCAGTTTCCCTACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTACT	441	1065
VK1-12_8_3_8	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMGCAA ATNHCTTCCCTACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAATNHCTTCCCTACT	442	1066
VK1-12_9_3_8	CCIGAAAGATTTTGCAACTTATTACTGTGAGSWMGCAA ATAGTYWCCCTACTTTTGGCGGAGGGACCAAG	CAGSWMGCAAAIAGIYWCCCTACT	443	1067
VK1-33_1_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMTACG ATAATCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAATCTCTWCACCT	444	1068
VK1-33_10_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGBHCN HCAATCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAICTCTWCACCT	445	1069
VK1-33_11_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGBHCN ATNHCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCTCTWCACCT	446	1070
VK1-33_12_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGBHCN ATAATYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCCTCTWCACCT	447	1071
VK1-33_13_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN HCNHCCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCTWCACCT	448	1072
VK1-33_14_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN HCAATYWCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCTWCACCT	449	1073
VK1-33_15_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN ATNHCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCTWCACCT	450	1074
VK1-33_2_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGBHCN ATAATCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCNHCCTCTCTWCACCT	451	1075
VK1-33_3_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN HCAATCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCNHCCTCTCTWCACCT	452	1076
VK1-33_4_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN ATNHCTCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCNHCCTCTCTWCACCT	453	1077
VK1-33_5_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGTACN ATAATYWCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCNHCCTCTCTWCACCT	454	1078
VK1-33_6_0_8	CCTGAAGATTTTGCAACTTATTACTGTGAGCAGBHCN ATAATCTCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCNHCCTCTCTWCACCT	455	1079

VK1-33_7_0_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACN HCAATCTCTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACNHCATCTCTWCACI	456	1080
VK1-33_8_0_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACG ATNHCCCTCTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACGAINHCCTCTWCACI	457	1081
VK1-33_9_0_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACG ATAATYWCCTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACGATAATYWCCTWCACI	458	1082
VK1-33_1_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACG ATAATCTCTWCACITTTGGCGGAGGGACCAAG	SWMSWMTACGATAATCTCTWCACI	459	1083
VK1-33_10_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGBHCN HCAATCTCTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATCTCTWCACI	460	1084
VK1-33_11_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGBHCN ATNHCCCTCTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATCTCTWCACI	461	1085
VK1-33_12_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGBHCN ATAATYWCCTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATCTCTWCACI	462	1086
VK1-33_13_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACN HCNHCCTCTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCCTCTWCACI	463	1087
VK1-33_14_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACN HCAATYWCCTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACNHCATCTCTWCACI	464	1088
VK1-33_15_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACG ATNHCCCTCTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACNHCATCTCTWCACI	465	1089
VK1-33_2_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGBHCN ATAATCTCTWCACITTTGGCGGAGGGACCAAG	SWMCAGBHCNHCATCTCTWCACI	466	1090
VK1-33_3_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACN HCAATCTCTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACNHCATCTCTWCACI	467	1091
VK1-33_4_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACG ATNHCCCTCTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACNHCATCTCTWCACI	468	1092
VK1-33_5_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGCAGTACG ATAATYWCCTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACGATAATYWCCTWCACI	469	1093
VK1-33_6_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMBHCG ATAATCTCTWCACITTTGGCGGAGGGACCAAG	CAGSWMBHCGATAATCTCTWCACI	470	1094
VK1-33_7_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACN HCAATCTCTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACNHCATCTCTWCACI	471	1095
VK1-33_8_1_8	CCTGAAGATATTGCAACATATTACTGTTCAGSWMTACG ATNHCCCTCTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACGATNHCCTCTWCACI	472	1096

VK1-33_9_1_8	CCTGAAGATATTGCAACATAATTACTGTGAGSWMTACG	CAGSWMTACGATAATYWCMIACACI	473	1097
VK1-33_1_2_8	ATAATYWCMTACATTTTGGCGGAGGACCAAG		474	1098
VK1-33_10_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGSWMTACG	SWMSWMTACGATAATCTCWGGACI	475	1099
VK1-33_11_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCATCTCWGGACI	476	1100
VK1-33_12_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGCAGBHCN	CAGCAGBHCNHCCTCWGGACI	477	1101
VK1-33_13_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCCTCWGGACT	478	1102
VK1-33_14_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGCAGTACN	CAGCAGTACNHCNHCCTCWGGACT	479	1103
VK1-33_15_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACNHCATCTCWGGACT	480	1104
VK1-33_2_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGCAGTACG	CAGCAGTACGATNHCYWCWGGACT	481	1105
VK1-33_3_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHCNHCATCTCWGGACT	482	1106
VK1-33_4_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGCAGTACN	SWMCAGTACNHCATCTCWGGACT	483	1107
VK1-33_5_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACGATNHCCTCWGGACT	484	1108
VK1-33_6_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGSWMBHCG	CAGSWMBHCNHCATCTCWGGACT	485	1109
VK1-33_7_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHCATCTCWGGACT	486	1110
VK1-33_8_2_8	CCTGAAGATATTGCAACATAATTACTGTGAGSWMTACG	CAGSWMTACGATNHCCTCWGGACT	487	1111
VK1-33_9_2_8	ATAATCTCWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACGATAATYWCWGGACT	488	1112
VK1-33_1_3_8	CCTGAAGATATTGCAACATAATTACTGTGAGSWMTACG	SWMSWMTACGATAATCTCCTACT	489	1113

VK1-33_10_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHN	CAGCAGBHCNHCATCTCCCTACT	490	1114
VK1-33_11_3_8	HCAATCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	491	1115
VK1-33_12_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	492	1116
VK1-33_13_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	493	1117
VK1-33_14_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	494	1118
VK1-33_15_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	495	1119
VK1-33_2_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	496	1120
VK1-33_3_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	497	1121
VK1-33_4_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	498	1122
VK1-33_5_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	499	1123
VK1-33_6_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	500	1124
VK1-33_7_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	501	1125
VK1-33_8_3_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	502	1126
VK1-33_9_3_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	503	1127
VK1-39_1_0_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	504	1128
VK1-39_10_0_8	ATNHCCTCCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATCTCCCTACT	505	1129
VK1-39_11_0_8	CCGGAAGAATATGGCAACAATTACTGTGACGAGBHCHG	CAGCAGBHCNHCATCTCCCTACT	506	1130

VK1- 39_12_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAT ACAGTBHCTWCACITTTGGCGGAGGACCAAG	CAGCAAVNATACAGTBHCTWCACI	507	1131
VK1- 39_13_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCB HCNHCACIWCACITTTGGCGGAGGACCAAG	CAGCAAAAGCBHCNHCACITTWCACT	508	1132
VK1- 39_14_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCB HCAGTBHCTWCACITTTGGCGGAGGACCAAG	CAGCAAAAGCBHCAGTBHCTWCACI	509	1133
VK1- 39_15_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACNHCACIWCACITTTGGCGGAGGACCAAG	CAGCAAAAGCTACNHCACITTWCACT	510	1134
VK1-39_2_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAT ACAGTACTIWCACITTTGGCGGAGGACCAAG	SWMCAAVNATACAGTACTTWCACT	511	1135
VK1-39_3_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCB HCAGTACTIWCACITTTGGCGGAGGACCAAG	SWMCAAAAGCBHCAGTACTTWCACT	512	1136
VK1-39_4_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACNHCACIWCACITTTGGCGGAGGACCAAG	SWMCAAAAGCTACNHCACITTWCACT	513	1137
VK1-39_5_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACAGTBHCTWCACITTTGGCGGAGGACCAAG	SWMCAAAAGCTACAGTBHCTWCACI	514	1138
VK1-39_6_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAT ACAGTACTIWCACITTTGGCGGAGGACCAAG	CAGSWMVNATACAGTACTTWCACT	515	1139
VK1-39_7_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCB HCAGTACTIWCACITTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTACTTWCACT	516	1140
VK1-39_8_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACNHCACIWCACITTTGGCGGAGGACCAAG	CAGSWMAGCTACNHCACITTWCACT	517	1141
VK1-39_9_0_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACAGTBHCTWCACITTTGGCGGAGGACCAAG	CAGSWMAGCTACAGTBHCTWCACI	518	1142
VK1-39_1_1_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCT ACAGTACTIWCACITTTGGCGGAGGACCAAG	SWMSWMAGCTACAGTACTIWCACI	519	1143
VK1- 39_10_1_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAB HCAGTACTIWCACITTTGGCGGAGGACCAAG	CAGCAAVNABHCAGTACTIWCACI	520	1144
VK1- 39_11_1_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAT ACNHCACIWCACITTTGGCGGAGGACCAAG	CAGCAAVNATACNHCACITTWCACT	521	1145
VK1- 39_12_1_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAVNAT ACAGTBHCTWCACITTTGGCGGAGGACCAAG	CAGCAAVNATACAGTBHCTWCACI	522	1146
VK1- 39_13_1_8	CCTGAAGATTTGCAACTTACTACTGTGCAAAAGCB HCNHCACIWCACITTTGGCGGAGGACCAAG	CAGCAAAAGCBHCNHCACITTWCACT	523	1147

VK1-39_14_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAAGCB	CAGCAAAGCBHCAGTBHCMTCACI	524	1148
VK1-39_15_1_8	HCAGTBHCMTCACITTTGGCGGAGGGACCAAG			
VK1-39_15_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAAGCT	CAGCAAAGCTACNHCBCBHCMTCACT	525	1149
VK1-39_15_1_8	ACNHCBCBHCMTCACTTTTGCGGAGGGACCAAG			
VK1-39_2_1_8	CCTGAAGATTTTGCAACTTACTACTGTSWMCAAVNAT	SWMCAAVNATACAGTACTMTCACT	526	1150
VK1-39_2_1_8	ACAGTACTMTCACTTTTGCGGAGGGACCAAG			
VK1-39_3_1_8	CCTGAAGATTTTGCAACTTACTACTGTSWMCAAAGCB	SWMCAAAGCBHCAGTACTMTCACT	527	1151
VK1-39_3_1_8	HCAGTACTMTCACTTTTGCGGAGGGACCAAG			
VK1-39_4_1_8	CCTGAAGATTTTGCAACTTACTACTGTSWMCAAAGCT	SWMCAAAGCTACNHCACMTCACT	528	1152
VK1-39_4_1_8	ACNHCACMTCACTTTTGCGGAGGGACCAAG			
VK1-39_5_1_8	CCTGAAGATTTTGCAACTTACTACTGTSWMCAAAGCT	SWMCAAAGCTACAGTBHCMTCACT	529	1153
VK1-39_5_1_8	ACAGTBHCMTCACITTTTGCGGAGGGACCAAG			
VK1-39_6_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMVNAT	CAGSWMVNATACAGTACTMTCACT	530	1154
VK1-39_6_1_8	ACAGTACTMTCACTTTTGCGGAGGGACCAAG			
VK1-39_7_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCB	CAGSWMAGCBHCAGTACTMTCACT	531	1155
VK1-39_7_1_8	HCAGTACTMTCACTTTTGCGGAGGGACCAAG			
VK1-39_8_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCT	CAGSWMAGCTACNHCACMTCACT	532	1156
VK1-39_8_1_8	ACNHCACMTCACTTTTGCGGAGGGACCAAG			
VK1-39_9_1_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCT	CAGSWMAGCTACAGTBHCMTCACI	533	1157
VK1-39_9_1_8	ACAGTBHCMTCACITTTTGCGGAGGGACCAAG			
VK1-39_1_2_8	CCTGAAGATTTTGCAACTTACTACTGTSWMSWMAGCT	SWMSWMAGCTACAGTACTWGGACT	534	1158
VK1-39_1_2_8	ACAGTACTWGGACTTTTGCGGAGGGACCAAG			
VK1-39_10_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAVNAB	CAGCAAAVNABHCAGTACTWGGACT	535	1159
VK1-39_10_2_8	HCAGTACTWGGACTTTTGCGGAGGGACCAAG			
VK1-39_11_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAVNAT	CAGCAAVNATACNHCACMTWGGACT	536	1160
VK1-39_11_2_8	ACNHCACMTWGGACTTTTGCGGAGGGACCAAG			
VK1-39_12_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAVNAT	CAGCAAVNATACAGTBHCWGGACT	537	1161
VK1-39_12_2_8	ACAGTBHCWGGACTTTTGCGGAGGGACCAAG			
VK1-39_13_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAAGCB	CAGCAAAGCBHCNHCACMTWGGACT	538	1162
VK1-39_13_2_8	HCNHCACMTWGGACTTTTGCGGAGGGACCAAG			
VK1-39_14_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAAGCB	CAGCAAAGCBHCAGTBHCWGGACT	539	1163
VK1-39_14_2_8	HCAGTBHCWGGACTTTTGCGGAGGGACCAAG			
VK1-39_15_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGCAAAGCT	CAGCAAAGCTACNHCBCBHCWGGACT	540	1164
VK1-39_15_2_8	ACNHCBCBHCWGGACTTTTGCGGAGGGACCAAG			

VK1-39_2_2_8	CCTGAAGATTTTGCAACTTACTACTGTSWCAAAT ACAGTACTGGACTTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACTWGGACT	541	1165
VK1-39_3_2_8	CCITGAAGATTTTGCAACTTACTACTGTSWCAAAGCB HCAGTACTWGGACTTTTGGCGAGGGACCAAG	SWMCAAAAGCBHCAGTACTWGGACT	542	1166
VK1-39_4_2_8	CCITGAAGATTTTGCAACTTACTACTGTSWCAAAGCT ACNHCACCTWGGACTTTTGGCGAGGGACCAAG	SWMCAAAAGCTACNHCACCTWGGACT	543	1167
VK1-39_5_2_8	CCTGAAGATTTTGCAACTTACTACTGTSWCAAAGCT ACAGTBHCWGGACTTTTGGCGAGGGACCAAG	SWMCAAAAGCTACAGTBHCWGGACT	544	1168
VK1-39_6_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMVNAT ACAGTACTWGGACTTTTGGCGAGGGACCAAG	CAGSWMVNATAACAGTACTWGGACT	545	1169
VK1-39_7_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCB HCAGTACTWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGCBHCAGTACTWGGACT	546	1170
VK1-39_8_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCT ACNHCACCTWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGCTACNHCACCTWGGACT	547	1171
VK1-39_9_2_8	CCTGAAGATTTTGCAACTTACTACTGTCAGSWMAGCT ACAGTBHCWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGCTACAGTBHCWGGACT	548	1172
VK1-39_1_3_8	CCITGAAGATTTTGCAACTTACTACTGTSWMSWMAGCT ACAGTACTCCTACTTTTGGCGAGGGACCAAG	SMSWMAGCTACAGTACTCCTACT	549	1173
VK1-39_10_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAVNAB HCAGTACTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNABHCAGTACTCCTACT	550	1174
VK1-39_11_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAVNAT ACNHCACCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNAIACNHCACCTCCTACT	551	1175
VK1-39_12_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAVNAT ACAGTBHCCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNATAACAGTBHCCCTACT	552	1176
VK1-39_13_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAAAGCB HCNHCACCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAAAGCBHCNHCACCTCCTACT	553	1177
VK1-39_14_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAAAGCB HCAGTBHCCCTACTTTTGGCGAGGGACCAAG	CAGCAAAAGCBHCAGTBHCCCTACT	554	1178
VK1-39_15_3_8	CCITGAAGATTTTGCAACTTACTACTGTCAGCAAAAGCT ACNHCBHCCCTACTTTTGGCGAGGGACCAAG	CAGCAAAAGCTACNHCBHCCCTACT	555	1179
VK1-39_2_3_8	CCTGAAGATTTTGCAACTTACTACTGTSWCAAAT ACAGTACTCCTACTTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACTCCTACT	556	1180
VK1-39_3_3_8	CCTGAAGATTTTGCAACTTACTACTGTSWCAAAGCB HCAGTACTCCTACTTTTGGCGAGGGACCAAG	SWMCAAAAGCBHCAGTACTCCTACT	557	1181

VK1-39_4_3_8	CC1GAAGATTTSCAACTTACTACTGTSMCAAAGCT	ACNHCACCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAAAGCTACNHCACCTCCTACT	558	1182
VK1-39_5_3_8	CC1GAAGATTTGCAACTTACTACTGTSMCAAAGCT	ACAGTBHCCCTACTTTTGGCGGAGGGACCAAG	SWMCAAAGCTACAGTBHCCCTACT	559	1183
VK1-39_6_3_8	CC1GAAGATTTGCAACTTACTACTGTGAGSMVUNAT	ACAGTACTCCTACTTTTGGCGGAGGGACCAAG	CAGSMVUNATACAGTACTCCTACT	560	1184
VK1-39_7_3_8	CC1GAAGATTTGCAACTTACTACTGTGAGSMVUNAT	HCAGTACTCCTACTTTTGGCGGAGGGACCAAG	CAGSMVUNATACAGTACTCCTACT	561	1185
VK1-39_8_3_8	CC1GAAGATTTGCAACTTACTACTGTGAGSMVUNAT	ACNHCACCTCCTACTTTTGGCGGAGGGACCAAG	CAGSMVUNATACAGTACTCCTACT	562	1186
VK1-39_9_3_8	CC1GAAGATTTGCAACTTACTACTGTGAGSMVUNAT	ACAGTBHCCCTACTTTTGGCGGAGGGACCAAG	CAGSMVUNATACAGTACTCCTACT	563	1187
VK2-28_1_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCCAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSSWMGCACCTCCAGACTTWCAC	564	1188
VK2-28_10_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	NACAGACTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVUNAMNACAGACTTWCAC	565	1189
VK2-28_11_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	ICSRMACCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVUNAMNACAGACTTWCAC	566	1190
VK2-28_12_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCCAGVBCCTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVUNAMNACAGACTTWCAC	567	1191
VK2-28_13_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	NASRMACCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCMNACAGACTTWCAC	568	1192
VK2-28_14_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	NACAGVBCCTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCMNACAGACTTWCAC	569	1193
VK2-28_15_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCSRMBCTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCMNACAGACTTWCAC	570	1194
VK2-28_2_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCCAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGVUNAMNACAGACTTWCAC	571	1195
VK2-28_3_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	NACAGACTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCMNACAGACTTWCAC	572	1196
VK2-28_4_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCSRMBCTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCMNACAGACTTWCAC	573	1197
VK2-28_5_0_8	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCAC	TCCAGVBCCTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCMNACAGACTTWCAC	574	1198

VK2-28_6_0_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMVMVNAC	ATGSMVMVNACTCCAGACTTTWCAC	575	1199
	TCCAGACTTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_7_0_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSMGCM	ATGSMGCMACAMNACAGACTTTWCAC	576	1200
	NACAGACTTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_8_0_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM	ATGSMGCMCACTCSRMACTTTWCAC	577	1201
	TCSRMACTTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_9_0_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM	ATGSMGCMCACTCCAGVBCMTWCAC	578	1202
	TCCAGVBCMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_1_1_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWCM	DTSSWCMCACTCCAGACTMTWCAC	579	1203
	TCCAGACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_10_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGVN	ATGCAGVNMAMNACAGACTMTWCAC	580	1204
	NACAGACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_11_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGVN	ATGCAGVNMCACTCSRMACTMTWCAC	581	1205
	TCSRMACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_12_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGVN	ATGCAGVNMCACTCCAGVBCMTWCAC	582	1206
	TCCAGVBCMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_13_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGCM	ATGCAGGCMAMNARMACTMTWCAC	583	1207
	NARMACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_14_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGCM	ATGCAGGCMAMNACAGVBCMTWCAC	584	1208
	NACAGVBCMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_15_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGCAGCM	ATGCAGGCMCACTCSRMBVBCMTWCAC	585	1209
	TCSRMBVBCMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_2_1_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVN	DTSCAGVNMCACTCCAGACTMTWCAC	586	1210
	TCCAGACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_3_1_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGCM	DTSCAGGCMAMNACAGACTMTWCAC	587	1211
	NACAGACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_4_1_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGCM	DTSCAGGCMCACTCSRMACTMTWCAC	588	1212
	TCSRMACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_5_1_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGCM	DTSCAGGCMCACTCCAGVBCMTWCAC	589	1213
	TCCAGVBCMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_6_1_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSMVMVN	ATGSMVMVNCACTCCAGACTMTWCAC	590	1214
	TCCAGACTMTWCACATTTTGGCGGAGGGACCAAG			
VK2-28_7_1_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM	ATGSMGCMAMNACAGACTMTWCAC	591	1215
	NACAGACTMTWCACATTTTGGCGGAGGGACCAAG			

VK2-28_8_1_8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMGAC TCSRMACTMTCACCTTTTGGCGGAGGACCAAG	ATGSMGCACTCSRMACTMTCACI	592	1216
VK2-28_9_1_8	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMGAC TCCAGVBCMTACCTTTTGGCGGAGGACCAAG	ATGSMGCACTCCAGVBCMTACI	593	1217
VK2-28_1_2_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMCAC TCCAGACTWGGACTTTTGGCGGAGGACCAAG	DTSSWMCACCTCCAGACTWGGACT	594	1218
VK2-28_10_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAM NACAGACTWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNAMNACAGACTWGGACT	595	1219
VK2-28_11_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAC TCSRMACTWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMACTWGGACT	596	1220
VK2-28_12_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAC TCCAGVBCWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNACTCCAGVBCWGGACT	597	1221
VK2-28_13_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAM NASRMACTWGGACTTTTGGCGGAGGACCAAG	ATGCAGGCAMNASRMACTWGGACT	598	1222
VK2-28_14_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAM NACAGVBCWGGACTTTTGGCGGAGGACCAAG	ATGCAGGCAMNACAGVBCWGGACT	599	1223
VK2-28_15_2_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAC TCSRMBWCGGACTTTTGGCGGAGGACCAAG	ATGCAGGCACCSRMBWCGGACT	600	1224
VK2-28_2_2_8	GCTGAGGATGTTGGGGTTTATTACTGCDISCAGVNAC TCCAGACTWGGACTTTTGGCGGAGGACCAAG	DTSCAGVNACTCCAGACTWGGACT	601	1225
VK2-28_3_2_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAM NACAGACTWGGACTTTTGGCGGAGGACCAAG	DTSCAGGCAMNACAGACTWGGACT	602	1226
VK2-28_4_2_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAC TCSRMACTWGGACTTTTGGCGGAGGACCAAG	DTSCAGGCACCSRMACTWGGACT	603	1227
VK2-28_5_2_8	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAC TCCAGVBCWGGACTTTTGGCGGAGGACCAAG	DTSCAGGCACCTCCAGVBCWGGACT	604	1228
VK2-28_6_2_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMVNVNAC TCCAGACTWGGACTTTTGGCGGAGGACCAAG	ATGSMVNVNACTCCAGACTWGGACT	605	1229
VK2-28_7_2_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM NACAGACTWGGACTTTTGGCGGAGGACCAAG	ATGSMGCMNACAGACTWGGACT	606	1230
VK2-28_8_2_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM TCSRMACTWGGACTTTTGGCGGAGGACCAAG	ATGSMGCACTCSRMACTWGGACT	607	1231
VK2-28_9_2_8	GCTGAGGATGTTGGGGTTTATTACTGCATGSMGCM TCCAGVBCWGGACTTTTGGCGGAGGACCAAG	ATGSMGCACTCCAGVBCWGGACT	608	1232

VK2-28_1_3_8	GCTGAGGATGTTGGGGTTTATTACTGCDISSWMGCAC TCCAGACTCCTACTTTTGGCGGAGGGACCAAG	DISSWMGCACCTCCAGACTCCTTACT	609	1233
VK2-28_10_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAM NACAGACICCTACTTTTGGCGGAGGGACCAAG	AIGCAGVNAMNACAGACTCCTTACT	610	1234
VK2-28_11_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAC TCSRMACTCCTACTTTTGGCGGAGGGACCAAG	AIGCAGVNACTCSRMACTCCTTACT	611	1235
VK2-28_12_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAC TCCAGVBCCCTACTTTTGGCGGAGGGACCAAG	AIGCAGVNACTCCAGVBCCCTTACT	612	1236
VK2-28_13_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAM NASRMACTCCTACTTTTGGCGGAGGGACCAAG	AIGCAGGCAMNASRMACTCCTTACT	613	1237
VK2-28_14_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAM NACAGVBCCCTACTTTTGGCGGAGGGACCAAG	AIGCAGGCAMNACAGVBCCCTTACT	614	1238
VK2-28_15_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAC TCSRMBVCCCTACTTTTGGCGGAGGGACCAAG	AIGCAGGCACITCSRMBVCCCTTACT	615	1239
VK2-28_2_3_8	GCTGAGGATGTTGGGGTTTATTACTGCDISCAGVNAC TCCAGACTCCTACTTTTGGCGGAGGGACCAAG	DISCAGVNACTCCAGACTCCTTACT	616	1240
VK2-28_3_3_8	GCTGAGGATGTTGGGGTTTATTACTGCDISCAGGCAM NACAGACTCCTACTTTTGGCGGAGGGACCAAG	DISCAGGCAMNACAGACTCCTTACT	617	1241
VK2-28_4_3_8	GCTGAGGATGTTGGGGTTTATTACTGCDISCAGGCAC TCSRMACTCCTACTTTTGGCGGAGGGACCAAG	DISCAGGCACITCSRMACTCCTTACT	618	1242
VK2-28_5_3_8	GCTGAGGATGTTGGGGTTTATTACTGCDISCAGGCAC TCCAGVBCCCTACTTTTGGCGGAGGGACCAAG	DISCAGGCACITCCAGVBCCCTTACT	619	1243
VK2-28_6_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVNAC TCCAGACTCCTACTTTTGGCGGAGGGACCAAG	ATGSWMVNACTCCAGACTCCTTACT	620	1244
VK2-28_7_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCAM NACAGACTCCTACTTTTGGCGGAGGGACCAAG	ATGSWMGCAMNACAGACTCCTTACT	621	1245
VK2-28_8_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCAC TCSRMACTCCTACTTTTGGCGGAGGGACCAAG	ATGSWMGCACITCSRMACTCCTTACT	622	1246
VK2-28_9_3_8	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCAC TCCAGVBCCCTACTTTTGGCGGAGGGACCAAG	ATGSWMGCACITCCAGVBCCCTTACT	623	1247
VK3-11_1_0_8	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWMAGAA GTAATTGGTWCATTTTGGCGGAGGGACCAAG	SWMSWMAGAAATAATTGGTWCAT	624	1248
VK3-11_10_0_8	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCN HCAATTGGTWCATTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATAATTGGTWCAT	625	1249

VK3- 11_11_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTNCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGTWCACCT	626	1250
VK3- 11_12_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATYCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYCTWCACCT	627	1251
VK3- 11_13_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCNCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCCTGGTWCACCT	628	1252
VK3- 11_14_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCAATYCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYCTWCACCT	629	1253
VK3- 11_15_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTNHCYCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYCTWCACCT	630	1254
VK3-11_2_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATTGGTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGTWCACCT	631	1255
VK3-11_3_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCAATGGTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATGGTWCACCT	632	1256
VK3-11_4_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTNCTGGTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATGGTWCACCT	633	1257
VK3-11_5_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTAATYCTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCCTGGTWCACCT	634	1258
VK3-11_6_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTGGTWCACCT	635	1259
VK3-11_7_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAN HCAATGGTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATGGTWCACCT	636	1260
VK3-11_8_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTNCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCCTGGTWCACCT	637	1261
VK3-11_9_0_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTAATYCTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYCTWCACCT	638	1262
VK3-11_1_1_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTAATTGGTWCACCTTTTGGCGGAGGGACCAAG	SWMSWMAGAAAGTAATTGGTWCACCT	639	1263
VK3- 11_10_1_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCHN HCAATGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATTTGGTWCACCT	640	1264
VK3- 11_11_1_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTNCTGGTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGTWCACCT	641	1265
VK3- 11_12_1_8	CCGGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATYCTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYCTWCACCT	642	1266

VK3-11_13_1_8	CCTGAAGATTTTCAGTTTATTACTGTACGACAGAN	CAGCAGAGANHCNHCCTGGMTCACT	643	1267
VK3-11_14_1_8	HCNCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCCTGGMTCACT	644	1268
VK3-11_15_1_8	CCTGAAGATTTTCAGTTTATTACTGTACGACAGAA	CAGCAGAGAGTAATYWCMTCACT	645	1269
VK3-11_2_1_8	GTNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAGTAATYWCMTCACT	646	1270
VK3-11_3_1_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAN	SWCAGBHCAGTAATYWCMTCACT	647	1271
VK3-11_4_1_8	HCNCTGGMTCACTTTTGGCGGAGGGACCAAG	SWCAGAGANHCNHCCTGGMTCACT	648	1272
VK3-11_5_1_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAA	SWCAGAGAGTAATYWCMTCACT	649	1273
VK3-11_6_1_8	GTNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATYWCMTCACT	650	1274
VK3-11_7_1_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAN	CAGSWMAGANHCNHCCTGGMTCACT	651	1275
VK3-11_8_1_8	HCNCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAGTAATYWCMTCACT	652	1276
VK3-11_9_1_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAA	CAGSWMAGAGTAATYWCMTCACT	653	1277
VK3-11_10_2_8	GTNHCYWCMTCACTTTTGGCGGAGGGACCAAG	SWMSWMAGAGTAATYWCMTCACT	654	1278
VK3-11_11_2_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAN	CAGCAGBHCNHCNHCCTGGMTCACT	655	1279
VK3-11_12_2_8	HCNCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTGGMTCACT	656	1280
VK3-11_13_2_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAN	CAGCAGBHCNHCNHCCTGGMTCACT	657	1281
VK3-11_14_2_8	HCNCTGGMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTGGMTCACT	658	1282
VK3-11_15_2_8	CCTGAAGATTTTCAGTTTATTACTGTSMCAGAGAN	CAGCAGBHCNHCNHCCTGGMTCACT	659	1283

VK3- 11_15_2_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTHCYWCWGGACTTTTGGCGAGGGACCAAG	CAGCAGAGAAGTNHCYWCWGGACT	660	1284
VK3-11_2_2_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GIAATTGGWGGACTTTTGGCGAGGGACCAAG	SWMCAGBHCAGTAATTGGWGGACT	661	1285
VK3-11_3_2_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCAATTGGWGGACTTTTGGCGAGGGACCAAG	SWMCAGAGANHCAATTGGWGGACT	662	1286
VK3-11_4_2_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTHCTGGWGGACTTTTGGCGAGGGACCAAG	SWMCAGAGAAGTNHCTGGWGGACT	663	1287
VK3-11_5_2_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTAATYWCWGGACTTTTGGCGAGGGACCAAG	SWMCAGAGAAGTAATYWCWGGACT	664	1288
VK3-11_6_2_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGSWMBHCA GTAATTGGWGGACTTTTGGCGAGGGACCAAG	CAGSWMBHCAGTAATTGGWGGACT	665	1289
VK3-11_7_2_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAN HCAATTGGWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGANHCAATTGGWGGACT	666	1290
VK3-11_8_2_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTHCTGGWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGAAGTNHCTGGWGGACT	667	1291
VK3-11_9_2_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTAATYWCWGGACTTTTGGCGAGGGACCAAG	CAGSWMAGAAGTAATYWCWGGACT	668	1292
VK3-11_1_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAA GTAATTGGCCTACTTTTGGCGAGGGACCAAG	SWMSWMAGAAGTAATTGGCCTACT	669	1293
VK3- 11_10_3_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCHN HCAATTGGCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCNHCATTTGGCCTACT	670	1294
VK3- 11_11_3_8	CCIGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTHCTGGCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTACT	671	1295
VK3- 11_12_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATYWCCTACTTTTGGCGAGGGACCAAG	CAGCAGBHCAGTAATYWCCTACT	672	1296
VK3- 11_13_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCNHTGGCCTACTTTTGGCGAGGGACCAAG	CAGCAGAGANHCNHCNCTGGCCTACT	673	1297
VK3- 11_14_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAN HCAATYWCCTACTTTTGGCGAGGGACCAAG	CAGCAGAGANHCAATYWCCTACT	674	1298
VK3- 11_15_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAA GTHCYWCCTACTTTTGGCGAGGGACCAAG	CAGCAGAGAAGTNHCYWCCTACT	675	1299
VK3-11_2_3_8	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCA GTAATTGGCCTACTTTTGGCGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTACT	676	1300

VK3-11_3_3_8	CCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAGAN	SWMCAGAGANHCAATTGGCCCTACT	677	1301
VK3-11_4_3_8	HCAATTGGCCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAGTNHCTGGCCCTACT	678	1302
VK3-11_5_3_8	CCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAGAA	SWMCAGAGAGTNHCTGGCCCTACT	679	1303
VK3-11_6_3_8	GINHCTGGCCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAGTAATYWCCTACT	680	1304
VK3-11_7_3_8	CCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAGAA	CAGSMBHCAGTAATTGGCCCTACT	681	1305
VK3-11_8_3_8	GTAATYWCCTACTTTTGGCGGAGGGACCAAG	CAGSMBHCAGTAATTGGCCCTACT	682	1306
VK3-11_9_3_8	CCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAGAA	CAGSMBHCAGTAATTGGCCCTACT	683	1307
VK3-15_1_0_8	GTAATYWCCTACTTTTGGCGGAGGGACCAAG	SWMSMTACAAATAATTGGTWCCTACT	684	1308
VK3-15_10_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAG	CAGCAGBHCNHCATTAATTGGTWCCTACT	685	1249
VK3-15_11_0_8	HCNHCATTAATTGGTWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATTAATTGGTWCCTACT	686	1309
VK3-15_12_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAG	CAGCAGBHCNHCATTAATTGGTWCCTACT	687	1310
VK3-15_13_0_8	HCAATAATYWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCATTAATTGGTWCCTACT	688	1311
VK3-15_14_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAG	CAGCAGTACNHCNHCATTAATTGGTWCCTACT	689	1073
VK3-15_15_0_8	ACNHCNHCATTAATTGGTWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCATTAATTGGTWCCTACT	690	1312
VK3-15_2_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAG	SWMCAGBHCATTAATTGGTWCCTACT	691	1313
VK3-15_3_0_8	HCAATAATYWCCTACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCNHCATTAATTGGTWCCTACT	692	1314
VK3-15_4_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSMWCAGAG	SWMCAGTACNHCNHCATTAATTGGTWCCTACT	693	1315

VK3-15_5_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMCAGT	SWMCAGTACAATAATYWCCTWCACT	694	1316
VK3-15_6_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMB	CAGSWMBHCAATAATTTGGTWCCT	695	1317
VK3-15_7_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMT	CAGSWMTACNHCATAATTTGGTWCCT	696	1318
VK3-15_8_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMT	CAGSWMTACNHCATAATTTGGTWCCT	697	1319
VK3-15_9_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMT	CAGSWMTACNHCATAATTTGGTWCCT	698	1320
VK3-15_1_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMSWMT	SWMSWMTACAATAATTTGGMTCACT	699	1321
VK3-15_10_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGB	CAGCAGBHCNHCATAATTTGGMTCACT	700	1264
VK3-15_11_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGB	CAGCAGBHCNHCATAATTTGGMTCACT	701	1322
VK3-15_12_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGB	CAGCAGBHCNHCATAATTTGGMTCACT	702	1323
VK3-15_13_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGT	CAGCAGTACNHCNHCATAATTTGGMTCACT	703	1324
VK3-15_14_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGT	CAGCAGTACNHCATAATTTGGMTCACT	704	1088
VK3-15_15_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGT	CAGCAGTACAATAATTTGGMTCACT	705	1325
VK3-15_2_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMCAGB	SWMCAGBHCATAATAATTTGGMTCACT	706	1326
VK3-15_3_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMCAGT	SWMCAGTACNHCATAATTTGGMTCACT	707	1327
VK3-15_4_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMCAGT	SWMCAGTACAATAATTTGGMTCACT	708	1328
VK3-15_5_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTSWMCAGT	SWMCAGTACAATAATTTGGMTCACT	709	1329
VK3-15_6_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMB	CAGSWMBHCAATAATTTGGMTCACT	710	1330

VK3-15_7_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGSWMT	CAGSWMTACNHCAATTGGMICACT	711	1331
VK3-15_8_1_8	ACNHCAATTGGMTACATTTTGGCGGAGGACCAAG	CAGSWMTACNHCAATTGGMICACT	712	1332
VK3-15_9_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGSWMT	CAGSWMTACAAATNHCTGGMICACT	713	1333
VK3-15_10_1_8	ACAAATTAATTCAGTTTGGCGGAGGACCAAG	CAGSWMTACAAATAATYWCMTACACT	714	1334
VK3-15_11_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGSWMT	SWMSWMTACAAATAATTTGGWGGACT	715	1279
VK3-15_12_1_8	ACAAATTAATTCAGTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATTGGWGGACT	716	1335
VK3-15_13_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGCAGB	CAGCAGBHCNHCAATTGGWGGACT	717	1336
VK3-15_14_1_8	HCNHCAATTGGWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATAATYWCWGGACT	718	1337
VK3-15_15_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGCAGT	CAGCAGTACNHCHCTGGWGGACT	719	1103
VK3-15_16_1_8	ACNHCAATTGGWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACNHCAATAATYWCWGGACT	720	1338
VK3-15_17_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGCAGT	CAGCAGTACAAATNHCHCTGGWGGACT	721	1339
VK3-15_18_1_8	ACAAATTAATTCAGTTTGGCGGAGGACCAAG	SWMCAGBHCNHCAATAATTTGGWGGACT	722	1340
VK3-15_19_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGCAGT	SWMCAGTACNHCAATTGGWGGACT	723	1341
VK3-15_20_1_8	ACNHCAATTGGWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATNHCTGGWGGACT	724	1342
VK3-15_21_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGSWMB	SWMCAGTACAAATAATYWCWGGACT	725	1343
VK3-15_22_1_8	HCNHCAATTGGWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHCNHCAATAATTTGGWGGACT	726	1344
VK3-15_23_1_8	CAGCTGGAAGATTTTGCAGTTTATTACTGTGAGSWMT	CAGSWMTACNHCAATTGGWGGACT	727	1345
VK3-15_24_1_8	ACNHCAATTGGWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATNHCTGGWGGACT		

VK3- 20_10_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGBHC HCAGTAGTTCACATTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTAGTTWCACI	745	1361
VK3- 20_11_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGBHC GAVNCAGTTCACATTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTAGTTWCACI	746	1362
VK3- 20_12_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGBHC GAAGTBHCTWCACATTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTAGTTWCACI	747	1363
VK3- 20_13_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGTACB HCVNCAGTTCACATTTTGGCGGAGGGACCAAG	CAGCAGTACBHCVCNAGTTWCACI	748	1364
VK3- 20_14_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGTACB HCAGTBHCTWCACATTTTGGCGGAGGGACCAAG	CAGCAGTACBHCAGTBHCTWCACI	749	1365
VK3- 20_15_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGCAGTACG GAVNCBHTWCACATTTTGGCGGAGGGACCAAG	CAGCAGTACGGAVNCBHTWCACI	750	1366
VK3-20_2_0_8	CCTGAAGATTTGCAGTGTATTACTGTSWMCAGBHC GAAGTAGTTCACATTTTGGCGGAGGGACCAAG	SWMCAGBHCBCAGTAGTTWCACI	751	1367
VK3-20_3_0_8	CCTGAAGATTTGCAGTGTATTACTGTSWMCAGTACB HCAGTAGTTCACATTTTGGCGGAGGGACCAAG	SWMCAGTACBHCAGTAGTTWCACI	752	1368
VK3-20_4_0_8	CCTGAAGATTTGCAGTGTATTACTGTSWMCAGTACG GAVNCAGTTCACATTTTGGCGGAGGGACCAAG	SWMCAGTACGGAVNCAGTTWCACI	753	1369
VK3-20_5_0_8	CCTGAAGATTTGCAGTGTATTACTGTSWMCAGTACG GAAGTBHCTWCACATTTTGGCGGAGGGACCAAG	SWMCAGTACGGAAAGTBHCTWCACI	754	1370
VK3-20_6_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGSWMBHC GAAGTAGTTCACATTTTGGCGGAGGGACCAAG	CAGSWMBHCBCAGTAGTTWCACI	755	1371
VK3-20_7_0_8	CCGGAAGATTTGCAGTGTATTACTGTCAGSWMTACB HCAGTAGTTCACATTTTGGCGGAGGGACCAAG	CAGSWMTACBHCAGTAGTTWCACI	756	1372
VK3-20_8_0_8	CCTGAAGATTTGCAGTGTATTACTGTCAGSWMTACG GAVNCAGTTCACATTTTGGCGGAGGGACCAAG	CAGSWMTACGGAVNCAGTTWCACI	757	1373
VK3-20_9_0_8	CCTGAAGATTTGCAGTGTATTACTGTCAGSWMTACG GAAGTBHCTWCACATTTTGGCGGAGGGACCAAG	CAGSWMTACGGAAAGTBHCTWCACI	758	1374
VK3-20_1_1_8	CCGGAAGATTTGCAGTGTATTACTGTSWMSWMTACG GAAGTAGTTCACATTTTGGCGGAGGGACCAAG	SWMSWMTACGGAAAGTAGTTCACI	759	1375
VK3- 20_10_1_8	CCTGAAGATTTGCAGTGTATTACTGTCAGCAGBHC HCAGTAGTTCACATTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTAGTTCACI	760	1376
VK3- 20_11_1_8	CCTGAAGATTTGCAGTGTATTACTGTCAGCAGBHC GAVNCAGTTCACATTTTGGCGGAGGGACCAAG	CAGCAGBHCBCAGTAGTTCACI	761	1377

VK3- 20_12_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGBHC	CAGCAGBHC	762	1378
VK3- 20_13_1_8	GAAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHC	763	1379
VK3- 20_14_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	764	1380
VK3- 20_15_1_8	HCVNCAGTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	765	1381
VK3- 20_16_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	766	1382
VK3- 20_17_1_8	HCAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	767	1383
VK3- 20_18_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	768	1384
VK3- 20_19_1_8	GAAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	769	1385
VK3- 20_20_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	770	1386
VK3- 20_21_1_8	HCAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	771	1387
VK3- 20_22_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	772	1388
VK3- 20_23_1_8	GAAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	773	1389
VK3- 20_24_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	774	1390
VK3- 20_25_1_8	HCAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	775	1391
VK3- 20_26_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	776	1392
VK3- 20_27_1_8	GAAGTBHCHMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACB	777	1393
VK3- 20_28_1_8	CCGGAAGATTTTGCAGTGTTACTTACTGTCTCAGCAGTACB	CAGCAGTACB	778	1394

VK3- 20_14_2_8	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACB HCAGTBHCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCAGTBHCWGGACT	779	1395
VK3- 20_15_2_8	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACG GAVNCBHCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACGGAVNCBHCWGGACT	780	1396
VK3-20_2_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGBHC GAAGTAGTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAGTWGGACT	781	1397
VK3-20_3_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACB HCAGTAGTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACBHCAGTAGTWGGACT	782	1398
VK3-20_4_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAVNCAGTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACGGAVNCAGTWGGACT	783	1399
VK3-20_5_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAAGTBHCWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACGGAAAGTBHCWGGACT	784	1400
VK3-20_6_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAAGTAGTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAGTWGGACT	785	1401
VK3-20_7_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACB HCAGTAGTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACBHCAGTAGTWGGACT	786	1402
VK3-20_8_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAVNCAGTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGGAVNCAGTWGGACT	787	1403
VK3-20_9_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAAGTBHCWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACGGAAAGTBHCWGGACT	788	1404
VK3-20_1_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAAGTAGTCCCTACTTTTGGCGGAGGGACCAAG	SWMSWMTACGGAAAGTAGTCCCTACT	789	1405
VK3- 20_10_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACB HCAGTAGTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAGTCCCTACT	790	1406
VK3- 20_11_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAVNCAGTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAGTCCCTACT	791	1407
VK3- 20_12_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAAGTBHCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAGTCCCTACT	792	1408
VK3- 20_13_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACB HCVNCAGTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCAGTAGTCCCTACT	793	1409
VK3- 20_14_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACB HCAGTBHCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACBHCAGTAGTCCCTACT	794	1410
VK3- 20_15_3_8	CCIGAAAGATTTTGCAGTGTATTACTGTSMCAGTACG GAVNCBHCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACGGAVNCBHCWGGACT	795	1411

VK3-20_2_3_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGBHC	SWMCAGBHC	796	1412
VK3-20_3_3_8	GAAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGBHC	797	1413
VK3-20_4_3_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACB	SWMCAGTACB	798	1414
VK3-20_5_3_8	HCAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	799	1415
VK3-20_6_3_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG	SWMCAGTACB	800	1416
VK3-20_7_3_8	GAAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	801	1417
VK3-20_8_3_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG	SWMCAGTACB	802	1418
VK3-20_9_3_8	GAAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	803	1419
Jumping Trimer				
VK1-05_t1_0_8	CCTGATGATTTTGCAGTGTATTACTGTSMCAGBHC	SWMCAGBHC	804	1420
VK1-05_t1_1_8	RMBCTACTWCACTTTTGGCGGAGGACCAAG	SWMCAGBHC	805	1421
VK1-05_t1_2_8	CCTGATGATTTTGCAGTGTATTACTGTSMCAGTACB	SWMCAGTACB	806	1422
VK1-05_t1_3_8	HCAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	807	1423
VK1-05_t2_0_8	CCTGATGATTTTGCAGTGTATTACTGTSMCAGTACG	SWMCAGTACB	808	1424
VK1-05_t2_1_8	GAAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	809	1425
VK1-05_t2_2_8	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACG	SWMCAGTACB	810	1426
VK1-05_t2_3_8	GAAGTAGTCTACTTTTGGCGGAGGACCAAG	SWMCAGTACB	811	1427

VK1- 05_t3_0_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCV RMAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCTWCACT	812	1428
VK1- 05_t3_1_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCV RMAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCTWCACT	813	1429
VK1- 05_t3_2_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCV RMAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCTWCACT	814	1430
VK1- 05_t3_3_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGBHCV RMAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTYWCTWCACT	815	1431
VK1- 05_t4_0_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMBBCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCWCTWCACT	816	1432
VK1- 05_t4_1_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMBBCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCWCTWCACT	817	1433
VK1- 05_t4_2_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMBBCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCWGGACT	818	1434
VK1- 05_t4_3_8	CCTGATGATTTTGCAACTTATTACTGCCAGCAGTACV RMBBCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCYWCWGGACT	819	1435
VK1- 12_t1_0_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCNHCITCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCNHCITCCTWCACT	820	1436
VK1- 12_t1_1_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCNHCITCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCNHCITCCTWCACT	821	1437
VK1- 12_t1_2_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCNHCITCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCNHCITCCTWCACT	822	1438
VK1- 12_t1_3_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCNHCITCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCNHCITCCTWCACT	823	1439
VK1- 12_t2_0_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTYWCTWCACT	824	1440
VK1- 12_t2_1_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTYWCTWCACT	825	1441
VK1- 12_t2_2_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTYWCTWCACT	826	1442
VK1- 12_t2_3_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN HCAGTYWCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTYWCTWCACT	827	1443
VK1- 12_t3_0_8	CCTGAAGATTTTGCAACTTATTACTGTGTCAGCAGRNAN ATNHCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCATNHCYWCWCTWCACT	828	1444

VK1- 12_t3_1_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGR _{NAA} ATNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGR _{NAAA} ATNHCYWCMTCACT	829	1445
VK1- 12_t3_2_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGR _{NAA} ATNHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGR _{NAAA} ATNHCYWCWGGACT	830	1446
VK1- 12_t3_3_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGR _{NAA} ATNHCYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGR _{NAAA} ATNHCYWCCCTACT	831	1447
VK1- 12_t4_0_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGGCAN HCNHCYWCWCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCYWCCTWCACT	832	1448
VK1- 12_t4_1_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGGCAN HCNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCYWCMTCACT	833	1449
VK1- 12_t4_2_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGGCAN HCNHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCYWCWGGACT	834	1450
VK1- 12_t4_3_8	CCTGAAGATTTTGGAACTTATTACTGTCTCAGCAGGCAN HCNHCYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCNHCYWCCCTACT	835	1451
VK1- 33_t1_0_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCNHCCTCWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	836	1452
VK1- 33_t1_1_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCNHCCTCMCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	837	1453
VK1- 33_t1_2_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCNHCCTCWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWGGACT	838	1454
VK1- 33_t1_3_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCNHCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTCCTACT	839	1455
VK1- 33_t2_0_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCAATYWCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	840	1456
VK1- 33_t2_1_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCAATYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	841	1457
VK1- 33_t2_2_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWGGACT	842	1458
VK1- 33_t2_3_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN HCAATYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCCCTACT	843	1459
VK1- 33_t3_0_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN ATNHCYWCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	844	1460
VK1- 33_t3_1_8	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCN ATNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCCTCTWCACT	845	1461

VK1- 33_t3_2_8	CCTGAAGATATTGCAACATATTACTGTACGAGBHCG ATNHCYWCWGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTATNHCYWCWGACT	846	1462
VK1- 33_t3_3_8	CCTGAAGATATTGCAACATATTACTGTACGAGBHCG ATNHCYWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTATNHCYWCCTACT	847	1463
VK1- 33_t4_0_8	CCTGAAGATATTGCAACATATTACTGTACGAGTACN HCNHCYWCITWCACITTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCITWCACIT	848	1464
VK1- 33_t4_1_8	CCTGAAGATATTGCAACATATTACTGTACGAGTACN HCNHCYWCITWCACITTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCITWCACIT	849	1465
VK1- 33_t4_2_8	CCTGAAGATATTGCAACATATTACTGTACGAGTACN HCNHCYWCWGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCWGACT	850	1466
VK1- 33_t4_3_8	CCTGAAGATATTGCAACATATTACTGTACGAGTACN HCNHCYWCCTACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTACT	851	1467
VK1- 39_t1_0_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCNHCACITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNABHCNHCACITWCACIT	852	1468
VK1- 39_t1_1_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCNHCACITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNABHCNHCACITWCACIT	853	1469
VK1- 39_t1_2_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCNHCACITWCAGCTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCNHCACITWCAGCT	854	1470
VK1- 39_t1_3_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCNHCACITCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCNHCACITCCTACT	855	1471
VK1- 39_t2_0_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCAGTBHCITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTBHCITWCACIT	856	1472
VK1- 39_t2_1_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCAGTBHCITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTBHCITWCACIT	857	1473
VK1- 39_t2_2_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCAGTBHCWGAGCTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTBHCWGAGCT	858	1474
VK1- 39_t2_3_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAB HCAGTBHCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTBHCCTACT	859	1475
VK1- 39_t3_0_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAT ACNHCBCITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCBCITWCACIT	860	1476
VK1- 39_t3_1_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAT ACNHCBCITWCACITTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCBCITWCACIT	861	1477
VK1- 39_t3_2_8	CCTGAAGATTTTGCACCTTACTACTGTACGCAAVNAT ACNHCBCITWCAGCTTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCBCITWCAGCT	862	1478

VK1- 39_t3_3_8	CCTGAAGATTTGCAACTTACTACTGTCTCAGCAAVNAT ACNHCBCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNATAACNHCBCCTACT	863	1479
VK1- 39_t4_0_8	CCTGAAGATTTGCAACTTACTACTGTCTCAGCAAGCB HCNHCBCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAAGCBHCNHCBCCTWCACT	864	1480
VK1- 39_t4_1_8	CCTGAAGATTTGCAACTTACTACTGTCTCAGCAAGCB HCNHCBCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAAGCBHCNHCBCCTWCACT	865	1481
VK1- 39_t4_2_8	CCTGAAGATTTGCAACTTACTACTGTCTCAGCAAGCB HCNHCBCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAAGCBHCNHCBCCTWCACT	866	1482
VK1- 39_t4_3_8	CCTGAAGATTTGCAACTTACTACTGTCTCAGCAAGCB HCNHCBCCTWCACTTTTGGCGGAGGGACCAAG	CAGCAAGCBHCNHCBCCTWCACT	867	1483
VK2- 28_t1_0_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NASRMACCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACCTWCACT	868	1484
VK2- 28_t1_1_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NASRMACCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACCTWCACT	869	1485
VK2- 28_t1_2_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NASRMACCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACCTWCACT	870	1486
VK2- 28_t1_3_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NASRMACCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACCTWCACT	871	1487
VK2- 28_t2_0_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NACAGVBCCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGVBCCTWCACT	872	1488
VK2- 28_t2_1_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NACAGVBCCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGVBCCTWCACT	873	1489
VK2- 28_t2_2_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NACAGVBCCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGVBCCTWCACT	874	1490
VK2- 28_t2_3_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAM NACAGVBCCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGVBCCTWCACT	875	1491
VK2- 28_t3_0_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAC TCSRMBCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMBCTWCACT	876	1492
VK2- 28_t3_1_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAC TCSRMBCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMBCTWCACT	877	1493
VK2- 28_t3_2_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAC TCSRMBCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMBCTWCACT	878	1494
VK2- 28_t3_3_8	GCTGAGGATGTTGGGTTTATTACTGCAIGCAGVNAC TCSRMBCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMBCTWCACT	879	1495

VK2- 28_t4_0_8	GCTGAGGATGTTGGGTTTTATTACTGCAIGCAGGCAM NASRMVBCTWCACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMVBCIWCCTACT	880	1496
VK2- 28_t4_1_8	GCTGAGGATGTTGGGTTTTATTACTGCAIGCAGGCAM NASRMVBCMIACCTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMVBCMICTACT	881	1497
VK2- 28_t4_2_8	GCTGAGGATGTTGGGTTTTATTACTGCAIGCAGGCAM NASRMVBCWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMVBCWGGGACT	882	1498
VK2- 28_t4_3_8	GCTGAGGATGTTGGGTTTTATTACTGCAIGCAGGCAM NASRMVBCCTACTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMVBCCTCTACT	883	1499
VK3- 11_t1_0_8	CCIGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCNHCIGGIWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGTWCCTACT	884	1500
VK3- 11_t1_1_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCNHCIGGMIACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGMICTACT	885	1501
VK3- 11_t1_2_8	CCIGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCNHCIGGAGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGAGGACT	886	1502
VK3- 11_t1_3_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCNHCIGGCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGCTCTACT	887	1503
VK3- 11_t2_0_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCAATYWCITWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCITWCACT	888	1456
VK3- 11_t2_1_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCAATYWCMIACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCMICTACT	889	1457
VK3- 11_t2_2_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCAATYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCWGGGACT	890	1458
VK3- 11_t2_3_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN HCAATYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCCCTCTACT	891	1459
VK3- 11_t3_0_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN GINHCYWCITWCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCITWCACT	892	1504
VK3- 11_t3_1_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN GINHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCMTCTACT	893	1505
VK3- 11_t3_2_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN GINHCYWCWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCWGGGACT	894	1506
VK3- 11_t3_3_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGBHCN GINHCYWCCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCYWCCTCTACT	895	1507
VK3- 11_t4_0_8	CCTGAAGATTTGCAGTTTATTACTGTCAGCAGAGAN HCNHCYWCITWCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCYWCITWCACT	896	1508

VK3- 11_t4_1_8	CCTGAAGATTTTGCAGTTTATTACTGTACGACAGAN HCNECYWCMTCACITTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCYWCMTCACT	897	1509
VK3- 11_t4_2_8	CCTGAAGATTTTGCAGTTTATTACTGTACGACAGAN HCNECYWCGWGACITTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCYWCWGGACI	898	1510
VK3- 11_t4_3_8	CCTGAAGATTTTGCAGTTTATTACTGTACGACAGAN HCNECYWCCCIACITTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCYWCCTTACT	899	1511
VK3- 15_t1_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNEHCNHCIGGTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGTWCACI	900	1500
VK3- 15_t1_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNEHCNHCIGGTCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGTCACI	901	1501
VK3- 15_t1_2_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNEHCNHCIGGAGACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGAGGACT	902	1502
VK3- 15_t1_3_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNEHCNHCIGGCTACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCIGGCTACT	903	1503
VK3- 15_t2_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNHCAATYWCITWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCITWCACI	904	1456
VK3- 15_t2_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNHCAATYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCMTCACT	905	1457
VK3- 15_t2_2_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNHCAATYWCWGACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCWGGACT	906	1458
VK3- 15_t2_3_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCNHCAATYWCCTTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATYWCCTTACT	907	1459
VK3- 15_t3_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCAATNHCYWCITWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATNHCYWCITWCACI	908	1512
VK3- 15_t3_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCAATNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATNHCYWCMTCACT	909	1513
VK3- 15_t3_2_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCAATNHCYWCWGACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATNHCYWCWGGACT	910	1514
VK3- 15_t3_3_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGB HCAATNHCYWCCTTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCNHCATNHCYWCCTTACT	911	1515
VK3- 15_t4_0_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGT ACNHCNHCYWCITWCACITTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCNHCYWCITWCACI	912	1464
VK3- 15_t4_1_8	CAGTCTGAAGATTTTGCAGTTTATTACTGTACGACAGT ACNHCNHCYWCMTCACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCNHCYWCMTCACT	913	1465

VK3- 15_t4_2_8	CAGTCGAGATTTTGCAGTTTATTACTGTACGAGT ACNHCNHCYWCWGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCYWCWGGACT	914	1466
VK3- 15_t4_3_8	CAGTCGAGATTTTGCAGTTTATTACTGTACGAGT ACNHCNHCYWCWCCCIACITTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCYWCWCCCTACT	915	1467
VK3- 20_t1_0_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	916	1516
VK3- 20_t1_1_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	917	1517
VK3- 20_t1_2_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	918	1518
VK3- 20_t1_3_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	919	1519
VK3- 20_t2_0_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	920	1520
VK3- 20_t2_1_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	921	1521
VK3- 20_t2_2_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	922	1522
VK3- 20_t2_3_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	923	1523
VK3- 20_t3_0_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	924	1524
VK3- 20_t3_1_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	925	1525
VK3- 20_t3_2_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	926	1526
VK3- 20_t3_3_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGBHCB HCVNCAGTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBHCVNCAGTTCWCACT	927	1527
VK3- 20_t4_0_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGTACB HCVNCBHCTWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCVNCBHCTWCACI	928	1528
VK3- 20_t4_1_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGTACB HCVNCBHCTWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCVNCBHCTWCACI	929	1529
VK3- 20_t4_2_8	CCGSAAGATTTTGCAGTGTATTACTGTACGAGTACB HCVNCBHCTWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCVNCBHCTWCACI	930	1530

VK3- 20_t4_3_8	CCTGAAGATTTTGCAGTGTTTACTGTACAGTACB HCVNCHCCCTACTTTTGGCGAGGGACCAAG	CAGCAGTACBHCNCHCCTACT	931	1531
VK4- 01_t1_0_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCNHCACITWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	932	1532
VK4- 01_t1_1_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCNHCACITWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	933	1533
VK4- 01_t1_2_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCNHCACITWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	934	1534
VK4- 01_t1_3_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCNHCACITWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	935	1535
VK4- 01_t2_0_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCAGTBHCTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	936	1520
VK4- 01_t2_1_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCAGTBHCTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	937	1521
VK4- 01_t2_2_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCAGTBHCTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	938	1522
VK4- 01_t2_3_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC HCAGTBHCTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	939	1523
VK4- 01_t3_0_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC ACNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	940	1536
VK4- 01_t3_1_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC ACNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	941	1537
VK4- 01_t3_2_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC ACNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	942	1538
VK4- 01_t3_3_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGBHC ACNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGBHCBCNHCACITWCACI	943	1539
VK4- 01_t4_0_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGTACB HCNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGTACBHCNHCBCHTWCACI	944	1540
VK4- 01_t4_1_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGTACB HCNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGTACBHCNHCBCHTWCACI	945	1541
VK4- 01_t4_2_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGTACB HCNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGTACBHCNHCBCHTWCACI	946	1542
VK4- 01_t4_3_8	GCTGAAGATGTGGCAGTTTATTACTGTACAGTACB HCNHCBCHTWCATTTTGGCGAGGGACCAAG	CAGCAGTACBHCNHCBCHTWCACI	947	1543

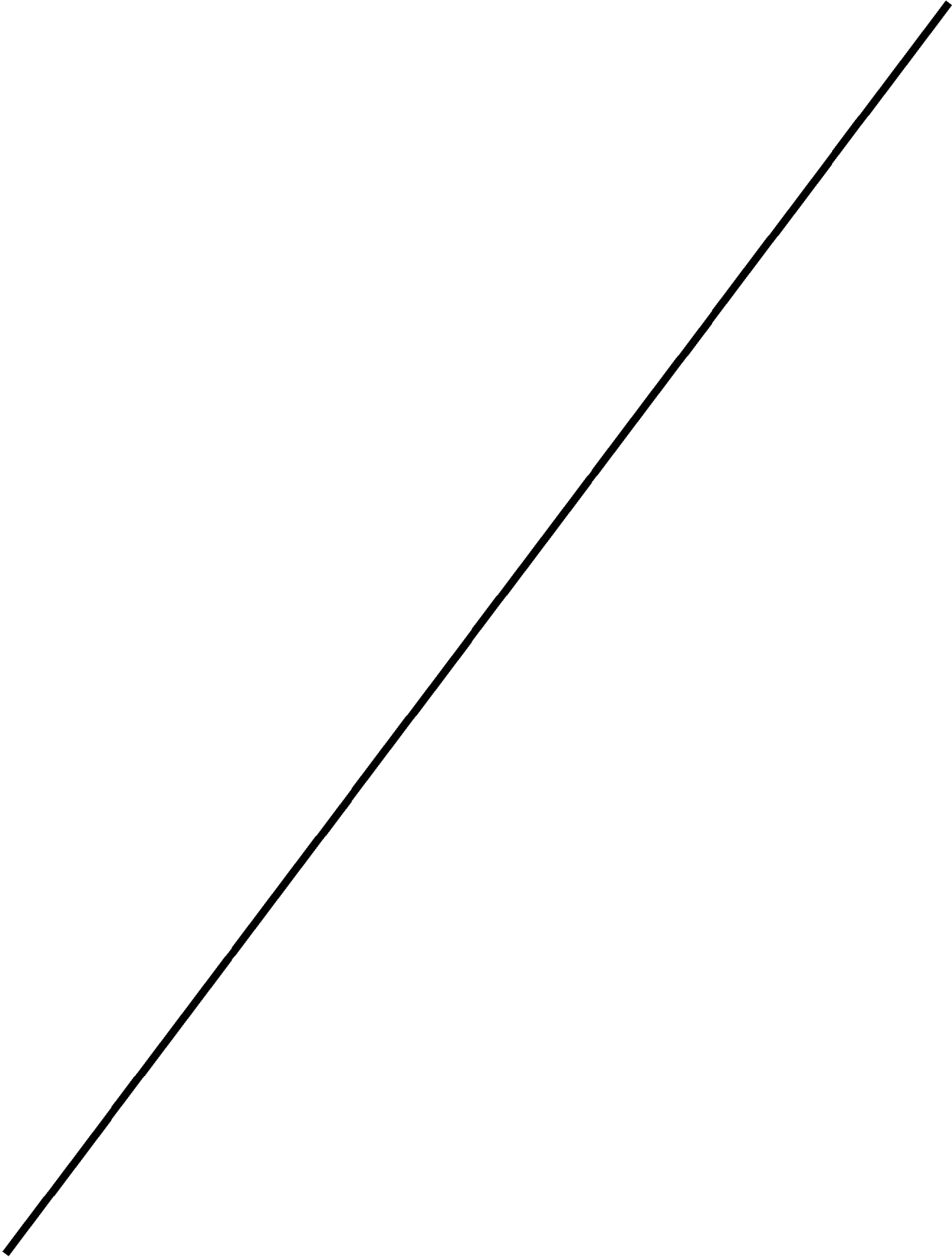


Table 6. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDRL3 length 9.

Name	Sequence of Synthesized Oligonucleotide	Portion of Oligonucleotide Corresponding to CDRL3 Proper	SEQ ID NO (Oligo)	SEQ ID NO (CDRL3 Portion)
Jumping Dimer				
VK1-05_1_0_9	CCIGATGATTTTGCACCTATTACTGCSWMSWMTACAATA GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	SWMSWMTACAATACTTACYCCTT WCACT	1544	2168
VK1-05_10_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGBHCVRMA GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACYCCTT WCACT	1545	2169
VK1-05_11_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGBHCATM BCTACYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCATMBCTACYCCTT WCACT	1546	2170
VK1-05_12_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGBHCATM GTYWCYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCATMAGTYWCYCCTT WCACT	1547	2171
VK1-05_13_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGTACVRMM BCTACYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGTACVRMBCTACYCCTT WCACT	1548	2172
VK1-05_14_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGTACVRMA GTYWCYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYCCTT WCACT	1549	2173
VK1-05_15_0_9	CCIGATGATTTTGCACCTATTACTGCCAGCAGTACAATM BCYWCYCTTTWCACCTTTGGCGGAGGACCAAG	CAGCAGTACAATMBCYWCYCCTT WCACT	1550	2174
VK1-05_2_0_9	CCIGATGATTTTGCACCTATTACTGCSWMCAGBHCATM GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	SWMCAGBHCATMAGTTACYCCTT WCACT	1551	2175
VK1-05_3_0_9	CCIGATGATTTTGCACCTATTACTGCSWMCAGTACVRMA GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYCCTT WCACT	1552	2176
VK1-05_4_0_9	CCIGATGATTTTGCACCTATTACTGCSWMCAGTACAATM BCTACYCTTTWCACCTTTGGCGGAGGACCAAG	SWMCAGTACAATMBCTACYCCTT WCACT	1553	2177
VK1-05_5_0_9	CCIGATGATTTTGCACCTATTACTGCSWMCAGTACAATA GTYWCYCTTTWCACCTTTGGCGGAGGACCAAG	SWMCAGTACAATAAGTYWCYCCTT WCACT	1554	2178
VK1-05_6_0_9	CCIGATGATTTTGCACCTATTACTGCCAGSMMBHCATM GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	CAGSMMBHCATMAGTTACYCCTT WCACT	1555	2179
VK1-05_7_0_9	CCIGATGATTTTGCACCTATTACTGCCAGSMMTACVRMA GTTACYCTTTWCACCTTTGGCGGAGGACCAAG	CAGSMMTACVRMAGTTACYCCTT WCACT	1556	2180

VK1-05_8_0_9	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATM BCTACYCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMTACAATMBCTACYCTT WCACT	1557	2181
VK1-05_9_0_9	CCGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATA GTWYCYCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMTACAATAAGTYWCYCTT WCACT	1558	2182
VK1-05_1_1_9	CCTGATGATTTTGGCAACTTATTACTGCSWMSWMTACAATA GTTACYCTMTCACTTTTGGCGGAGGACCAAG	SWMSWMTACAATAAGTTACYCTM TCACT	1559	2183
VK1-05_10_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVRMA GTTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACYCTM TCACT	1560	2184
VK1-05_11_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVAATM BCTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCVAATMBCTACYCTM TCACT	1561	2185
VK1-05_12_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVAATA GTWYCYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCVAATAAGTYWCYCTM TCACT	1562	2186
VK1-05_13_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMM BCTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTACYCTM TCACT	1563	2187
VK1-05_14_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMA GTWYCYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYCTM TCACT	1564	2188
VK1-05_15_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACAATM BCYWCYCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGTACAATMBCYWCYCTM TCACT	1565	2189
VK1-05_2_1_9	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACVRMA GTTACYCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGBHCVAATAAGTTACYCTM TCACT	1566	2190
VK1-05_3_1_9	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACVRMA GITACYCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYCTM TCACT	1567	2191
VK1-05_4_1_9	CCGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATM BCTACYCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACAATMBCTACYCTM TCACT	1568	2192
VK1-05_5_1_9	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATA GTWYCYCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACAATAAGTYWCYCTM TCACT	1569	2193
VK1-05_6_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMBHCVAATA GTTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHCVAATAAGTTACYCTM TCACT	1570	2194
VK1-05_7_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACVRMA GITACYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTTACYCTM TCACT	1571	2195
VK1-05_8_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATM BCTACYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACAATMBCTACYCTM TCACT	1572	2196
VK1-05_9_1_9	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATA GTWYCYCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACAATAAGTYWCYCTM TCACT	1573	2197

VK1-05_1_2_9	CCIGATGATTTTGCACCTTATTACTGCSWMSWMTACAATA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	SWMSWMTACAATAAGTTACYCCTW GGACT	1574	2198
VK1-05_10_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCVRMA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTTACYCCTW GGACT	1575	2199
VK1-05_11_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCVAATM BCTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVAATMBCTACYCCTW GGACT	1576	2200
VK1-05_12_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCVAATA GTYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVAATAGTYWCYCCTW GGACT	1577	2201
VK1-05_13_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACVRMM BCTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMMBCTACYCCTW GGACT	1578	2202
VK1-05_14_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACVRMA GTYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACVRMAGTYWCYCCTW GGACT	1579	2203
VK1-05_15_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACAATM BCYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACAATMBCYWCYCCTW GGACT	1580	2204
VK1-05_2_2_9	CCTGATGATTTTGCACCTTATTACTGCSWMCAGBHCVAATA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCVAATAGTTACYCCTW GGACT	1581	2205
VK1-05_3_2_9	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACVRMA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACVRMAGTTACYCCTW GGACT	1582	2206
VK1-05_4_2_9	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACAATM BCTACYCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACAATMBCTACYCCTW GGACT	1583	2207
VK1-05_5_2_9	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACAATA GTYWCYCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACAATAAGTYWCYCCTW GGACT	1584	2208
VK1-05_6_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGSWMHBHCVAATA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMHBHCVAATAGTTACYCCTW GGACT	1585	2209
VK1-05_7_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACVRMA GTTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACVRMAGTTACYCCTW GGACT	1586	2210
VK1-05_8_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACAATM BCTACYCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACAATMBCTACYCCTW GGACT	1587	2211
VK1-05_9_2_9	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACAATA GTYWCYCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACAATAAGTYWCYCCTW GGACT	1588	2212
VK1-05_1_3_9	CCTGATGATTTTGCACCTTATTACTGCSWMSWMTACAATA GTTACYCTCCTACTTTTGGCGGAGGGACCAAG	SWMSWMTACAATAAGTTACYCCTC CTACT	1589	2213
VK1-05_10_3_9	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCVRMA GTTACYCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTTACYCCTC CTACT	1590	2214

VK1- 05_11_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGCAGBHCATM BCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCATMBCTACYCTC CTACT	1591	2215
VK1- 05_12_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGCAGBHCATM GTWYCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCATAGTYWCYCTC CTACT	1592	2216
VK1- 05_13_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGCAGTACVRMM BCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTACYCTC CTACT	1593	2217
VK1- 05_14_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGCAGTACVRMA GTWYCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYCTC CTACT	1594	2218
VK1- 05_15_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGCAGTACAAATM BCYWCYCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACAAATMBCYWCYCTC CTACT	1595	2219
VK1-05_2_3_9	CCTGATGAIITTTGCAACTTATTACTGCSWNCAGBHCATM GTTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGBHCATAGTTACYCTC CTACT	1596	2220
VK1-05_3_3_9	CCTGATGAIITTTGCAACTTATTACTGCSWNCAGTACVRMA GTTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYCTC CTACT	1597	2221
VK1-05_4_3_9	CCTGATGAIITTTGCAACTTATTACTGCSWNCAGTACAAATM BCTACYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACAAATMBCTACYCTC CTACT	1598	2222
VK1-05_5_3_9	CCTGATGAIITTTGCAACTTATTACTGCSWNCAGTACAAATA GTWYCYCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACAAIAGTYWCYCTC CTACT	1599	2223
VK1-05_6_3_9	CCIGATGAIITTTGCAACTTATTACTGCCAGSWMBHCATM GTTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMBHCATAGTTACYCTC CTACT	1600	2224
VK1-05_7_3_9	CCIGATGAIITTTGCAACTTATTACTGCCAGSWMTACVRMA GTTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTTACYCTC CTACT	1601	2225
VK1-05_8_3_9	CCIGATGAIITTTGCAACTTATTACTGCCAGSWMTACAAATM BCTACYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATMBCTACYCTC CTACT	1602	2226
VK1-05_9_3_9	CCTGATGAIITTTGCAACTTATTACTGCCAGSWMTACAAATA GTWYCYCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACAAATAGTYWCYCTC CTACT	1603	2227
VK1-12_1_0_9	CCTGAAGAIITTTGCAACTTATTACTGTSWNSWMGCAAAATA GTTTCCCTTWCATTTTGGCGGAGGACCAAG	SWMSWMGCAAAATAGTTTCCCTT WCACT	1604	2228
VK1- 12_10_0_9	CCTGAAGAIITTTGCAACTTATTACTGTACAGCAGRANHCATM GTTTCCCTTWCATTTTGGCGGAGGACCAAG	CAGCAGRANHCAGTTTCCCTT WCACT	1605	2229
VK1- 12_11_0_9	CCTGAAGAIITTTGCAACTTATTACTGTACAGCAGRANAAATN HCTTCCCTTWCATTTTGGCGGAGGACCAAG	CAGCAGRANAAATNHCTTCCCTT WCACT	1606	2230
VK1- 12_12_0_9	CCTGAAGAIITTTGCAACTTATTACTGTACAGCAGRANAAATA GTWCCCTTWCATTTTGGCGGAGGACCAAG	CAGCAGRANAAATAGTYWCCTT WCACT	1607	2231

VK1- 12_13_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGGCGCANHCN HCTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCCTTCCCTT WCACT	1608	2232
VK1- 12_14_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGGCGCANHCA GTYWCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCANHCAGTYWCCCTT WCACT	1609	2233
VK1- 12_15_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGGCGCAAAATN HCYWCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCAGAAATNHCYWCCCTT WCACT	1610	2234
VK1-12_2_0_9	CCTGAAGATTTTGCAACTTATTACTGTSMCAGRANAAATA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGRANAAATAGTTTCCCTT WCACT	1611	2235
VK1-12_3_0_9	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCANHCA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCANHCAGTTTCCCTT WCACT	1612	2236
VK1-12_4_0_9	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCAAAATN HCTTCCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCAAAATNHCCTTCCCTT WCACT	1613	2237
VK1-12_5_0_9	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCAAAATA GTYWCCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCAAAATAGTYWCCCTT WCACT	1614	2238
VK1-12_6_0_9	CCTGAAGATTTTGCAACTTATTACTGTSMCAGRANAAATA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSMRANAAATAGTTTCCCTT WCACT	1615	2239
VK1-12_7_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGSMGCANHCA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSMGCANHCAGTTTCCCTT WCACT	1616	2240
VK1-12_8_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGSMGCAAAATN HCTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSMGCAAAATNHCCTTCCCTT WCACT	1617	2241
VK1-12_9_0_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGSMGCAAAATA GTYWCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSMGCAAAATAGTYWCCCTT WCACT	1618	2242
VK1-12_1_1_9	CCTGAAGATTTTGCAACTTATTACTGTSMGCAAAATA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	SWMSWGCAAAATAGTTTCCCTT TCACT	1619	2243
VK1- 12_10_1_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGCAGRANANHCA GTTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRANANHCAGTTTCCCTT TCACT	1620	2244
VK1- 12_11_1_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGCAGRANAAATN HCTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRANAAATNHCCTTCCCTT TCACT	1621	2245
VK1- 12_12_1_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGCAGRANAAATA GTYWCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRANAAATAGTYWCCCTT TCACT	1622	2246
VK1- 12_13_1_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGCAGGCANHCN HCTTCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCCTTCCCTT TCACT	1623	2247
VK1- 12_14_1_9	CCTGAAGATTTTGCAACTTATTACTGTGCAGCAGGCANHCA GTYWCCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCANHCAGTYWCCCTT TCACT	1624	2248

VK1-12_15_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCYWCCCTM TCACT	1625	2249
VK1-12_2_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGRNAAAATAGTTTCCCTM TCACT	1626	2250
VK1-12_3_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTM TCACT	1627	2251
VK1-12_4_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATNHCTTCCCTM TCACT	1628	2252
VK1-12_5_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGGCAAAATAGTYWCCCTM TCACT	1629	2253
VK1-12_6_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMRNAAATAGTTTCCCTM TCACT	1630	2254
VK1-12_7_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTM TCACT	1631	2255
VK1-12_8_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTM TCACT	1632	2256
VK1-12_9_1_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTM TCACT	1633	2257
VK1-12_1_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMSWMCANHCAGTTTCCCTM GGACT	1634	2258
VK1-12_10_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNANHCAGTTTCCCTM GGACT	1635	2259
VK1-12_11_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATNHCTTCCCTM GGACT	1636	2260
VK1-12_12_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGRNAAAATAGTYWCCCTM GGACT	1637	2261
VK1-12_13_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCTTCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTTTCCCTM GGACT	1638	2262
VK1-12_14_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCANHCAGTYWCCCTM GGACT	1639	2263
VK1-12_15_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN HCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGGCAAAATNHCTTCCCTM GGACT	1640	2264
VK1-12_2_2_9	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATN GTTTCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGRNAAAATAGTTTCCCTM GGACT	1641	2265

VK1-12_3_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTWGGACITTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTW GGACT	1642	2266
VK1-12_4_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTWGGACITTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTW GGACT	1643	2267
VK1-12_5_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTWCCCTWGGACITTTTGGCGGAGGGACCAAG	SWMCAGGCANHCAGTTTCCCTW GGACT	1644	2268
VK1-12_6_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTWGGACITTTTGGCGGAGGGACCAAG	CAGSWMRNAAATAGTTTCCCTW GGACT	1645	2269
VK1-12_7_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTWGGACITTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTW GGACT	1646	2270
VK1-12_8_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTWGGACITTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTW GGACT	1647	2271
VK1-12_9_2_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTWCCCTWGGACITTTTGGCGGAGGGACCAAG	CAGSWMGCANHCAGTTTCCCTW GGACT	1648	2272
VK1-12_1_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMSWGCANHCAGTTTCCCTC CTACT	1649	2273
VK1-12_10_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1650	2274
VK1-12_11_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1651	2275
VK1-12_12_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1652	2276
VK1-12_13_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1653	2277
VK1-12_14_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1654	2278
VK1-12_15_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGRANHCAGTTTCCCTC CTACT	1655	2279
VK1-12_2_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGRANHCAGTTTCCCTC CTACT	1656	2280
VK1-12_3_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA GTTTCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGRANHCAGTTTCCCTC CTACT	1657	2281
VK1-12_4_3_9	CCTGAAGATTTTGCAACTTATTACTGTGTSWMCAGGCANHCA HCTTCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGRANHCAGTTTCCCTC CTACT	1658	2282

VK1-12_5_3_9	CCTGAAGATTTTGCAACITATTACTGTGTCAGGCAAAATA GTWCCCTCCTACTTTTGGCGGAGGACCAAG	SWCAGGCAAAATAGTYWCCCTC CTACT	1659	2283
VK1-12_6_3_9	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMRNAAATA GTTCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMRNAAATAGTTTCCCTC CTACT	1660	2284
VK1-12_7_3_9	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMGCAHCA GTTCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMGCAHCAAGTTTCCCTC CTACT	1661	2285
VK1-12_8_3_9	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMGCAAAATN HCTCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMGCAAAATNHCTTCCCTC CTACT	1662	2286
VK1-12_9_3_9	CCTGAAGATTTTGCAACTTATTACTGTGAGSWMGCAAAATA GTWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMGCAAAATAGTYWCCCTC CTACT	1663	2287
VK1-33_1_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMTACGATA ATCTCCCTTWCACITTTTGGCGGAGGACCAAG	SWMSMTACGATAATCTCCCTT WCACT	1664	2288
VK1-33_10_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGBHCNHCA ATCTCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTT WCACT	1665	2289
VK1-33_11_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGBHCGATN HCTCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCGATNHCTTCCCTT WCACT	1666	2290
VK1-33_12_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGBHCGATA ATYWCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCGATAATYWCCCTT WCACT	1667	2291
VK1-33_13_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGTACNHCHN HCTCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACNHCHNCTTCCCTT WCACT	1668	2292
VK1-33_14_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGTACNHCA ATYWCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACNHCAATYWCCCTT WCACT	1669	2293
VK1-33_15_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGCAGTACGATN HCYWCCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACGATNHCHYWCCCTT WCACT	1670	2294
VK1-33_2_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMGCAHCGATA ATCTCCCTTWCACITTTTGGCGGAGGACCAAG	SWCAGBHCAGATAATCTCCCTT WCACT	1671	2295
VK1-33_3_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMGCAGTACNHCA ATCTCCCTTWCACITTTTGGCGGAGGACCAAG	SWCAGTACNHCAATCTCCCTT WCACT	1672	2296
VK1-33_4_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMGATACGATN HCTCCCTTWCACITTTTGGCGGAGGACCAAG	SWCAGTACGATNHCTTCCCTT WCACT	1673	2297
VK1-33_5_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMGATACGATA ATYWCCCTTWCACITTTTGGCGGAGGACCAAG	SWCAGTACGATAATYWCCCTT WCACT	1674	2298
VK1-33_6_0_9	CCTGAAGATTTTGCAACATATTACTGTGAGSWMBHCGATA ATCTCCCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMBHCGATAATCTCCCTT WCACT	1675	2299

VK1-33_7_0_9	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTCACNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMTCACNHCAATCTCCCTT WCACT	1676	2300
VK1-33_8_0_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGSWMTCACGATN HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMTCACGATNHCCCTCCCTT WCACT	1677	2301
VK1-33_9_0_9	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTCACGATA ATYWCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMTCACGATAATYWCCCTT WCACT	1678	2302
VK1-33_1_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTCACGATA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	SMSWMTCAGATAATCTCCCTT TCACT	1679	2303
VK1-33_10_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGCAGBHCNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATCTCCCTT TCACT	1680	2304
VK1-33_11_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCA HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATCTCCCTT TCACT	1681	2305
VK1-33_12_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCA ATYWCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATCTCCCTT TCACT	1682	2306
VK1-33_13_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGCAGTACNHCHN HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCHNCCCTCCCTT TCACT	1683	2307
VK1-33_14_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGTACNHCA ATYWCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCAATYWCCCTT TCACT	1684	2308
VK1-33_15_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGCAGTACGATN HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACGATNHCHNCCCTT TCACT	1685	2309
VK1-33_2_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCNHCAATCTCCCTT TCACT	1686	2310
VK1-33_3_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGCAGTACNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAATCTCCCTT TCACT	1687	2311
VK1-33_4_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGTACGATN HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACGATNHCCCTCCCTT TCACT	1688	2312
VK1-33_5_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGCAGTACGATA ATYWCCCTTTCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACGATAATYWCCCTT TCACT	1689	2313
VK1-33_6_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMBHCNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCNHCAATCTCCCTT TCACT	1690	2314
VK1-33_7_1_9	CCTGAAGAATATTGCAACATATTACTGTCTCAGSWMTCACNHCA ATCTCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMTCACNHCAATCTCCCTT TCACT	1691	2315
VK1-33_8_1_9	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTCACGATN HCCICCCCTTTCACCTTTTGGCGGAGGGACCAAG	CAGSWMTCAGATNHCCCTCCCTT TCACT	1692	2316

VK1-33_9_1_9	CCTGAAGAATATTGCAACATATTACTGTGAGSWMTACGATA ATYWCCCTNICACTTTTGGCGGAGGGACCAAG	CAGSWMTACGATAATYWCCCTW TCACT	1693	2317
VK1-33_1_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGSWMTACGATA ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAATCTCCCTW GGACT	1694	2318
VK1-33_10_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGBHCNHCA ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATCTCCCTW GGACT	1695	2319
VK1-33_11_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGBHCNHCA HCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATCTCCCTW GGACT	1696	2320
VK1-33_12_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGBHCNHCA ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATYWCCCTW GGACT	1697	2321
VK1-33_13_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN HCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCHNCTCCCTW GGACT	1698	2322
VK1-33_14_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCHNCTCCCTW GGACT	1699	2323
VK1-33_15_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN HCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCHNCTCCCTW GGACT	1700	2324
VK1-33_2_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCNHCAATCTCCCTW GGACT	1701	2325
VK1-33_3_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCHNCTCCCTW GGACT	1702	2326
VK1-33_4_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN HCTCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCHNCTCCCTW GGACT	1703	2327
VK1-33_5_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCHNCTCCCTW GGACT	1704	2328
VK1-33_6_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMBHCNHCAATCTCCCTW GGACT	1705	2329
VK1-33_7_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCHNCTCCCTW GGACT	1706	2330
VK1-33_8_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN HCTCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCHNCTCCCTW GGACT	1707	2331
VK1-33_9_2_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCHNCTCCCTW GGACT	1708	2332
VK1-33_1_3_9	CCTGAAGAATATTGCAACATATTACTGTGAGCAGTACNHCHN ATCTCCCTCTACTTTTGGCGGAGGGACCAAG	SWMSWMTACGATAATCTCCCTC CTACT	1709	2333

VK1- 33_10_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1710	2334
VK1- 33_11_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1711	2335
VK1- 33_12_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATYWCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1712	2336
VK1- 33_13_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1713	2337
VK1- 33_14_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATYWCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1714	2338
VK1- 33_15_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCCTC CTACT	1715	2339
VK1-33_2_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATCTCCCTCCTACITTTTGGCGGAGGACCAAG	SWMCAGBHCNHCAATCTCCCTC CTACT	1716	2340
VK1-33_3_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATCTCCCTCCTACITTTTGGCGGAGGACCAAG	SWMCAGBHCNHCAATCTCCCTC CTACT	1717	2341
VK1-33_4_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCTCCCTCCTACITTTTGGCGGAGGACCAAG	SWMCAGBHCNHCAATCTCCCTC CTACT	1718	2342
VK1-33_5_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATYWCCCTCCTACITTTTGGCGGAGGACCAAG	SWMCAGBHCNHCAATCTCCCTC CTACT	1719	2343
VK1-33_6_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGSWMBHCNHCAATCTCCCTC CTACT	1720	2344
VK1-33_7_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGSWMBHCNHCAATCTCCCTC CTACT	1721	2345
VK1-33_8_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCTCCCTCCTACITTTTGGCGGAGGACCAAG	CAGSWMBHCNHCAATCTCCCTC CTACT	1722	2346
VK1-33_9_3_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA ATYWCCCTCCTACITTTTGGCGGAGGACCAAG	CAGSWMBHCNHCAATCTCCCTC CTACT	1723	2347
VK1-39_1_0_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	SWMSWAGCTACAGTACTCCTT WCACT	1724	305
VK1- 39_10_0_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNABHCAGTACTCCTT WCACT	1725	314
VK1- 39_11_0_9	CCTGAAGATAATTGCAACATATTACTGTGTCAGCAGBHCNHCA HCACICCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNABHCAGTACTCCTT WCACT	1726	315

VK1- 39_12_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTBHECCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT WCACT	1727	316
VK1- 39_13_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT WCACT	1728	317
VK1- 39_14_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTBHECCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT WCACT	1729	318
VK1- 39_15_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCBHECCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT WCACT	1730	319
VK1-39_2_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	SWMCAAVNATAACAGTBHCCCTT WCACT	1731	306
VK1-39_3_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	SWMCAAVNATAACAGTBHCCCTT WCACT	1732	307
VK1-39_4_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCACTCCTTWCACITTTTGGCGGAGGACCAAG	SWMCAAVNATAACAGTBHCCCTT WCACT	1733	308
VK1-39_5_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTBHECCCTTWCACITTTTGGCGGAGGACCAAG	SWMCAAVNATAACAGTBHCCCTT WCACT	1734	309
VK1-39_6_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMVNATAACAGTBHCCCTT WCACT	1735	310
VK1-39_7_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTBHCCCTT WCACT	1736	311
VK1-39_8_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTBHCCCTT WCACT	1737	312
VK1-39_9_0_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTBHECCCTTWCACITTTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTBHCCCTT WCACT	1738	313
VK1-39_1_1_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	SWMSWMAGCBHCAGTBHCCCTT TCACCT	1739	2348
VK1- 39_10_1_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNABHCAGTBHCCCTT TCACCT	1740	2349
VK1- 39_11_1_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT TCACCT	1741	2350
VK1- 39_12_1_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA GTBHECCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT TCACCT	1742	2351
VK1- 39_13_1_9	CCTGAAGATTTTGGCACTTACTACTGTCTAGCAAVNATAACA HCACTCCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNATAACAGTBHCCCTT TCACCT	1743	2352

VK1- 39_14_1_9	CCGGAAGATTTGCAACTTACTACTGTGAGCAAAAGCBHCA GTBHCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAAGCBHCAGTBHCCCTM TCACI	1744	2353
VK1- 39_15_1_9	CCGGAAGATTTGCAACTTACTACTGTGAGCAAAAGCTACN HCBHCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAAGCTACNHCBHCCCTM TCACI	1745	2354
VK1-39_2_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAVNATACA GTACTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAAVNATACAGTACTCCTM TCACI	1746	2355
VK1-39_3_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA GTACTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAAAAGCBHCAGTACTCCTM TCACI	1747	2356
VK1-39_4_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCTACN HCACTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAAAAGCTACNHCACI CCTM TCACI	1748	2357
VK1-39_5_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCTACA GTBHCCTMTACATTTTGGCGGAGGACCAAG	SWMCAAAAGCTACAGTBHCCCTM TCACI	1749	2358
VK1-39_6_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAVNATACA GTACTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMVNATACAGTACTCCTM TCACI	1750	2359
VK1-39_7_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA GTACTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTACTCCTM TCACI	1751	2360
VK1-39_8_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA HCACTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMAGCTACNHCACI CCTM TCACI	1752	2361
VK1-39_9_1_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA GTBHCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMAGCTACAGTBHCCCTM TCACI	1753	2362
VK1-39_1_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA GTACTCCTMTACATTTTGGCGGAGGACCAAG	SWMSWMAGCTACAGTACTCCTM GGACT	1754	2363
VK1- 39_10_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAVNABHCA GTACTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAVNABHCAGTACTCCTM GGACT	1755	2364
VK1- 39_11_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAVNATACN HCACTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAVNATACNHCACI CCTM GGACT	1756	2365
VK1- 39_12_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAVNATACA GTBHCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAVNATACAGTBHCCCTM GGACT	1757	2366
VK1- 39_13_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA HCACTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAAGCBHCNHCACI CCTM GGACT	1758	2367
VK1- 39_14_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA GTBHCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAAGCBHCAGTBHCCCTM GGACT	1759	2368
VK1- 39_15_2_9	CCGGAAGATTTGCAACTTACTACTGTGCAAAAGCBHCA HCBHCCTMTACATTTTGGCGGAGGACCAAG	CAGCAAAAGCTACNHCBHCCCTM GGACT	1760	2369

VK1-39_2_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTACTCCTWGGACITTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACICCTW GGACT	1761	2370
VK1-39_3_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTACTCCTWGGACITTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACICCTW GGACT	1762	2371
VK1-39_4_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA HCACTCCTWGGACITTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACICCTW GGACT	1763	2372
VK1-39_5_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTBHCCCTWGGACITTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACICCTW GGACT	1764	2373
VK1-39_6_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTACTCCTWGGACITTTTGGCGAGGGACCAAG	CAGSWMVNATAACAGTACICCTW GGACT	1765	2374
VK1-39_7_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGSWMAGCBHCA GTACTCCTWGGACITTTTGGCGAGGGACCAAG	CAGSWMAGCBHCACTACICCTW GGACT	1766	2375
VK1-39_8_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGSWMAGCBHCA HCACTCCTWGGACITTTTGGCGAGGGACCAAG	CAGSWMAGCBHCACTACICCTW GGACT	1767	2376
VK1-39_9_2_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGSWMAGCBHCA GTBHCCCTWGGACITTTTGGCGAGGGACCAAG	CAGSWMAGCBHCACTACICCTW GGACT	1768	2377
VK1-39_1_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGSWMAGCBHCA GTACTCCTCCTACTTTTGGCGAGGGACCAAG	SWMSWMAGCBHCACTACICCTC CTACT	1769	2378
VK1-39_10_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNABHCA GTACTCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNABHCACTACICCTC CTACT	1770	2379
VK1-39_11_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACN HCACTCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNATAACNHCACCTCCTC CTACT	1771	2380
VK1-39_12_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTBHCCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNATAACAGTBHCCCTC CTACT	1772	2381
VK1-39_13_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNABHCA HCACTCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNABHCACTACICCTC CTACT	1773	2382
VK1-39_14_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNABHCA GTBHCCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNABHCACTACICCTC CTACT	1774	2383
VK1-39_15_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNABHCA HCBHCCCTCCTACTTTTGGCGAGGGACCAAG	CAGCAAVNABHCACTACICCTC CTACT	1775	2384
VK1-39_2_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNATAACA GTACTCCTCCTACTTTTGGCGAGGGACCAAG	SWMCAAVNATAACAGTACICCTC CTACT	1776	2385
VK1-39_3_3_9	CCTGAAGAIITTTGCAACTTACTACTGTGTCAGCAAVNABHCA GTACTCCTCCTACTTTTGGCGAGGGACCAAG	SWMCAAVNABHCACTACICCTC CTACT	1777	2386

VK1-39_4_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWCAAAGCTACN HCACTCCTCTACTTTTGGCGGAGGGACCAAG	SWCAAAGCTACNHCACCTCTC CTACT	1778	2387
VK1-39_5_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWCAAAGCTACA GTBHCCCTCTACTTTTGGCGGAGGGACCAAG	SWCAAAGCTACAGTBHCCCTC CTACT	1779	2388
VK1-39_6_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWVMVNATACA GTACTCCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTCCTC CTACT	1780	2389
VK1-39_7_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWVMAGCBHCA GTACTCCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCACTACICCTC CTACT	1781	2390
VK1-39_8_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWVMAGCTACN HCACTCCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACCTC CTACT	1782	2391
VK1-39_9_3_9	CCTGAAGATTTTGCAACTTTACTACTGTGTSWVMAGCTACA GTBHCCCTCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCCTC CTACT	1783	2392
VK2-28_1_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTSSWMGCACTCC AGACTCCTTWCACCTTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTCCTT WCACT	1784	2393
VK2-28_10_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGVNAMNAC AGACTCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCTT WCACT	1785	2394
VK2-28_11_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGVNACTCS RMACTCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCTT WCACT	1786	2395
VK2-28_12_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGVNACTCC AGVBCCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCTT WCACT	1787	2396
VK2-28_13_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGGCAMNAS RMACTCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNASRMACTCCTT WCACT	1788	2397
VK2-28_14_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGGCAMNAC AGVBCCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCTT WCACT	1789	2398
VK2-28_15_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTGCAGGCACTCS RMVBCCCTTWCACCTTTTGGCGGAGGGACCAAG	ATGCAGGCACCTCSRMVBCCCTT WCACT	1790	2399
VK2-28_2_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTDISCAGVNACTCC AGACTCCTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCTT WCACT	1791	2400
VK2-28_3_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTDISCAGGCAMNAC AGACTCCTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCTT WCACT	1792	2401
VK2-28_4_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTDISCAGGCACCTCS RMACTCCTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCSRMACTCCTT WCACT	1793	2402
VK2-28_5_0_9	GCTGAGGATGTTGGGGTTTATTACTGCTDISCAGGCACCTCC AGVBCCCTTWCACCTTTTGGCGGAGGGACCAAG	DTSCAGGCACCTCCAGVBCCCTT WCACT	1794	2403

VK2-28_6_0_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1795	2404
VK2-28_7_0_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1796	2405
VK2-28_8_0_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC RMACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1797	2406
VK2-28_9_0_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGVBCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1798	2407
VK2-28_1_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1799	2408
VK2-28_10_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1800	2409
VK2-28_11_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC RMACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1801	2410
VK2-28_12_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGVBCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1802	2411
VK2-28_13_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC RMACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1803	2412
VK2-28_14_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGVBCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1804	2413
VK2-28_15_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC RMACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1805	2414
VK2-28_2_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1806	2415
VK2-28_3_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1807	2416
VK2-28_4_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC RMACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1808	2417
VK2-28_5_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGVBCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1809	2418
VK2-28_6_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1810	2419
VK2-28_7_1_9	GCTGAGGAGTGTGGGGTTTTATTACTGCATGSMVMNACTCC AGACTCCTTWCACITTTTGGCGGAGGACCAAG	ATGSMVMNACTCCAGACICCTT WCACT	1811	2420

VK2-28_8_1_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMGCACTCS RMACTCCTMTCACITTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCCTM TCACT	1812	2421
VK2-28_9_1_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMGCACTCC AGVBCCCTMTCACITTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCCTM TCACT	1813	2422
VK2-28_1_2_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACTCC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTCCTW GGACT	1814	2423
VK2-28_10_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGVAMNAC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVAMNACAGACTCCTW GGACT	1815	2424
VK2-28_11_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGVAMNACTCS RMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVAMNACTCSRMACTCCTW GGACT	1816	2425
VK2-28_12_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGVAMNACTCC AGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVAMNACTCCAGVBCCCTW GGACT	1817	2426
VK2-28_13_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGGAMNAS RMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCGAMNASRMACTCCTW GGACT	1818	2427
VK2-28_14_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGGAMNAC AGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCGAMNACAGVBCCCTW GGACT	1819	2428
VK2-28_15_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGCGGACTCS RMVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGGCGACTCSRMBVBCCCTW GGACT	1820	2429
VK2-28_2_2_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVAMNACTCC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGVAMNACTCCAGACTCCTW GGACT	1821	2430
VK2-28_3_2_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCGAMNAC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCGAMNACAGACTCCTW GGACT	1822	2431
VK2-28_4_2_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCS RMACTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCGACTCSRMACTCCTW GGACT	1823	2432
VK2-28_5_2_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACTCC AGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	DTSCAGGCGACTCCAGVBCCCTW GGACT	1824	2433
VK2-28_6_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMVAMNACTCC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMVAMNACTCCAGACTCCTW GGACT	1825	2434
VK2-28_7_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMGCGAMNAC AGACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCGAMNACAGACTCCTW GGACT	1826	2435
VK2-28_8_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMGCACTCS RMACTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCCTW GGACT	1827	2436
VK2-28_9_2_9	GCTGAGGATGTTGGGGTTTATTACTGCGATGSMGCACTCC AGVBCCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCCTW GGACT	1828	2437

VK2-28_1_3_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACTCC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	DTSSWMGCACTCCAGACTCCCTC CTACT	1829	2438
VK2-28_10_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGVNAMNACAGACTCCCTC CTACT	1830	2439
VK2-28_11_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC RMACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGVNACTCSRMACTCCCTC CTACT	1831	2440
VK2-28_12_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC AGVBCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGVNACTCCAGVBCCCTC CTACT	1832	2441
VK2-28_13_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC RMACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGGCAMNARMACTCCCTC CTACT	1833	2442
VK2-28_14_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC AGVBCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGGCAMNACAGVBCCCTC CTACT	1834	2443
VK2-28_15_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAVAMNAC RMACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGCAGGCCTCSRMVBCCCTC CTACT	1835	2444
VK2-28_2_3_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVAMNAC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	DTSCAGVNACTCCAGACTCCCTC CTACT	1836	2445
VK2-28_3_3_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVAMNAC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	DTSCAGGCAMNACAGACTCCCTC CTACT	1837	2446
VK2-28_4_3_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVAMNAC RMACTCCTCCTACITTTTGGCGGAGGGACCAAG	DTSCAGGCCTCSRMACTCCCTC CTACT	1838	2447
VK2-28_5_3_9	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVAMNAC AGVBCCTCCTACITTTTGGCGGAGGGACCAAG	DTSCAGGCCTCCAGVBCCCTC CTACT	1839	2448
VK2-28_6_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVAMNAC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGSWMVNACTCCAGACTCCCTC CTACT	1840	2449
VK2-28_7_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVAMNAC AGACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGSWMGCACTCCAGACTCCCTC CTACT	1841	2450
VK2-28_8_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVAMNAC RMACTCCTCCTACITTTTGGCGGAGGGACCAAG	ATGSWMGCACTCSRMACTCCCTC CTACT	1842	2451
VK2-28_9_3_9	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVAMNAC AGVBCCTCCTACITTTTGGCGGAGGGACCAAG	ATGSWMGCACTCCAGVBCCCTC CTACT	1843	2452
VK3-11_1_0_9	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWAGAGTA ATTGGCCTTWCACITTTTGGCGGAGGGACCAAG	SWMSWAGAGTAATTGGCCCTT WCACIT	1844	2453
VK3-11_10_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCAGAGBHCNHC ATTGGCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATTGGCCCTT WCACIT	1845	2454

VK3- 11_11_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTN HCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCCTT WCACT	1846	2455
VK3- 11_12_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCCTT WCACT	1847	2456
VK3- 11_13_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCN HCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCTGGCCCTT WCACT	1848	2457
VK3- 11_14_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCA ATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTT WCACT	1849	2458
VK3- 11_15_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN HCYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGAGAAAGTNHCYWCCCTT WCACT	1850	2459
VK3-11_2_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTTGGCCCTT WCACT	1851	2460
VK3-11_3_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCA ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTTGGCCCTT WCACT	1852	2461
VK3-11_4_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN HCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTNHCTGGCCCTT WCACT	1853	2462
VK3-11_5_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN ATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGAGAAAGTAATYWCCCTT WCACT	1854	2463
VK3-11_6_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCAGTAATTTGGCCCTT WCACT	1855	2464
VK3-11_7_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGANHCA ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTTGGCCCTT WCACT	1856	2465
VK3-11_8_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAAGTN HCYGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTNHCTGGCCCTT WCACT	1857	2466
VK3-11_9_0_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAAGTN ATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMAGAAAGTAATYWCCCTT WCACT	1858	2467
VK3-11_1_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAAGTN ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMSWMAGAAAGTAATTTGGCCCTT TCACT	1859	2468
VK3- 11_10_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCA ATTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATTTGGCCCTT TCACT	1860	2469
VK3- 11_11_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTN HCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCCTT TCACT	1861	2470
VK3- 11_12_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCCTT TCACT	1862	2471

VK3- 11_13_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCN HCTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCTGGCCTM TCACT	1863	2472
VK3- 11_14_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCA ATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTM TCACT	1864	2473
VK3- 11_15_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN HCYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAGTNHCYWCCCTM TCACT	1865	2474
VK3-11_2_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATTGGCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTM TCACT	1866	2475
VK3-11_3_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCA ATTGGCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGCCTM TCACT	1867	2476
VK3-11_4_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN HCTGGCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTNHCNHCTGGCCTM TCACT	1868	2477
VK3-11_5_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTA ATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTAATYWCCCTM TCACT	1869	2478
VK3-11_6_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSMBHCAGTAATTGGCCTM TCACT	1870	2479
VK3-11_7_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGANHCA ATTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGCCTM TCACT	1871	2480
VK3-11_8_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAAGTN HCTGGCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAGTNHCNHCTGGCCTM TCACT	1872	2481
VK3-11_9_1_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAGTA ATYWCCCTMTCACTTTTGGCGGAGGGACCAAG	CAGSWMAGAAGTAATYWCCCTM TCACT	1873	2482
VK3-11_1_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGSWMAGAAGTA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	SWMSWMAGAAGTAATTGGCCTW GGACT	1874	2483
VK3- 11_10_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAATTGGCCTW GGACT	1875	2484
VK3- 11_11_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTN HCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCNHCTGGCCTW GGACT	1876	2485
VK3- 11_12_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTA ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCCTW GGACT	1877	2486
VK3- 11_13_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCN HCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCTGGCCTW GGACT	1878	2487
VK3- 11_14_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCA ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTW GGACT	1879	2488

VK3- 11_15_2_9	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGAAAGTN HCYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAGTNHCYWCCCTW GGACT	1880	2489
VK3-11_2_2_9	CCTGAAGATTTTGCAGTTTATTACTGTSMCAGBHCAGTA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTW GGACT	1881	2490
VK3-11_3_2_9	CCTGAAGATTTTGCAGTTTATTACTGTSMCAGAGANHCA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGCCTW GGACT	1882	2491
VK3-11_4_2_9	CCTGAAGATTTTGCAGTTTATTACTGTSMCAGAGAAAGTN HCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTNHCTGGCCTW GGACT	1883	2492
VK3-11_5_2_9	CCCTGAAGATTTTGCAGTTTATTACTGTSMCAGAGAAAGTA ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	SWMCAGAGAAGTAATYWCCCTW GGACT	1884	2493
VK3-11_6_2_9	CCCTGAAGATTTTGCAGTTTATTACTGTGAGSMWBHCAGTA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSMWBHCAGTAATTGGCCTW GGACT	1885	2494
VK3-11_7_2_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGSMWAGANHCA ATTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSMWAGANHCAATTGGCCTW GGACT	1886	2495
VK3-11_8_2_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGSMWAGAAAGTN HCTGGCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSMWAGAAGTNECTGGCCTW GGACT	1887	2496
VK3-11_9_2_9	CCCTGAAGATTTTGCAGTTTATTACTGTGAGSMWAGAAAGTA ATYWCCCTWGGACTTTTGGCGGAGGGACCAAG	CAGSMWAGAAGTAATYWCCCTW GGACT	1888	2497
VK3-11_1_3_9	CCTGAAGATTTTGCAGTTTATTACTGTSMWSMWAGAAAGTA ATTGGCCTCCTACTTTTGGCGGAGGGACCAAG	SWMSMWAGAAGTAATTGGCCTC CTACT	1889	2498
VK3- 11_10_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCNEHCA ATTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCNEHCAATTGGCCTC CTACT	1890	2499
VK3- 11_11_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCAGTN HCTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCCTC CTACT	1891	2500
VK3- 11_12_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCAGTA ATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCCTC CTACT	1892	2501
VK3- 11_13_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGAGANHCN HCTGGCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNCTGGCCTC CTACT	1893	2502
VK3- 11_14_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGAGANHCA ATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAATYWCCCTC CTACT	1894	2503
VK3- 11_15_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGAGCAGAGAAAGTN HCYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGCAGAGAAGTNHCYWCCCTC CTACT	1895	2504
VK3-11_2_3_9	CCTGAAGATTTTGCAGTTTATTACTGTSMCAGBHCAGTA ATTGGCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCCTC CTACT	1896	2505

VK3-11_3_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGANHCA ATTGGCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGANHCAATTGGCCCTC CTACT	1897	2506
VK3-11_4_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGAGTN HCTGGCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAGTNHCTGGCCCTC CTACT	1898	2507
VK3-11_5_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGAGTA ATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	SWMCAGAGAGTAATYWCCCTC CTACT	1899	2508
VK3-11_6_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMBHCAGTA ATTGGCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSMBHCAGTAATTGGCCCTC CTACT	1900	2509
VK3-11_7_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGANHCA ATTGGCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGANHCAATTGGCCCTC CTACT	1901	2510
VK3-11_8_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGTN HCTGGCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGAGTNHCTGGCCCTC CTACT	1902	2511
VK3-11_9_3_9	CCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGTA ATYWCCCTCCTACTTTTGGCGGAGGGACCAAG	CAGSWMAGAGTAATYWCCCTC CTACT	1903	2512
VK3-15_1_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGTSWMCAGAGTA ATAATGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMSWMTACAAATAATTGGCCCTT WCACT	1904	2513
VK3-15_10_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCN HCAATGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCATAATTGGCCCTT WCACT	1905	2454
VK3-15_11_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCN ATNHCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCATAATNECTGGCCCTT WCACT	1906	2514
VK3-15_12_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCN ATAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCATAATAATYWCCCTT WCACT	1907	2515
VK3-15_13_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGTACN HCNHCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCNHCATAATTGGCCCTT WCACT	1908	2516
VK3-15_14_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGTACN HCAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCATAATYWCCCTT WCACT	1909	2293
VK3-15_15_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGTACN ATNHCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCATAATYWCCCTT WCACT	1910	2517
VK3-15_2_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGBHCN ATAATGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCATAATAATTGGCCCTT WCACT	1911	2518
VK3-15_3_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGTACN HCAATGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCATAATTGGCCCTT WCACT	1912	2519
VK3-15_4_0_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTGAGCAGTACN ATNHCTGGCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACATAATNECTGGCCCTT WCACT	1913	2520

VK3-15_5_0_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMCAGTACA ATAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACAATAATYWCCCTT WCACT	1914	2521
VK3-15_6_0_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMBHCA ATAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCAATAATIGGCCTT WCACT	1915	2522
VK3-15_7_0_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN HCAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMTACNHCAATIGGCCTT WCACT	1916	2523
VK3-15_8_0_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACA ATNHCTGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMTACAATNHCTGGCCTT WCACT	1917	2524
VK3-15_9_0_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACA ATAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMTACAATAATYWCCCTT WCACT	1918	2525
VK3-15_1_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACA ATAATGGCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMSWMTACAATAATIGGCCTT TCACT	1919	2526
VK3-15_10_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN HCAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNEHCAATIGGCCTT TCACT	1920	2469
VK3-15_11_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATNHCTGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNEHCTGGCCTT TCACT	1921	2527
VK3-15_12_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGBHCNEHCTGGCCTT TCACT	1922	2528
VK3-15_13_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN HCAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCTGGCCTT TCACT	1923	2529
VK3-15_14_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN HCAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACNHCAATAATYWCCCTT TCACT	1924	2308
VK3-15_15_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATNHCTGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGCAGTACAATNHCTGGCCTT TCACT	1925	2530
VK3-15_2_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGBHCNEHCTGGCCTT TCACT	1926	2531
VK3-15_3_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN HCAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACNHCAATAATIGGCCTT TCACT	1927	2532
VK3-15_4_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATNHCTGGCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACAATNHCTGGCCTT TCACT	1928	2533
VK3-15_5_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATAATYWCCCTTWCACCTTTTGGCGGAGGGACCAAG	SWMCAGTACAATAATYWCCCTT TCACT	1929	2534
VK3-15_6_1_9	CAGTCIGAAGATTTTGCAGTTTATTACTGTGTSWMTACN ATAATIGGCCTTWCACCTTTTGGCGGAGGGACCAAG	CAGSWMBHCAATAATIGGCCTT TCACT	1930	2535

VK3-15_7_1_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGSWMTACN HCAATGGCCCTMCACATTTTGGCGAGGGACCAAG	CAGSWMTACNHCATTTGGCCCTM TCACT	1931	2536
VK3-15_8_1_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGSWMTACA ATNHCCTGGCCCTMCACATTTTGGCGAGGGACCAAG	CAGSWMTACNHCATTTGGCCCTM TCACT	1932	2537
VK3-15_9_1_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGSWMTACA ATAATYWCCCTMCACATTTTGGCGAGGGACCAAG	CAGSWMTACNHCATTTGGCCCTM TCACT	1933	2538
VK3-15_1_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGSWMTACA ATAATYWCCCTMCACATTTTGGCGAGGGACCAAG	SWMSWMTACNHCATTTGGCCCTW GGACT	1934	2539
VK3-15_10_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN HCAATGGCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGBCHNHCATTTGGCCCTW GGACT	1935	2484
VK3-15_11_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATNHCCTGGCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGBCHNHCATTTGGCCCTW GGACT	1936	2540
VK3-15_12_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGBCHNHCATTTGGCCCTW GGACT	1937	2541
VK3-15_13_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN HCNHCCTGGCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCATTTGGCCCTW GGACT	1938	2542
VK3-15_14_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN HCAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCATTTGGCCCTW GGACT	1939	2323
VK3-15_15_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATNHCCTGGCCCTWGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCATTTGGCCCTW GGACT	1940	2543
VK3-15_2_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	SWMCAGBCHNHCATTTGGCCCTW GGACT	1941	2544
VK3-15_3_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN HCAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATTTGGCCCTW GGACT	1942	2545
VK3-15_4_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATNHCCTGGCCCTWGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATTTGGCCCTW GGACT	1943	2546
VK3-15_5_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATTTGGCCCTW GGACT	1944	2547
VK3-15_6_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	CAGSWMBCHNHCATTTGGCCCTW GGACT	1945	2548
VK3-15_7_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN HCAATYWCCCTWGGACTTTTGGCGAGGGACCAAG	CAGSWMTACNHCATTTGGCCCTW GGACT	1946	2549
VK3-15_8_2_9	CAGICTGAAGATTTTGCAGTTTATTACTGTGTCAGCBHCHN ATNHCCTGGCCCTWGGACTTTTGGCGAGGGACCAAG	CAGSWMTACNHCATTTGGCCCTW GGACT	1947	2550

VK3-15_9_2_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMTACA ATAATYWCCCTWGGACITTTTGGCGGAGGACCAAG	CAGSWMTACAAATAATYWCCCTC GGACT	1948	2551
VK3-15_1_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGSWMTACA ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	SWMSWMTACAAATAATYGCCCTC CTACT	1949	2552
VK3-15_10_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCN HCAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCNECAATTTGGCCCTC CTACT	1950	2499
VK3-15_11_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCN ATNHCTGGCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCNECAATTTGGCCCTC CTACT	1951	2553
VK3-15_12_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGBHCN ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCNECAATTTGGCCCTC CTACT	1952	2554
VK3-15_13_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN HCNECTGGCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACNEHCNECTGGCCCTC CTACT	1953	2555
VK3-15_14_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN HCAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACNEHCNECTGGCCCTC CTACT	1954	2338
VK3-15_15_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATNHCTGGCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGTACNEHCNECTGGCCCTC CTACT	1955	2556
VK3-15_2_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGBHCNECAATTTGGCCCTC CTACT	1956	2557
VK3-15_3_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN HCAATWCCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACNEHCNECTGGCCCTC CTACT	1957	2558
VK3-15_4_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATNHCTGGCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACNEHCNECTGGCCCTC CTACT	1958	2559
VK3-15_5_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	SWMCAGTACNEHCNECTGGCCCTC CTACT	1959	2560
VK3-15_6_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMBHCNECAATTTGGCCCTC CTACT	1960	2561
VK3-15_7_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN HCAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACNEHCNECTGGCCCTC CTACT	1961	2562
VK3-15_8_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATNHCTGGCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACNEHCNECTGGCCCTC CTACT	1962	2563
VK3-15_9_3_9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACN ATAATWCCCTCCTACTTTTGGCGGAGGACCAAG	CAGSWMTACNEHCNECTGGCCCTC CTACT	1963	2564
VK3-20_1_0_9	CCTGAAGATTTTGCAGTGTATTACTGTSWMSWMTACGGAA GTAGTCTTWCACITTTTGGCGGAGGACCAAG	SWMSWMTACGGAGTAGTCTT WCACT	1964	2565

VK3- 20_10_0_9	CCIGAAGAIITTTGCAGTGTATTACTGTTCAGCAGBHCBCBHA GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBCBHCAGTAGTCCCTT WCACT	1965	2566
VK3- 20_11_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGBHCBCBHA NCAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBCBHCAGTAGTCCCTT WCACT	1966	2567
VK3- 20_12_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGBHCBCBHA GTBHCCTTWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBCBHCAGTAGTCCCTT WCACT	1967	2568
VK3- 20_13_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV NCAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1968	2569
VK3- 20_14_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV GTBHCCTTWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1969	2570
VK3- 20_15_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV NCBHCCTTWCACITTTTGGCGAGGGACCAAG	CAGCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1970	2571
VK3-20_2_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	SWMCAGBHCBCBHCAGTAGTCCCTT WCACT	1971	2572
VK3-20_3_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	SWMCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1972	2573
VK3-20_4_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV NCAGTCCCTIWCACITTTTGGCGAGGGACCAAG	SWMCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1973	2574
VK3-20_5_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV GTBHCCTTWCACITTTTGGCGAGGGACCAAG	SWMCAGTACBHCBCBHCAGTAGTCCCTT WCACT	1974	2575
VK3-20_6_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGCAGTACBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGSWMBCBHCBCBHCAGTAGTCCCTT WCACT	1975	2576
VK3-20_7_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGSWMBCBHCBCBHCAGTAGTCCCTT WCACT	1976	2577
VK3-20_8_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV NCAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGSWMBCBHCBCBHCAGTAGTCCCTT WCACT	1977	2578
VK3-20_9_0_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV GTBHCCTTWCACITTTTGGCGAGGGACCAAG	CAGSWMBCBHCBCBHCAGTAGTCCCTT WCACT	1978	2579
VK3-20_1_1_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	SWMSWMTACBHCBCBHCAGTAGTCCCTT TCACT	1979	2580
VK3- 20_10_1_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV GTAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBCBHCAGTAGTCCCTT TCACT	1980	2581
VK3- 20_11_1_9	CCTGAAGAIITTTGCAGTGTATTACTGTTCAGSWMBCBHCV NCAGTCCCTIWCACITTTTGGCGAGGGACCAAG	CAGCAGBHCBCBHCAGTAGTCCCTT TCACT	1981	2582

VK3- 20_12_1_9	CCIGAAGATTTTGCAGTGTATTACTGTACAGCBHCGGAA GTBHCCTTMCACITTTGGCGGAGGACCAAG	CAGCAGBHCAGGAAAGTBHCCCTM TCACT	1982	2583
VK3- 20_13_1_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACBHC NCAGTCCCTMCACITTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNCAAGTCCCTM TCACT	1983	2584
VK3- 20_14_1_9	CCIGAAGATTTTGCAGTGTATTACTGTACAGTACBHC GTBHCCTTMCACITTTGGCGGAGGACCAAG	CAGCAGTACBHCAGTBHCCCTM TCACT	1984	2585
VK3- 20_15_1_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACGGAV NCBHCCTTMCACITTTGGCGGAGGACCAAG	CAGCAGTACGGAVNCBHCCTM TCACT	1985	2586
VK3-20_20_2_1_9	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGBHC GTAGTCCCTMCACITTTGGCGGAGGACCAAG	SMCAGBHCAGGAGTAGTCCCTM TCACT	1986	2587
VK3-20_20_3_1_9	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACBHC GTAGTCCCTMCACITTTGGCGGAGGACCAAG	SMCAGTACBHCAGTAGTCCCTM TCACT	1987	2588
VK3-20_20_4_1_9	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACGGAV NCAGTCCCTMCACITTTGGCGGAGGACCAAG	SMCAGTACGGAVNCAGTCCCTM TCACT	1988	2589
VK3-20_20_5_1_9	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACGGAA GTBHCCTTMCACITTTGGCGGAGGACCAAG	SMCAGTACGGAVNCAGTBHCCCTM TCACT	1989	2590
VK3-20_20_6_1_9	CCTGAAGATTTTGCAGTGTATTACTGTAGSWMBHC GTAGTCCCTMCACITTTGGCGGAGGACCAAG	CAGSWMBHCAGGAGTAGTCCCTM TCACT	1990	2591
VK3-20_20_7_1_9	CCTGAAGATTTTGCAGTGTATTACTGTAGSWMTACBHC GTAGTCCCTMCACITTTGGCGGAGGACCAAG	CAGSWMTACBHCAGTAGTCCCTM TCACT	1991	2592
VK3-20_20_8_1_9	CCTGAAGATTTTGCAGTGTATTACTGTAGSWMTACGGAV NCAGTCCCTMCACITTTGGCGGAGGACCAAG	CAGSWMTACGGAVNCAGTCCCTM TCACT	1992	2593
VK3-20_20_9_1_9	CCTGAAGATTTTGCAGTGTATTACTGTAGSWMTACGGAA GTBHCCTTMCACITTTGGCGGAGGACCAAG	CAGSWMTACGGAAAGTBHCCCTM TCACT	1993	2594
VK3-20_20_1_2_9	CCTGAAGATTTTGCAGTGTATTACTGTSMWSMTACGGAA GTAGTCCCTWGGACTTTTGGCGGAGGACCAAG	SMWSMTACGGAAAGTAGTCCCTW GGACT	1994	2595
VK3- 20_10_2_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGBHC GTAGTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGTAGTCCCTW GGACT	1995	2596
VK3- 20_11_2_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGBHC NCAGTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGGAVNCAGTCCCTW GGACT	1996	2597
VK3- 20_12_2_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGBHC GTBHCCTTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGGAAAGTBHCCCTW GGACT	1997	2598
VK3- 20_13_2_9	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACBHC NCAGTCCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNCAAGTCCCTW GGACT	1998	2599

VK3- 20_14_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	1999	2600
VK3- 20_15_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2000	2601
VK3-20_2_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2001	2602
VK3-20_3_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2002	2603
VK3-20_4_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2003	2604
VK3-20_5_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2004	2605
VK3-20_6_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2005	2606
VK3-20_7_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2006	2607
VK3-20_8_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2007	2608
VK3-20_9_2_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2008	2609
VK3-20_1_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2009	2610
VK3- 20_10_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2010	2611
VK3- 20_11_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2011	2612
VK3- 20_12_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2012	2613
VK3- 20_13_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2013	2614
VK3- 20_14_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2014	2615
VK3- 20_15_3_9	CCIGAAGATTTTGCAGTGTATTTACTGTCTCAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	CAGCAGTACBHCAGTGTATTTGGCGAGGGACCAAG	2015	2616

VK1-05 t3 2 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGBHCVRMAGTYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTYWCYCTWGG ACT	2034	2635
VK1-05 t3 3 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGBHCVRMAGTYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTYWCYCTWGG ACT	2035	2636
VK1-05 t4 0 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2036	2637
VK1-05 t4 1 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2037	2638
VK1-05 t4 2 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2038	2639
VK1-05 t4 3 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2039	2640
VK1-12 t1 0 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2040	2641
VK1-12 t1 1 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2041	2642
VK1-12 t1 2 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2042	2643
VK1-12 t1 3 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2043	2644
VK1-12 t2 0 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2044	2645
VK1-12 t2 1 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2045	2646
VK1-12 t2 2 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2046	2647
VK1-12 t2 3 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2047	2648
VK1-12 t3 0 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2048	2649
VK1-12 t3 1 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2049	2650
VK1-12 t3 2 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2050	2651
VK1-12 t3 3 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2051	2652
VK1-12 t4 0 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2052	2653
VK1-12 t4 1 9	CCATGATCAATTTGCAACTTATTACTGCCAGCAGTACVRMMBCYWC YCTWGCACATTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTWGG ACT	2053	2654

VK1-12 t4 2 9	CCTGAAGATTTTCCAACTTATTACTGTCTCAGCAGGCANHCNHCYWC CCTWGCACITTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCYWCCTTGGG ACT	2054	2655
VK1-12 t4 3 9	CCTGAAGATTTTGCACACTTATTACTGTCTCAGCAGGCANHCNHCYWC CCTCCTACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCYWCCTTCCCT ACT	2055	2656
VK1-33 t1 0 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2056	2657
VK1-33 t1 1 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2057	2658
VK1-33 t1 2 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2058	2659
VK1-33 t1 3 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2059	2660
VK1-33 t2 0 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2060	2661
VK1-33 t2 1 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2061	2662
VK1-33 t2 2 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2062	2663
VK1-33 t2 3 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2063	2664
VK1-33 t3 0 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2064	2665
VK1-33 t3 1 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2065	2666
VK1-33 t3 2 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2066	2667
VK1-33 t3 3 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2067	2668
VK1-33 t4 0 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2068	2669
VK1-33 t4 1 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2069	2670
VK1-33 t4 2 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2070	2671
VK1-33 t4 3 9	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHCNHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTCCCTTWC ACT	2071	2672
VK1-39 t1 0 9	CCTGAAGATTTTGCACACTTACTACTGTCTCAGCAAVNABHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNABHCNHCCTTCCCTTWC ACT	2072	320
VK1-39 t1 1 9	CCTGAAGATTTTGCACACTTACTACTGTCTCAGCAAVNABHCNHCCTC CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAAVNABHCNHCCTTCCCTTWC ACT	2073	2673

VK1-39 t1 2 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTCCTTGG ACT	2074	2674
VK1-39 t1 3 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTCCTCCT ACT	2075	2675
VK1-39 t2 0 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTWGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTWC ACT	2076	321
VK1-39 t2 1 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTMT ACT	2077	2676
VK1-39 t2 2 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTGG ACT	2078	2677
VK1-39 t2 3 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTCCCT ACT	2079	2678
VK1-39 t3 0 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTWGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTWC ACT	2080	322
VK1-39 t3 1 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTMT ACT	2081	2679
VK1-39 t3 2 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTGG ACT	2082	2680
VK1-39 t3 3 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTCCCT ACT	2083	2681
VK1-39 t4 0 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTWGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTWC ACT	2084	323
VK1-39 t4 1 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTTMTCACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTMT ACT	2085	2682
VK1-39 t4 2 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTWGGACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTGG ACT	2086	2683
VK1-39 t4 3 9	CCTGAAGATTTTGGCAACTTACTACTGTTCAGCAAVNABHCHNCACCT CCTCCCTACTTTTGGCGGAGGGACCAAG	CAGCAAVNABHCHNCACCTBHCCTTCCCT ACT	2087	2684
VK2-28 t1 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTTWGACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTTWC ACT	2088	2685
VK2-28 t1 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTTMT ACT	2089	2686
VK2-28 t1 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTWGGACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTTGG ACT	2090	2687
VK2-28 t1 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTCCCTACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTCCT ACT	2091	2688
VK2-28 t2 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTTWGACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTTWC ACT	2092	2689
VK2-28 t2 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNASRMACT CCTTMTCACTTTTGGCGGAGGGACCAAG	ATGCAGVNAMNASRMACTCCTTMT ACT	2093	2690

VK2-28 t2 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTWGGACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTGG ACT	2094	2691
VK2-28 t2 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTCTACTTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTCT ACT	2095	2692
VK2-28 t3 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2096	2693
VK2-28 t3 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2097	2694
VK2-28 t3 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2098	2695
VK2-28 t3 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2099	2696
VK2-28 t4 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2100	2697
VK2-28 t4 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2101	2698
VK2-28 t4 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2102	2699
VK2-28 t4 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2103	2700
VK3-11 t1 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2104	2701
VK3-11 t1 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2105	2702
VK3-11 t1 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2106	2703
VK3-11 t1 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2107	2704
VK3-11 t2 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2108	2661
VK3-11 t2 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2109	2662
VK3-11 t2 2 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2110	2663
VK3-11 t2 3 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2111	2664
VK3-11 t3 0 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2112	2705
VK3-11 t3 1 9	GCTGAGGATGTTGGGGTTTATTACTGTCATGTCAGVNMNACAGVBC CCTTWCACATTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGVBCCCTTWC ACT	2113	2706

VK3-11 t3 2 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTNHCYWC CTWGGACATTTTGGCGGAGGACCAAG	CAGCAGBHCAGTNHCYWCCTTGGG ACT	2114	2707
VK3-11 t3 3 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCAGTNHCYWC CTCTCTACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGTNHCYWCCTTGGG ACT	2115	2708
VK3-11 t4 0 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGANHCNHCYWC CTTWCACATTTTGGCGGAGGACCAAG	CAGCAGANHCNHCYWCCTTGGG ACT	2116	2709
VK3-11 t4 1 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGANHCNHCYWC CTTWCACATTTTGGCGGAGGACCAAG	CAGCAGANHCNHCYWCCTTGGG ACT	2117	2710
VK3-11 t4 2 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGANHCNHCYWC CTTWCACATTTTGGCGGAGGACCAAG	CAGCAGANHCNHCYWCCTTGGG ACT	2118	2711
VK3-11 t4 3 9	CCGCAAGATTTTGCAGTTTATTACTGTCTCAGCAGANHCNHCYWC CTTWCACATTTTGGCGGAGGACCAAG	CAGCAGANHCNHCYWCCTTGGG ACT	2119	2712
VK3-15 t1 0 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2120	2701
VK3-15 t1 1 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2121	2702
VK3-15 t1 2 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2122	2703
VK3-15 t1 3 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2123	2704
VK3-15 t2 0 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2124	2661
VK3-15 t2 1 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2125	2662
VK3-15 t2 2 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2126	2663
VK3-15 t2 3 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2127	2664
VK3-15 t3 0 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2128	2713
VK3-15 t3 1 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2129	2714
VK3-15 t3 2 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2130	2715
VK3-15 t3 3 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2131	2716
VK3-15 t4 0 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2132	2669
VK3-15 t4 1 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHC TGGCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTTGGG ACT	2133	2670

VK3-15 t4 2 9	CAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTACNHCNHC YWCCTWGGACTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTWGG ACT	2134	2671
VK3-15 t4 3 9	CAGTCTGAACATTTTGCAGTTTATTACTGTCAGCAGTACNHCNHC YWCCTTCTTACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCYWCCTTCTT ACT	2135	2672
VK3-20 t1 0 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2136	2717
VK3-20 t1 1 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2137	2718
VK3-20 t1 2 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2138	2719
VK3-20 t1 3 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2139	2720
VK3-20 t2 0 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2140	2721
VK3-20 t2 1 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2141	2722
VK3-20 t2 2 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2142	2723
VK3-20 t2 3 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2143	2724
VK3-20 t3 0 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2144	2725
VK3-20 t3 1 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2145	2726
VK3-20 t3 2 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2146	2727
VK3-20 t3 3 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2147	2728
VK3-20 t4 0 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNHCBCVNCAGTCCCTTWC ACT	2148	2729
VK3-20 t4 1 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNHCBCVNCAGTCCCTTWC ACT	2149	2730
VK3-20 t4 2 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNHCBCVNCAGTCCCTTWC ACT	2150	2731
VK3-20 t4 3 9	CCCTGAAGATTTTGCAGTGTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGTACBHCVCNHCBCVNCAGTCCCTTWC ACT	2151	2732
VK4-01 t1 0 9	GCTGAAGATGTGGCAGTTTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2152	2733
VK4-01 t1 1 9	GCTGAAGATGTGGCAGTTTATTACTGTACGACGABHCBVCNCAGT CCTTWCACITTTTGGCGGAGGACCAAG	CAGCAGBHCBCVNCAGTCCCTTWC ACT	2153	2734

VK4-01 t1 2 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAC TCCCTWGG ACT	2154	2735
VK4-01 t1 3 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTCCTAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAC TCCCTCCT ACT	2155	2736
VK4-01 t2 0 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2156	2721
VK4-01 t2 1 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2157	2722
VK4-01 t2 2 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2158	2723
VK4-01 t2 3 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2159	2724
VK4-01 t3 0 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2160	2737
VK4-01 t3 1 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2161	2738
VK4-01 t3 2 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2162	2739
VK4-01 t3 3 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2163	2740
VK4-01 t4 0 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2164	2741
VK4-01 t4 1 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2165	2742
VK4-01 t4 2 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2166	2743
VK4-01 t4 3 9	GC TGAAGA TGTGGCAG TTTATTACTGTTCAGCAGBHCBCNHCAC T CCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHCBCNHCAG TBHCCCCTWGC ACT	2167	2744

Table 7. Oligonucleotide sequences for exemplary VK jumping dimer and trimer sequences with CDR3 length 10.

Name	Sequence of Synthesized Oligonucleotide	Portion of Oligonucleotide Corresponding to CDR3 Proper	SEQ ID NO (Oligo)	SEQ ID NO (CDR3 Portion)
Jumping Dimer				
VK1-05_1_0_10	CCTGATGATTTTGCAACTTATTACTGCSWMSWMTACAATAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	SWMSWMTACAATAGTTACYC TCCTTWCAC	2745	3213
VK1-05_10_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGBHCVRMAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCVRMAGTTACYC TCCTTWCAC	2746	3214
VK1-05_11_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGBHCATMB CTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCATMBCTACYC TCCTTWCAC	2747	3215
VK1-05_12_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGBHCATAG TYWCYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGBHCATAGTYWCYC TCCTTWCAC	2748	3216
VK1-05_13_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGTACVRMNB CTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACVRMNBCTACYC TCCTTWCAC	2749	3217
VK1-05_14_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGTACVRMAG TYWCYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACVRMAGTYWCYC TCCTTWCAC	2750	3218
VK1-05_15_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGTACAATNB CYWCYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGCAGTACAATMBCYWCYC TCCTTWCAC	2751	3219
VK1-05_2_0_10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACAATAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACAATAGTTACYC TCCTTWCAC	2752	3220
VK1-05_3_0_10	CCTGATGATTTTGCAACTTATTACTGCSWMCAGTACVRMAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACVRMAGTTACYC TCCTTWCAC	2753	3221
VK1-05_4_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGTACAATNB CTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACAATMBCTACYC TCCTTWCAC	2754	3222
VK1-05_5_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGAGTACAATAG TYWCYCTCCTTWCACITTTGGCGGAGGGACCAAG	SWMCAGTACAATAGTYWCYC TCCTTWCAC	2755	3223
VK1-05_6_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMBHCATAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGSWMBHCATAGTTACYC TCCTTWCAC	2756	3224
VK1-05_7_0_10	CCTGATGATTTTGCAACTTATTACTGCCAGSWMTACVRMAG TTACYCTCCTTWCACITTTGGCGGAGGGACCAAG	CAGSWMTACVRMAGTTACYC TCCTTWCAC	2757	3225

VK1-05_8_0_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATMB CTACYCICCTIWCACITTTGGCGGAGGACCAAG	CAGSWIACAATMBCTACYC TCCITWCAC	2758	3226
VK1-05_9_0_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTA2AATAG TYWCYCTCCTIWCACITTTGGCGGAGGACCAAG	CAGSWIACAATAGTYWCYC TCCITWCAC	2759	3227
VK1-05_1_1_10	CCTGATGATTTTGGCAACTTATTACTGCSWMSWMTACAATAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	SWSWMTACAATAGTITACYC TCCITMCAC	2760	3228
VK1-05_10_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCVRMAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTITACYC TCCITMCAC	2761	3229
VK1-05_11_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATMB CTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGBHCATMBCTACYC TCCITMCAC	2762	3230
VK1-05_12_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGBHCATAG TYWCYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGBHCATAGTYWCYC TCCITMCAC	2763	3231
VK1-05_13_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMMB CTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCTACYC TCCITMCAC	2764	3232
VK1-05_14_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMAG TYWCYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYC TCCITMCAC	2765	3233
VK1-05_15_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACAATMB CYWCYCTCCTMTACITTTGGCGGAGGACCAAG	CAGCAGTACAATMBCYWCYC TCCITMCAC	2766	3234
VK1-05_2_1_10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGBHCATAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	SWMCAGBHCATAGTITACYC TCCITMCAC	2767	3235
VK1-05_3_1_10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACVRMAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTITACYC TCCITMCAC	2768	3236
VK1-05_4_1_10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATMB CTACYCTCCTMTACITTTGGCGGAGGACCAAG	SWMCAGTACAATMBCTACYC TCCITMCAC	2769	3237
VK1-05_5_1_10	CCTGATGATTTTGGCAACTTATTACTGCSWMCAGTACAATAG TYWCYCTCCTMTACITTTGGCGGAGGACCAAG	SWMCAGTACAATAGTYWCYC TCCITMCAC	2770	3238
VK1-05_6_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMBHCATAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGSWMBHCATAGTITACYC TCCITMCAC	2771	3239
VK1-05_7_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACVRMAG TTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTITACYC TCCITMCAC	2772	3240
VK1-05_8_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATMB CTACYCTCCTMTACITTTGGCGGAGGACCAAG	CAGSWMTACAATMBCTACYC TCCITMCAC	2773	3241
VK1-05_9_1_10	CCTGATGATTTTGGCAACTTATTACTGCCAGSWMTACAATAG TYWCYCTCCTMTACITTTGGCGGAGGACCAAG	CAGSWMTACAATAGTYWCYC TCCITMCAC	2774	3242

VK1-05_1_2_10	CCTGATGATTTTGCACCTTATTACTGCSWMSWMTACAATAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMSWMTACAATAGTTACYC TCCTWGGACT	2775	3243
VK1-05_10_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCVRMAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMAGTTACYC TCCTWGGACT	2776	3244
VK1-05_11_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCATMB CTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCATMBCTACYC TCCTWGGACT	2777	3245
VK1-05_12_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGBHCATAG TYWCYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCATAGTYWCYC TCCTWGGACT	2778	3246
VK1-05_13_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACVRMB CTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMBCTACYC TCCTWGGACT	2779	3247
VK1-05_14_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACVRMAG TYWCYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMAGTYWCYC TCCTWGGACT	2780	3248
VK1-05_15_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGCAGTACAATMB CYWCYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACAATMBCYWCYC TCCTWGGACT	2781	3249
VK1-05_2_2_10	CCTGATGATTTTGCACCTTATTACTGCSWMCAGBHCATAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHCATAGTTACYC TCCTWGGACT	2782	3250
VK1-05_3_2_10	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACVRMAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACVRMAGTTACYC TCCTWGGACT	2783	3251
VK1-05_4_2_10	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACAATMB CTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACAATMBCTACYC TCCTWGGACT	2784	3252
VK1-05_5_2_10	CCTGATGATTTTGCACCTTATTACTGCSWMCAGTACAATAG TYWCYCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACAATAGTYWCYC TCCTWGGACT	2785	3253
VK1-05_6_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGSWMBHCATAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHCATAGTTACYC TCCTWGGACT	2786	3254
VK1-05_7_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACVRMAG TTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACVRMAGTTACYC TCCTWGGACT	2787	3255
VK1-05_8_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACAATMB CTACYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACAATMBCTACYC TCCTWGGACT	2788	3256
VK1-05_9_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGSWMTACAATAG TYWCYCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACAATAGTYWCYC TCCTWGGACT	2789	3257
VK1-12_1_0_10	CCTGAAGATTTTGCACCTTATTACTGTSWMSWMCACAATAG TTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMSWMCACAATAGTTCCC TCCTTWCACCT	2790	3258
VK1-12_10_0_10	CCTGAAGATTTTGCACCTTATTACTGTCAGCAGRANHCAG TTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRANHCAGTTTCCC TCCTTWCACCT	2791	3259

VK1-12_11_0_10	CCGGAAGATTTTGGCAACTTATTACTGTCTCAGCAGRNAAATNH CTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRNAAATNHCTTCCC TCCTTWCAC	2792	3260
VK1-12_12_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGRNAAATAG TYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRNAAATAGTYWCCC TCCTTWCAC	2793	3261
VK1-12_13_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGANHCNH CTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCGANHCNHCCTTCCC TCCTTWCAC	2794	3262
VK1-12_14_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGANHCAG TYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCGANHCAGTYWCCC TCCTTWCAC	2795	3263
VK1-12_15_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGAAATNH CYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGGCGAAATNHCYWCCC TCCTTWCAC	2796	3264
VK1-12_2_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGAAATAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGRNAAATAGTITCCC TCCTTWCAC	2797	3265
VK1-12_3_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGANHCAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCGANHCAGTITCCC TCCTTWCAC	2798	3266
VK1-12_4_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGAAATNH CTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCGAAATNHCCTTCCC TCCTTWCAC	2799	3267
VK1-12_5_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGAAATAG TYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMCAGGCGAAATAGTYWCCC TCCTTWCAC	2800	3268
VK1-12_6_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGGCGAAATAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSWMRNAAATAGTITCCC TCCTTWCAC	2801	3269
VK1-12_7_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGSWMGCGANHCAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSWMGCGANHCAGTITCCC TCCTTWCAC	2802	3270
VK1-12_8_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGSWMGCGAAATNH CTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSWMGCGAAATNHCCTTCCC TCCTTWCAC	2803	3271
VK1-12_9_0_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGSWMGCGAAATAG TYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGSWMGCGAAATAGTYWCCC TCCTTWCAC	2804	3272
VK1-12_1_1_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGSWMGCGAAATAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	SWMSWMGCGAAATAGTITCCC TCCTTWCAC	2805	3273
VK1-12_10_1_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGRNAANHCAG TTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRNAANHCAGTITCCC TCCTTWCAC	2806	3274
VK1-12_11_1_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGRNAAATNH CTTCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRNAAATNHCCTTCCC TCCTTWCAC	2807	3275
VK1-12_12_1_10	CCTGAAGATTTTGGCAACTTATTACTGTCTCAGCAGRNAAATAG TYWCCCTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGRNAAATAGTYWCCC TCCTTWCAC	2808	3276

VK1-12_13_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCCTTCCC TCCTMTCACT	2809	3277
VK1-12_14_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCAGTYWCCC TCCTMTCACT	2810	3278
VK1-12_15_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCYWCCC TCCTMTCACT	2811	3279
VK1-12_2_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGRNAAATAGTTTCCC TCCTMTCACT	2812	3280
VK1-12_3_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGGCANHCAGTTTCCC TCCTMTCACT	2813	3281
VK1-12_4_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGGCANHCNHCCTTCCC TCCTMTCACT	2814	3282
VK1-12_5_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGGCANHCAGTYWCCC TCCTMTCACT	2815	3283
VK1-12_6_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMRNAAATAGTTTCCC TCCTMTCACT	2816	3284
VK1-12_7_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMGCANHCAGTTTCCC TCCTMTCACT	2817	3285
VK1-12_8_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMGCANHCNHCCTTCCC TCCTMTCACT	2818	3286
VK1-12_9_1_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMGCANHCAGTYWCCC TCCTMTCACT	2819	3287
VK1-12_1_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMSWGCANHCAGTTTCCC TCCTMTCACT	2820	3288
VK1-12_10_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCAGTTTCCC TCCTMTCACT	2821	3289
VK1-12_11_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCC TCCTMTCACT	2822	3290
VK1-12_12_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCAGTYWCCC TCCTMTCACT	2823	3291
VK1-12_13_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCNH CTTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCNHCCTTCCC TCCTMTCACT	2824	3292
VK1-12_14_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCANHCAG TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGGCANHCAGTYWCCC TCCTMTCACT	2825	3293

VK1-12_15_2_10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGGCAAAATNH CYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGGCAAAATNHCHYWCCC TCCTWGGACT	2826	3294
VK1-12_2_2_10	CCTGAAGATTTTGCAACTTATTACTGTSMCAGRNAATAG TTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGRNAATAGTTTCCC TCCTWGGACT	2827	3295
VK1-12_3_2_10	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCANHCAG TTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGGCANHCAGTTTCCC TCCTWGGACT	2828	3296
VK1-12_4_2_10	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCAAAATNH CTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGGCAAAATNHCTTCCC TCCTWGGACT	2829	3297
VK1-12_5_2_10	CCTGAAGATTTTGCAACTTATTACTGTSMCAGGCAAAATAG TYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGGCAAAATAGTYWCCC TCCTWGGACT	2830	3298
VK1-12_6_2_10	CCTGAAGATTTTGCAACTTATTACTGTAGSWMRNAATAG TTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMRNAATAGTTTCCC TCCTWGGACT	2831	3299
VK1-12_7_2_10	CCTGAAGATTTTGCAACTTATTACTGTAGSWMGCANHCAG TTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMGCANHCAGTTTCCC TCCTWGGACT	2832	3300
VK1-12_8_2_10	CCTGAAGATTTTGCAACTTATTACTGTAGSWMGCAAATNH CTTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMGCAAATNHCTTCCC TCCTWGGACT	2833	3301
VK1-12_9_2_10	CCTGAAGATTTTGCAACTTATTACTGTAGSWMGCAAATAG TYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMGCAAATAGTYWCCC TCCTWGGACT	2834	3302
VK1-33_1_0_10	CCTGAAGATTTTGCAACTTATTACTGTSMWSMTACGATAA TCTCCCTCCTTWCACITTTGGCGGAGGACCAAG	SWMSMTACGATAAICTCCC TCCTTWCACIT	2835	3303
VK1-33_10_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCBHCHNHCAG TCTCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGBCHNHCAGTCTCCC TCCTTWCACIT	2836	3304
VK1-33_11_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCBHCHGATNH CCTCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGBCHGATNHCTCCC TCCTTWCACIT	2837	3305
VK1-33_12_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCBHCHGATAA TYWCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGBCHGATAATYWCCC TCCTTWCACIT	2838	3306
VK1-33_13_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCAGTACNHCNE CCTCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGTACNHCNHCCTCCC TCCTTWCACIT	2839	3307
VK1-33_14_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCAGTACNHCAG TYWCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGTACNHCAGTACNHCAG TCCTTWCACIT	2840	3308
VK1-33_15_0_10	CCTGAAGATTTTGCAACTTATTACTGTAGCAGTACGATNH CYWCCCTCCTTWCACITTTGGCGGAGGACCAAG	CAGCAGTACGATNHCHYWCCC TCCTTWCACIT	2841	3309
VK1-33_2_0_10	CCTGAAGATTTTGCAACTTATTACTGTSMCAGBCHGATAA TCTCCCTCCTTWCACITTTGGCGGAGGACCAAG	SWMCAGBCHGATAAICTCCC TCCTTWCACIT	2842	3310

VK1-33_3_0_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACNHCAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACNHCAAICTCCC TCCCTTWCAC	2843	3311
VK1-33_4_0_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACGATNH CCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACGATNHCCCTCCC TCCCTTWCAC	2844	3312
VK1-33_5_0_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACGATAA TYWCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACGATAATYWCCC TCCCTTWCAC	2845	3313
VK1-33_6_0_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSMBHCGATAATCTCCC TCCCTTWCAC	2846	3314
VK1-33_7_0_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSMBHCGATAATCTCCC TCCCTTWCAC	2847	3315
VK1-33_8_0_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH CCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSMBHCGATNHCCCTCCC TCCCTTWCAC	2848	3316
VK1-33_9_0_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATAA TYWCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSMBHCGATAATYWCCC TCCCTTWCAC	2849	3317
VK1-33_1_1_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACGATAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWSMCAGTACGATAICTCCC TCCCTTWCAC	2850	3318
VK1-33_10_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAAICTCCC TCCCTTWCAC	2851	3319
VK1-33_11_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH CCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAAICTCCC TCCCTTWCAC	2852	3320
VK1-33_12_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATAA TYWCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCAGATAATYWCCC TCCCTTWCAC	2853	3321
VK1-33_13_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH CCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACNHHCNHCCTCCC TCCCTTWCAC	2854	3322
VK1-33_14_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH TYWCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACNHCAAICTCCC TCCCTTWCAC	2855	3323
VK1-33_15_1_10	CCTGAAGATATTGCAACATATTACTGTGAGSMBHCGATNH CYWCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACGATNHHCYWCCC TCCCTTWCAC	2856	3324
VK1-33_2_1_10	CCTGAAGATATTGCAACATATTACTGTSMCAGBHCAGATAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGBHCAGATAAICTCCC TCCCTTWCAC	2857	3325
VK1-33_3_1_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACNECAA TCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACNHCAAICTCCC TCCCTTWCAC	2858	3326
VK1-33_4_1_10	CCTGAAGATATTGCAACATATTACTGTSMCAGTACGATNH CCTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACGATNHCCCTCCC TCCCTTWCAC	2859	3327

VK1-33_5_1_10	CCTGAAGATATTGCAACATATTACTGTGTSWMCAGTACGATAA TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGTACGATAAATYWCCC TCCTMTCACT	2860	3328
VK1-33_6_1_10	CCTGAAGATATTGCAACATATTACTGTGTSWMBHCGATAA TCTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHCGATAATCTCCC TCCTMTCACT	2861	3329
VK1-33_7_1_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACNHCAA TCTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACNHCAATCTCCC TCCTMTCACT	2862	3330
VK1-33_8_1_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATNH CCTCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACGATNHCCCTCCC TCCTMTCACT	2863	3331
VK1-33_9_1_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TYWCCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMTACGATAAATYWCCC TCCTMTCACT	2864	3332
VK1-33_1_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMSWMTACGATAATCTCCC TCCTWGGACT	2865	3333
VK1-33_10_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACNHCAA TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAATCTCCC TCCTWGGACT	2866	3334
VK1-33_11_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATNH CCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGATNHCCCTCCC TCCTWGGACT	2867	3335
VK1-33_12_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGATAAATYWCCC TCCTWGGACT	2868	3336
VK1-33_13_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACNHCA CCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACNHCAATYWCCC TCCTWGGACT	2869	3337
VK1-33_14_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACNHCAA TYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACNHCAATYWCCC TCCTWGGACT	2870	3338
VK1-33_15_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATNH CYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACGATNHCAATYWCCC TCCTWGGACT	2871	3339
VK1-33_2_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGBHCAGATAATCTCCC TCCTWGGACT	2872	3340
VK1-33_3_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACNHCAA TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACNHCAATCTCCC TCCTWGGACT	2873	3341
VK1-33_4_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATNH CCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACGATNHCCCTCCC TCCTWGGACT	2874	3342
VK1-33_5_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TYWCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACGATAAATYWCCC TCCTWGGACT	2875	3343
VK1-33_6_2_10	CCTGAAGATATTGCAACATATTACTGTGTSWMTACGATAA TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHCGATAATCTCCC TCCTWGGACT	2876	3344

VK1-33_7_2_10	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2877	3345
VK1-33_8_2_10	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2878	3346
VK1-33_9_2_10	CCTGAAGATATTGCAACATATTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2879	3347
VK1-39_1_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2880	3348
VK1-39_10_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2881	3349
VK1-39_11_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2882	3350
VK1-39_12_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2883	3351
VK1-39_13_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2884	3352
VK1-39_14_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2885	3353
VK1-39_15_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2886	3354
VK1-39_2_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2887	3355
VK1-39_3_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2888	3356
VK1-39_4_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2889	3357
VK1-39_5_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2890	3358
VK1-39_6_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2891	3359
VK1-39_7_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2892	3360
VK1-39_8_0_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMTACNHC TCTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACNHC TCTTWWGACT	2893	3361

VK1-39_9_0_10	CCIGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCTACAG TBHCCCTCCTTWCACATTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCC TCCTTWCAC	2894	3362
VK1-39_1_1_10	CCIGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCTACAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTCC TCCTTMTCACT	2895	3363
VK1-39_10_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAVNABHCAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAVNABHCAGTACTCC TCCTTMTCACT	2896	3364
VK1-39_11_1_10	CCIGAAGATTTTGGCAACTTACTACTGTCTCAGCAAVNATACNH CACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAVNATACNHCACTCC TCCTTMTCACT	2897	3365
VK1-39_12_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAVNATACAG TBHCCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAVNATACAGTBHCCC TCCTTMTCACT	2898	3366
VK1-39_13_1_10	CCIGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCBHCNH CACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAGCBHCNHCACTCC TCCTTMTCACT	2899	3367
VK1-39_14_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCBHCAG TBHCCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAGCBHCAGTBHCCC TCCTTMTCACT	2900	3368
VK1-39_15_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCTACNH CBHCCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGCAAGCTACNHCBHCCC TCCTTMTCACT	2901	3369
VK1-39_2_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAVNATACAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	SWMCAAVNATACAGTACTCC TCCTTMTCACT	2902	3370
VK1-39_3_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCBHCAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	SWMCAAGCBHCAGTACTCC TCCTTMTCACT	2903	3371
VK1-39_4_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCTACNH CACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	SWMCAAGCTACNHCACTCC TCCTTMTCACT	2904	3372
VK1-39_5_1_10	CCIGAAGATTTTGGCAACTTACTACTGTCTCAGCAAGCTACAG TBHCCCTCCTMTACATTTTGGCGGAGGGACCAAG	SWMCAAGCTACAGTBHCCC TCCTTMTCACT	2905	3373
VK1-39_6_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGSWMVNATACAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGSWMVNATACAGTACTCC TCCTTMTCACT	2906	3374
VK1-39_7_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCBHCAG TACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGSWMAGCBHCAGTACTCC TCCTTMTCACT	2907	3375
VK1-39_8_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCTACNH CACTCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGSWMAGCTACNHCACTCC TCCTTMTCACT	2908	3376
VK1-39_9_1_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCTACAG TBHCCCTCCTMTACATTTTGGCGGAGGGACCAAG	CAGSWMAGCTACAGTBHCCC TCCTTMTCACT	2909	3377
VK1-39_1_2_10	CCTGAAGATTTTGGCAACTTACTACTGTCTCAGSWMAGCTACAG TACTCCTCCTMTGAGACTTTTGGCGGAGGGACCAAG	SWMSWMAGCTACAGTACTCC TCCTTGGGACT	2910	3378

VK1-39_10_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAVNABHCAG TACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAVNABHCAGTACTCC TCCTWGGACT	2911	3379
VK1-39_11_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAVNATACNH CACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAVNATACNHCACTCC TCCTWGGACT	2912	3380
VK1-39_12_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAVNATACAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAVNATACAGTBHCCC TCCTWGGACT	2913	3381
VK1-39_13_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCBHCNH CACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAGCBHCNHCACTCC TCCTWGGACT	2914	3382
VK1-39_14_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCBHCAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAGCBHCAGTBHCCC TCCTWGGACT	2915	3383
VK1-39_15_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCTACNH CBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAAGCTACNHCBHCCC TCCTWGGACT	2916	3384
VK1-39_2_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAVNATACAG TACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAAVNATACAGTACTCC TCCTWGGACT	2917	3385
VK1-39_3_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCBHCAG TACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAAGCBHCAGTACTCC TCCTWGGACT	2918	3386
VK1-39_4_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCTACNH CACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAAGCTACNHCACTCC TCCTWGGACT	2919	3387
VK1-39_5_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGCAAGCTACAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SWMCAAGCTACAGTBHCCC TCCTWGGACT	2920	3388
VK1-39_6_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMVNATACAG TACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMVNATACAGTACTCC TCCTWGGACT	2921	3389
VK1-39_7_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMAGCBHCAG TACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMAGCBHCAGTACTCC TCCTWGGACT	2922	3390
VK1-39_8_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMAGCTACNH CACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMAGCTACNHCACTCC TCCTWGGACT	2923	3391
VK1-39_9_2_10	CCTGAAGATTTTGCAACTTACTACTGTCTCAGSWMAGCTACAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMAGCTACAGTBHCCC TCCTWGGACT	2924	3392
VK2-28_1_0_10	GCTGAGGATGTTGGGTTTATTACTGCDTSSWMGCACCTCA GACTCCTCCTWCACTTTTGGCGGAGGACCAAG	DTSSWMGCACCTCAGACTCC TCCTTWCACT	2925	3393
VK2-28_10_0_10	GCTGAGGATGTTGGGTTTATTACTGCGATGCAGVNMNACA GACTCCTCCTWCACTTTTGGCGGAGGACCAAG	ATGCAGVNMNACAGACTCC TCCTTWCACT	2926	3394
VK2-28_11_0_10	GCTGAGGATGTTGGGTTTATTACTGCGATGCAGVNACTCSR MACTCCTCCTWCACTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMACTCC TCCTTWCACT	2927	3395

VK2-28_12_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNACTCCA GVBCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGCAGVNACTCCAGVBCCC TCCTTWCAC	2928	3396
VK2-28_13_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCAMNASR MACTCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGCAGGCAMNASRVACICC TCCTTWCAC	2929	3397
VK2-28_14_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCAMNACA GVBCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGCAGGCAMNACAGVBCCC TCCTTWCAC	2930	3398
VK2-28_15_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCACCTCSR MVBCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGCAGGCACCTCSRMVBCCC TCCTTWCAC	2931	3399
VK2-28_2_0_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCA GACTCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	DTSCAGVNACTCCAGACICC TCCTTWCAC	2932	3400
VK2-28_3_0_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACA GACTCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	DTSCAGGCAMNACAGACICC TCCTTWCAC	2933	3401
VK2-28_4_0_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCSR MACTCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	DTSCAGGCACCTCSRMACICC TCCTTWCAC	2934	3402
VK2-28_5_0_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACCTCCA GVBCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	DTSCAGGCACCTCCAGVBCCC TCCTTWCAC	2935	3403
VK2-28_6_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMVNACTCCA GACTCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGSWMVNACTCCAGACTCC TCCTTWCAC	2936	3404
VK2-28_7_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCAMNACA GACTCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGSWMGCAMNACAGACTCC TCCTTWCAC	2937	3405
VK2-28_8_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCACCTCSR MACTCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGSWMGCACCTCSRMACICC TCCTTWCAC	2938	3406
VK2-28_9_0_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGSWMGCACCTCCA GVBCCCTCCTTWCACATTTTGGCGGAGG3ACCAAG	ATGSWMGCACCTCCAGVBCCC TCCTTWCAC	2939	3407
VK2-28_1_1_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACCTCCA GACTCCTCCTMTACATTTTGGCGGAGG3ACCAAG	DTSSWMGCACCTCCAGACTCC TCCTMTAC	2940	3408
VK2-28_10_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNAAMNACA GACTCCTCCTMTACATTTTGGCGGAGG3ACCAAG	ATGCAGVNAAMNACAGACTCC TCCTMTAC	2941	3409
VK2-28_11_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNACTCSR MACTCCTCCTMTACATTTTGGCGGAGG3ACCAAG	ATGCAGVNACTCSRMACICC TCCTMTAC	2942	3410
VK2-28_12_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGVNACTCCA GVBCCCTCCTMTACATTTTGGCGGAGG3ACCAAG	ATGCAGVNACTCCAGVBCCC TCCTMTAC	2943	3411
VK2-28_13_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAIGCAGGCAMNASR MACTCCTCCTMTACATTTTGGCGGAGG3ACCAAG	ATGCAGGCAMNASRMACICC TCCTMTAC	2944	3412

VK2-28_14_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCAMNACA GVBCCTCCTMTCACITTTGGCGGAGGACCAAG	ATGCAGGCAMNACAGVBCCC TCCTMTCACIT	2945	3413
VK2-28_15_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCACITCSR MVBCCCTCCTMTCACITTTGGCGGAGGACCAAG	ATGCAGGCACITCSRMVBCCC TCCTMTCACIT	2946	3414
VK2-28_2_1_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGVNACTCCA GACTCCTCCTMTCACITTTGGCGGAGGACCAAG	DTSCAGVNACTCCAGACTCC TCCTMTCACIT	2947	3415
VK2-28_3_1_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCAMNACA GACTCCTCCTMTCACITTTGGCGGAGGACCAAG	DTSCAGGCAMNACAGACTCC TCCTMTCACIT	2948	3416
VK2-28_4_1_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACITCSR MACTCCTCCTMTCACITTTGGCGGAGGACCAAG	DTSCAGGCACITCSRMVACTCC TCCTMTCACIT	2949	3417
VK2-28_5_1_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSCAGGCACITCCA GVBCCTCCTMTCACITTTGGCGGAGGACCAAG	DTSCAGGCACITCCAGVBCCC TCCTMTCACIT	2950	3418
VK2-28_6_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMVNACTCCA GACICCTCCTMTCACITTTGGCGGAGGACCAAG	ATGSMVNACTCCAGACTCC TCCTMTCACIT	2951	3419
VK2-28_7_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMGCAMNACA GACTCCTCCTMTCACITTTGGCGGAGGACCAAG	ATGSMGCAMNACAGACTCC TCCTMTCACIT	2952	3420
VK2-28_8_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMGCACITCSR MACTCCTCCTMTCACITTTGGCGGAGGACCAAG	ATGSMGCACITCSRMVACTCC TCCTMTCACIT	2953	3421
VK2-28_9_1_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGSMGCACITCCA GVBCCTCCTMTCACITTTGGCGGAGGACCAAG	ATGSMGCACITCCAGVBCCC TCCTMTCACIT	2954	3422
VK2-28_1_2_10	GCTGAGGATGTTGGGGTTTATTACTGCDTSSWMGCACITCCA GACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	DTSSWMGCACITCCAGACTCC TCCTWGGACT	2955	3423
VK2-28_10_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNAMNACA GACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNAMNACAGACTCC TCCTWGGACT	2956	3424
VK2-28_11_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNACTCSR MACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMVACTCC TCCTWGGACT	2957	3425
VK2-28_12_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGVNACTCCA GVBCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGVNACTCCAGVBCCC TCCTWGGACT	2958	3426
VK2-28_13_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCAMNASR MACTCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGGCAMNASRMVACTCC TCCTWGGACT	2959	3427
VK2-28_14_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCAMNACA GVBCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGGCAMNACAGVBCCC TCCTWGGACT	2960	3428
VK2-28_15_2_10	GCTGAGGATGTTGGGGTTTATTACTGCAATGCAGGCACITCSR MVBCCCTCCTWGGACTTTTGGCGGAGGACCAAG	ATGCAGGCACITCSRMVBCCC TCCTWGGACT	2961	3429

VK2-28_2_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTDTSAGVNACTCCA GACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSAGVNACTCCAGACTCC TCCTWGGACT	2962	3430
VK2-28_3_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTDTSAGGGMNACA GACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSAGGGMNACAGACTCC TCCTWGGACT	2963	3431
VK2-28_4_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTDTSAGGCACTCSR MACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSAGGCACTCSRMACTCC TCCTWGGACT	2964	3432
VK2-28_5_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTDTSAGGCACTCCA GVBCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	DTSAGGCACTCCAGVBCCC TCCTWGGACT	2965	3433
VK2-28_6_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTATGSMVNACTCCA GACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMVNACTCCAGACTCC TCCTWGGACT	2966	3434
VK2-28_7_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTATGSMGCACTCSR GACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCC TCCTWGGACT	2967	3435
VK2-28_8_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTATGSMGCACTCSR MACTCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCSRMACTCC TCCTWGGACT	2968	3436
VK2-28_9_2_10	GCTGAGGATGTTGGGGTTTATTACTGCTATGSMGCACTCCA GVBCCCTCCTWGGACTTTTGGCGGAGGGACCAAG	ATGSMGCACTCCAGVBCCC TCCTWGGACT	2969	3437
VK3-11_1_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTSMSWMAAGATAA TTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	SMSWMAAGATAATTGGCC TCCTTWCACIT	2970	3438
VK3-11_10_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCBHCNHCAA TTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCNHCAAATTGGCC TCCTTWCACIT	2971	3439
VK3-11_11_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCBHCAGTNH CTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTNHCTGGCC TCCTTWCACIT	2972	3440
VK3-11_12_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCBHCAGTAA TYWCCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGBHCAGTAATYWCCC TCCTTWCACIT	2973	3441
VK3-11_13_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCAGAGANHCNH CTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGAGANHCNHCCTGGCC TCCTTWCACIT	2974	3442
VK3-11_14_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCAGAGANHCAA TYWCCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGAGANHCAAATYWCCC TCCTTWCACIT	2975	3443
VK3-11_15_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTCAGCAGAGAGTNH CYWCCCTCCTTWCACITTTTGGCGGAGGGACCAAG	CAGCAGAGAGAGTNHCYWCCC TCCTTWCACIT	2976	3444
VK3-11_2_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTSWMCAGBHCAGTAA TTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	SWMCAGBHCAGTAATTGGCC TCCTTWCACIT	2977	3445
VK3-11_3_0_10	CCTGAAGATTTTGCAGTTTTATTACTGTSWMCAGAGNHCAA TTGGCCTCCTTWCACITTTTGGCGGAGGGACCAAG	SWMCAGAGNHCAAATTGGCC TCCTTWCACIT	2978	3446

VK3-11_4_0_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTNH CTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGAGAAGTNHCTGGCC TCCTTWCACI	2979	3447
VK3-11_5_0_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTAA TYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGAGAAGTAAATYWCCC TCCTTWCACI	2980	3448
VK3-11_6_0_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMBHCAGTAA TISGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGSWMBHCAGTAAITGGCC TCCTTWCACI	2981	3449
VK3-11_7_0_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWMAGANECAA TISGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGSWMAGANECAATITGGCC TCCTTWCACI	2982	3450
VK3-11_8_0_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWWAGAAGTNE CTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGSWWAGAAGTNHCTGGCC TCCTTWCACI	2983	3451
VK3-11_9_0_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGSWWAGAAGTAA TYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGSWWAGAAGTAAATYWCCC TCCTTWCACI	2984	3452
VK3-11_1_1_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMSWAGAAGTAA TTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMSWAGAAGTAAITGGCC TCCTTWCACI	2985	3453
VK3-11_10_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCBHCNHCAA TTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCBHCNHCAATITGGCC TCCTTWCACI	2986	3454
VK3-11_11_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCBHCAGTNH CTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCBHCAGTNHCTGGCC TCCTTWCACI	2987	3455
VK3-11_12_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCBHCAGTAA TYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCBHCAGTAAATYWCCC TCCTTWCACI	2988	3456
VK3-11_13_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCNH CTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCAGAGANHCNHCTGGCC TCCTTWCACI	2989	3457
VK3-11_14_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGANHCAA TYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCAGAGANHCAAATYWCCC TCCTTWCACI	2990	3458
VK3-11_15_1_10	CCTGAAGATTTTGCAGTTTATTACTGTCAGCAGAGAAGTNH CYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	CAGCAGAGAAGTNHCYWCCC TCCTTWCACI	2991	3459
VK3-11_2_1_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGBHCAGTAA TTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGBHCAGTAAITGGCC TCCTTWCACI	2992	3460
VK3-11_3_1_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGANHCAA TTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGAGANHCAAATITGGCC TCCTTWCACI	2993	3461
VK3-11_4_1_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTNH CTGGCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGAGAAGTNHCTGGCC TCCTTWCACI	2994	3462
VK3-11_5_1_10	CCTGAAGATTTTGCAGTTTATTACTGTSWMCAGAGAAGTAA TYWCCCTCCTTWCACITTTTGGCGGAGG3ACCAAG	SWMCAGAGAAGTAAATYWCCC TCCTTWCACI	2995	3463

VK3-11_6_1_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMBHCAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHCAGTAAATTGGCC TCCIMTCACT	2996	3464
VK3-11_7_1_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMAGANHCAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMAGANHCAAATTGGCC TCCIMTCACT	2997	3465
VK3-11_8_1_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMAGAAAGTINH TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMAGAAAGTINHCTGGCC TCCIMTCACT	2998	3466
VK3-11_9_1_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMAGAAAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMAGAAAGTAAATYWCCC TCCIMTCACT	2999	3467
VK3-11_1_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMAGAAAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMSWMAGAAAGTAAATTGGCC TCCIWGGACT	3000	3468
VK3-11_10_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGBHCNHCAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCAAATTGGCC TCCIWGGACT	3001	3469
VK3-11_11_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGBHCAGTINH TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGTINHCTGGCC TCCIWGGACT	3002	3470
VK3-11_12_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGBHCAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGTAAATYWCCC TCCIWGGACT	3003	3471
VK3-11_13_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGANHCNH TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGAGANHCNHCNCTGGCC TCCIWGGACT	3004	3472
VK3-11_14_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGAGANHCAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGAGANHCNHCNCTGGCC TCCIWGGACT	3005	3473
VK3-11_15_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGAGAAAGTINH TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGCAGAGAAAGTINHCTYWCCC TCCIWGGACT	3006	3474
VK3-11_2_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGBHCAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGBHCAGTAAATTGGCC TCCIWGGACT	3007	3475
VK3-11_3_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGAGANHCAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGAGANHCNHCNCTGGCC TCCIWGGACT	3008	3476
VK3-11_4_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGAGAAAGTINH TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGAGAAAGTINHCTGGCC TCCIWGGACT	3009	3477
VK3-11_5_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGCAGAGAAAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	SWMCAGAGAAAGTAAATYWCCC TCCIWGGACT	3010	3478
VK3-11_6_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMBHCAGTAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMBHCAGTAAATTGGCC TCCIWGGACT	3011	3479
VK3-11_7_2_10	CCTGAAGATTTTGCAAGTTTATTACTGTCTCAGSWMAGANHCAA TTGGCCTCCTMTCACTTTTGGCGGAGGACCAAG	CAGSWMAGANHCNHCNCTGGCC TCCIWGGACT	3012	3480

VK3-15_1_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMSWMTACAAATAATTGGCC TCCCTMTCACT	3030	3496
VK3-15_10_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCATTAATTGGCC TCCCTMTCACT	3031	3454
VK3-15_11_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCATTAATTGGCC TCCCTMTCACT	3032	3497
VK3-15_12_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCATTAATTGGCC TCCCTMTCACT	3033	3498
VK3-15_13_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3034	3499
VK3-15_14_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3035	3323
VK3-15_15_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3036	3500
VK3-15_2_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAGBHCNHCATTAATTGGCC TCCCTMTCACT	3037	3501
VK3-15_3_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3038	3502
VK3-15_4_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3039	3503
VK3-15_5_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMCAGTACNHCNHCATTAATTGGCC TCCCTMTCACT	3040	3504
VK3-15_6_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMBHCNHCATTAATTGGCC TCCCTMTCACT	3041	3505
VK3-15_7_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMTACNHCNHCATTAATTGGCC TCCCTMTCACT	3042	3506
VK3-15_8_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMTACNHCNHCATTAATTGGCC TCCCTMTCACT	3043	3507
VK3-15_9_1_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSWMTACNHCNHCATTAATTGGCC TCCCTMTCACT	3044	3508
VK3-15_1_2_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	SWMSWMTACNHCNHCATTAATTGGCC TCCCTMTCACT	3045	3509
VK3-15_10_2_10	CAGTCIGAAGATTTTIGCAGTIIATTTACIGTISWMSWMTACAA TAAITGGCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGBHCNHCATTAATTGGCC TCCCTMTCACT	3046	3469

VK3-15_11_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGBHC TNCCTGGCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGCAGBHCATNHCCTGGCC TCCTWGGACT	3047	3510
VK3-15_12_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGBHC TAATYWCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGCAGBHCATNHCCTGGCC TCCTWGGACT	3048	3511
VK3-15_13_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE CNCCTGGCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCNHCCTGGCC TCCTWGGACT	3049	3512
VK3-15_14_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE CAATYWCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCATNHCCTGGCC TCCTWGGACT	3050	3513
VK3-15_15_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TNCCTGGCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGCAGTACNHCATNHCCTGGCC TCCTWGGACT	3051	3514
VK3-15_2_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGBHC TAATYGCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	SWMCAGBHCATNHCCTGGCC TCCTWGGACT	3052	3515
VK3-15_3_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE CAATYGCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATNHCCTGGCC TCCTWGGACT	3053	3516
VK3-15_4_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TNCCTGGCTCCTTGGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATNHCCTGGCC TCCTWGGACT	3054	3517
VK3-15_5_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TAATYWCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	SWMCAGTACNHCATNHCCTGGCC TCCTWGGACT	3055	3518
VK3-15_6_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TAATYGCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGSWMBHCATNHCCTGGCC TCCTWGGACT	3056	3519
VK3-15_7_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE CAATYGCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGSWMTACNHCATNHCCTGGCC TCCTWGGACT	3057	3520
VK3-15_8_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TNCCTGGCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGSWMTACNHCATNHCCTGGCC TCCTWGGACT	3058	3521
VK3-15_9_2_10	CAGTCGAAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TAATYWCCCTCCTTGGGACTTTTGGCGAGGGACCAAG	CAGSWMTACNHCATNHCCTGGCC TCCTWGGACT	3059	3522
VK3-20_1_0_10	CCTGAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TAGTCTCCTTWCCTTTTGGCGAGGGACCAAG	SWMSWMTACGGAAGTAGTCTCCTTWCCTTTTGGCGAGGGACCAAG	3060	3523
VK3-20_10_0_10	CCTGAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TAGTCTCCTTWCCTTTTGGCGAGGGACCAAG	CAGCAGBHCATNHCCTGGCC TCCTWGGACT	3061	3524
VK3-20_11_0_10	CCTGAAGATTTTCAGTTTATTACTGTCAGCAGTACNE CAGTCTCCTTWCCTTTTGGCGAGGGACCAAG	CAGCAGBHCATNHCCTGGCC TCCTWGGACT	3062	3525
VK3-20_12_0_10	CCTGAAGATTTTCAGTTTATTACTGTCAGCAGTACNE TBHCCCTCCTTWCCTTTTGGCGAGGGACCAAG	CAGCAGBHCATNHCCTGGCC TCCTWGGACT	3063	

VK3-20_13_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGTACBHCVN CAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACBHCVNCAGTCC TCCTTWCAC	3064	3526
VK3-20_14_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGTACBHCAG TBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACBHCAGTBHCCC TCCTTWCAC	3065	3527
VK3-20_15_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGTACGGAVN CBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACGGAVNCBHCCC TCCTTWCAC	3066	3528
VK3-20_2_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTSWMCAGBHCAGGAA TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGBHCAGGAAAGTAGTCC TCCTTWCAC	3067	3529
VK3-20_3_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTSWMCAGTACBHCAG TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACBHCAGTAGTCC TCCTTWCAC	3068	3530
VK3-20_4_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTSWMCAGTACGGAVN CAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACGGAVNCAGTCC TCCTTWCAC	3069	3531
VK3-20_5_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTSWMCAGTACGGAA TBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMCAGTACGGAAAGTBHCCC TCCTTWCAC	3070	3532
VK3-20_6_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGSWMBHCAGGAA TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSWMBHCAGGAAAGTAGTCC TCCTTWCAC	3071	3533
VK3-20_7_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGSWMTACBHCAG TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSWMTACBHCAGTAGTCC TCCTTWCAC	3072	3534
VK3-20_8_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGSWMTACGGAVN CAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSWMTACGGAVNCAGTCC TCCTTWCAC	3073	3535
VK3-20_9_0_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGSWMTACGGAA TBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGSWMTACGGAAAGTBHCCC TCCTTWCAC	3074	3536
VK3-20_1_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTSWMSWMTACGGAA TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	SWMSWMTACGGAAAGTAGTCC TCCTTWCAC	3075	3537
VK3-20_10_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGBHCAG TAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCAGTAGTCC TCCTTWCAC	3076	3538
VK3-20_11_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGBHCAGGAVN CAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCAGGAVNCAGTCC TCCTTWCAC	3077	3539
VK3-20_12_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGBHCAGGAA TBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGBHCAGGAAAGTBHCCC TCCTTWCAC	3078	3540
VK3-20_13_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGTACBHCVN CAGTCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACBHCVNCAGTCC TCCTTWCAC	3079	3541
VK3-20_14_1_10	CCTGAAGATTTTGCAGTGATTAATCTGTCAGCAGTACBHCAG TBHCCCTCCTTWCACATTTTGGCGGAGGACCAAG	CAGCAGTACBHCAGTBHCCC TCCTTWCAC	3080	3542

VK3-20_15_1_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGCAGTACGGAVN CBHCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGCAGTACGGAVNCBHCCC TCCTMTCACT	3081	3543
VK3-20_2_1_10	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGBHCAGGAAAG TAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	SMCAGBHCAGGAAAGTAGTCC TCCTMTCACT	3082	3544
VK3-20_3_1_10	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACBHCAG TAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	SMCAGTACBHCAGTAGTCC TCCTMTCACT	3083	3545
VK3-20_4_1_10	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACGGAVN CAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	SMCAGTACGGAVNCAGTCC TCCTMTCACT	3084	3546
VK3-20_5_1_10	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGTACGGAAAG TBHCCCTCCTMTACATTTTGGCGGAGGACCAAG	SMCAGTACGGAAAGTBHCCC TCCTMTCACT	3085	3547
VK3-20_6_1_10	CCTGAAGATTTTGCAGTGTATTACTGTACGSMBHCAGGAAAG TAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSMBHCAGGAAAGTAGTCC TCCTMTCACT	3086	3548
VK3-20_7_1_10	CCTGAAGATTTTGCAGTGTATTACTGTACGSMWTACBHCAG TAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSMWTACBHCAGTAGTCC TCCTMTCACT	3087	3549
VK3-20_8_1_10	CCTGAAGATTTTGCAGTGTATTACTGTACGSMWTACGGAVN CAGTCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSMWTACGGAVNCAGTCC TCCTMTCACT	3088	3550
VK3-20_9_1_10	CCTGAAGATTTTGCAGTGTATTACTGTACGSMWTACGGAAAG TBHCCCTCCTMTACATTTTGGCGGAGGACCAAG	CAGSMWTACGGAAAGTBHCCC TCCTMTCACT	3089	3551
VK3-20_1_2_10	CCTGAAGATTTTGCAGTGTATTACTGTSMMSMWTACGGAAAG TAGTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SMMSMWTACGGAAAGTAGTCC TCCTWGGACT	3090	3552
VK3-20_10_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGCBHCAG TAGTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGTAGTCC TCCTWGGACT	3091	3553
VK3-20_11_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGCBHCAGGAVN CAGTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGAVNCAGTCC TCCTWGGACT	3092	3554
VK3-20_12_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGCBHCAGGAAAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCAGGAAAGTBHCCC TCCTWGGACT	3093	3555
VK3-20_13_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACBHCVN CAGTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACBHCVNCAGTCC TCCTWGGACT	3094	3556
VK3-20_14_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACBHCAG TBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACBHCAGTBHCCC TCCTWGGACT	3095	3557
VK3-20_15_2_10	CCTGAAGATTTTGCAGTGTATTACTGTACAGTACGGAVN CBHCCCTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACGGAVNCBHCCC TCCTWGGACT	3096	3558
VK3-20_2_2_10	CCTGAAGATTTTGCAGTGTATTACTGTSMCAGBHCAGGAAAG TAGTCCCTCCTWGGACTTTTGGCGGAGGACCAAG	SMCAGBHCAGGAAAGTAGTCC TCCTWGGACT	3097	3559

VK3-20_3_2_10	CCTGAAGATTTTGCAGIGTATTACTGTSMWCAGTACBHCAG TAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACBHCAGTAGTCC TCCTWGGACT	3098	3560
VK3-20_4_2_10	CCITGAAGATTTTGCAGIGTATTACTGTSMWCAGTACGGAVN CAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACGGAVNCAGTCC TCCTWGGACT	3099	3561
VK3-20_5_2_10	CCTGAAGATTTTGCAGIGTATTACTGTSMWCAGTACGGAAAG TBHCCCTCTWGGACTTTTGGCGGAGGACCAAG	SWMCAGTACGGAAAGTBHCCC TCCTWGGACT	3100	3562
VK3-20_6_2_10	CCTGAAGATTTTGCAGIGTATTACTGTAGSWMBHCAGGAAAG TAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMBHCAGGAAAGTAGTCC TCCTWGGACT	3101	3563
VK3-20_7_2_10	CCITGAAGATTTTGCAGIGTATTACTGTAGSWMTACBHCAG TAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACBHCAGTAGTCC TCCTWGGACT	3102	3564
VK3-20_8_2_10	CCTGAAGATTTTGCAGIGTATTACTGTAGSWMTACGGAVN CAGTCTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACGGAVNCAGTCC TCCTWGGACT	3103	3565
VK3-20_9_2_10	CCITGAAGATTTTGCAGIGTATTACTGTAGSWMTACGGAAAG TBHCCCTCTCTWGGACTTTTGGCGGAGGACCAAG	CAGSWMTACGGAAAGTBHCCC TCCTWGGACT	3104	3566
Jumping Trimer				
VK1-05_t1_0_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TTWCACCT	3105	3567
VK1-05_t1_1_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TMTCACT	3106	3568
VK1-05_t1_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TWGGACT	3107	3569
VK1-05_t2_0_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TTWCACCT	3108	3570
VK1-05_t2_1_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TMTCACT	3109	3571
VK1-05_t2_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TWGGACT	3110	3572
VK1-05_t3_0_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TTWCACCT	3111	3573
VK1-05_t3_1_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TMTCACT	3112	3574
VK1-05_t3_2_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGBHCVRMMBECTACYCTCC TWGGACT	3113	3575
VK1-05_t4_0_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBECTACYCTCC TTWCACCT	3114	3576
VK1-05_t4_1_10	CCTGATGATTTTGCACCTTATTACTGCCAGAGBHCVRMMBECTACY CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBECTACYCTCC TMTCACT	3115	3577

VK1-05 t4 2 10	CCTGATGATTTTGGCAACTTATTACTGCCAGCAGTACVRMMBCYWCY CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGTACVRMMBCYWCYCTCC TWGGACT	3116	3578
VK1-12 t1 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TTWCACI	3117	3579
VK1-12 t1 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3118	3580
VK1-12 t1 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3119	3581
VK1-12 t2 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3120	3582
VK1-12 t2 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3121	3583
VK1-12 t2 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3122	3584
VK1-12 t3 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3123	3585
VK1-12 t3 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3124	3586
VK1-12 t3 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3125	3587
VK1-12 t4 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3126	3588
VK1-12 t4 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3127	3589
VK1-12 t4 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3128	3590
VK1-33 t1 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3129	3591
VK1-33 t1 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3130	3592
VK1-33 t1 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3131	3593
VK1-33 t2 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3132	3594
VK1-33 t2 1 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTMTACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3133	3595
VK1-33 t2 2 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TWGGACT	3134	3596
VK1-33 t3 0 10	CCTGAAGATTTTGCAACTTATTACTGTCTCAGCAGRNANHCNHCCTTCC CTCCTTWACACTTTTGGCGGAGGACCAAG	CAGCAGRNANHCNHCCTTCCCTCC TMTCACT	3135	3597

VK1-33 t3 1 10	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHGATNHCYWCC CTCCTTMCACCTTTTGGGGAGGGACCAAG	CAGCAGBHGATNHCYWCCCTCC TMTCACT	3136	3598
VK1-33 t3 2 10	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHGATNHCYWCC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAGBHGATNHCYWCCCTCC TWGGACT	3137	3599
VK1-33 t4 0 10	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGBHGATNHCYWCC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCCTCC TTWCAC	3138	3600
VK1-33 t4 1 10	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGTACNHCNHCYWCC CTCCTTMCACCTTTTGGGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCCTCC TMTCACT	3139	3601
VK1-33 t4 2 10	CCTGAAGATATTGCAACATATTACTGTCTCAGCAGTACNHCNHCYWCC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAGTACNHCNHCYWCCCTCC TWGGACT	3140	3602
VK1-39 t1 0 10	CC1GAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TTWCAC	3141	3603
VK1-39 t1 1 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTMCACCTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TMTCACT	3142	3604
VK1-39 t1 2 10	CC1GAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TWGGACT	3143	3605
VK1-39 t2 0 10	CC1GAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TTWCAC	3144	3606
VK1-39 t2 1 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTMCACCTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TMTCACT	3145	3607
VK1-39 t2 2 10	CC1GAAGATTTTGCACCTTACTACTGTCTCAGCAAVNABHCNHCACCTC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAAVNABHCNHCACCTCCTCC TWGGACT	3146	3608
VX1-39 t3 0 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAVNATACNHCBHCC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAAVNATACNHCBHCCCTCC TTWCAC	3147	3609
VX1-39 t3 1 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAVNATACNHCBHCC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAAVNATACNHCBHCCCTCC TWGGACT	3148	3610
VX1-39 t3 2 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAVNATACNHCBHCC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAAVNATACNHCBHCCCTCC TMTCACT	3149	3611
VX1-39 t4 0 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAAAGCBHCNHCBHCC CTCCTTWCACTTTTGGGGAGGGACCAAG	CAGCAAAAGCBHCNHCBHCCCTCC TTWCAC	3150	3612
VX1-39 t4 1 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAAAGCBHCNHCBHCC CTCCTTMCACCTTTTGGGGAGGGACCAAG	CAGCAAAAGCBHCNHCBHCCCTCC TMTCACT	3151	3613
VX1-39 t4 2 10	CCTGAAGATTTTGCACCTTACTACTGTCTCAGCAAAAGCBHCNHCBHCC CTCCTTWGGACTTTTGGGGAGGGACCAAG	CAGCAAAAGCBHCNHCBHCCCTCC TWGGACT	3152	3614
VK2-28 t1 0 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVAMNASRMACTC CTCCTTWCACTTTTGGGGAGGGACCAAG	ATGCAGVAMNASRMACTCCTCC TTWCAC	3153	3615
VK2-28 t1 1 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVAMNASRMACTC CTCCTTMCACCTTTTGGGGAGGGACCAAG	ATGCAGVAMNASRMACTCCTCC TMTCACT	3154	3616
VK2-28 t1 2 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVAMNASRMACTC CTCCTTWGGACTTTTGGGGAGGGACCAAG	ATGCAGVAMNASRMACTCCTCC TWGGACT	3155	3617

VK2-28 t2 0 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNACAGVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNAMNACAGVBCCCTCC TTTWCACCT	3156	3618
VK2-28 t2 1 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNACAGVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNAMNACAGVBCCCTCC TTTWCACCT	3157	3619
VK2-28 t2 2 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNAMNACAGVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNAMNACAGVBCCCTCC TTTWCACCT	3158	3620
VK2-28 t3 0 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCSRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMBVBCCCTCC TTTWCACCT	3159	3621
VK2-28 t3 1 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCSRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMBVBCCCTCC TTTWCACCT	3160	3622
VK2-28 t3 2 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGVNACTCSRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGVNACTCSRMBVBCCCTCC TTTWCACCT	3161	3623
VK2-28 t4 0 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCAMNASRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGGCAMNASRMBVBCCCTCC TTTWCACCT	3162	3624
VK2-28 t4 1 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCAMNASRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGGCAMNASRMBVBCCCTCC TTTWCACCT	3163	3625
VK2-28 t4 2 10	GCTGAGGATGTTGGGGTTTATTACTGTCATGCAGGCAMNASRMBVBCC CTCCTTWCACCTTTTGGCGGAGGACCAAG	ATGCAGGCAMNASRMBVBCCCTCC TTTWCACCT	3164	3626
VK3-11 t1 0 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3165	3627
VK3-11 t1 1 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3166	3628
VK3-11 t1 2 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3167	3629
VK3-11 t2 0 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3168	3594
VK3-11 t2 1 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3169	3595
VK3-11 t2 2 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3170	3596
VK3-11 t3 0 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3171	3630
VK3-11 t3 1 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3172	3631
VK3-11 t3 2 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3173	3632
VK3-11 t4 0 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3174	3633
VK3-11 t4 1 10	CCTGAAGATTTTGCAGTTTATTACTGTCAGGBHCNHCNHCNCTGGC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCNCTGGCCTCC TTTWCACCT	3175	3634

VK3-11 t4 2 10	CCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGAGANHCNHCYWCC CTCCTWGGACTTTTGGCGGAGGACCAAG	CAGCAGAGANHCNHCYWCCCTCC TWGGACT	3176	3635
VK3-15 t1 0 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TTWCACCT	3177	3627
VK3-15 t1 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TMTCACT	3178	3628
VK3-15 t1 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TWGGACT	3179	3629
VK3-15 t2 0 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TTWCACCT	3180	3594
VK3-15 t2 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TMTCACT	3181	3595
VK3-15 t2 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TWGGACT	3182	3596
VK3-15 t3 0 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TTWCACCT	3183	3636
VK3-15 t3 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TMTCACT	3184	3637
VK3-15 t3 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TWGGACT	3185	3638
VK3-15 t4 0 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TTWCACCT	3186	3600
VK3-15 t4 1 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TMTCACT	3187	3601
VK3-15 t4 2 10	CAGTCTGAAGATTTTGCAGTTTATTACTGTCTCAGCAGBHCNHCNHC GGCCTCCTTWCACCTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCCTGCCCTCC TWGGACT	3188	3602
VK3-20 t1 0 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TTWCACCT	3189	3639
VK3-20 t1 1 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TMTCACT	3190	3640
VK3-20 t2 0 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TWGGACT	3191	3641
VK3-20 t2 1 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TMTCACT	3192	3642
VK3-20 t2 2 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TWGGACT	3193	3643
VK3-20 t3 0 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TMTCACT	3194	3644
VK3-20 t3 1 10	CCTGAAGATTTTGCAGTGTATTACTGTCTCAGCAGBHCNHCNHC CTCCTTWCACCTTTTGGCGGAGGACCAAG	CAGCAGBHCNHCNHCAGTCCCTCC TWGGACT	3195	3645

VK3-20 t3 1 10	CCTGAAGATTTTGCAGTGTATTACTGTCTAGCAGBHC GGAVNCBHCC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC GGAVNCBHCCCTCC TMTCACT	3196	3646
VK3-20 t3 2 10	CCTGAAGATTTTGCAGTGTATTACTGTCTAGCAGBHC GGAVNCBHCC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGBHC GGAVNCBHCCCTCC TWGGACT	3197	3647
VK3-20 t4 0 10	CCTGAAGATTTTGCAGTGTATTACTGTCTAGCAGBHC VNCBHCC CTCCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC VNCBHCCCTCC TTCAC T	3198	3648
VK3-20 t4 1 10	CCTGAAGATTTTGCAGTGTATTACTGTCTAGCAGTACBHC VNCBHCC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC VNCBHCCCTCC TMTCACT	3199	3649
VK3-20 t4 2 10	CCTGAAGATTTTGCAGTGTATTACTGTCTAGCAGTACBHC VNCBHCC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC VNCBHCCCTCC TWGGACT	3200	3650
VK4-01 t1 0 10	CCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TTCAC T	3201	3651
VK4-01 t1 1 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TMTCACT	3202	3652
VK4-01 t1 2 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TWGGACT	3203	3653
VK4-01 t2 0 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TTCAC T	3204	3642
VK4-01 t2 1 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TMTCACT	3205	3643
VK4-01 t2 2 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC HCNHCACTC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGBHC HCNHCACTCCCTCC TWGGACT	3206	3644
VK4-01 t3 0 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC TACNHC BHCC CTCCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC TACNHC BHCCCTCC TTCAC T	3207	3654
VK4-01 t3 1 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC TACNHC BHCC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGBHC TACNHC BHCCCTCC TMTCACT	3208	3655
VK4-01 t3 2 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGBHC TACNHC BHCC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGBHC TACNHC BHCCCTCC TWGGACT	3209	3656
VK4-01 t4 0 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGTACBHC HCNHCBHCC CTCCTTWCAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC HCNHCBHCCCTCC TTCAC T	3210	3657
VK4-01 t4 1 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGTACBHC HCNHCBHCC CTCCTTMCAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC HCNHCBHCCCTCC TMTCACT	3211	3658
VK4-01 t4 2 10	GCTGAAGATGTGGCAGTTTATTACTGTCTAGCAGTACBHC HCNHCBHCC CTCCTWGGAC TTTTGGCGGAGGACCAAG	CAGCAGTACBHC HCNHCBHCCCTCC TWGGACT	3212	3659

Table 8. Number of unique CDR13 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 TS21. Amino acid mismatches in the theoretical design are indicated in bold.

Test Case	CDRH3 Sequence from HPS	SEQ ID NO	Mismatches	TN1	SEQ ID NO	DH	SEQ ID NO	N2	H3-JH	SEQ ID NO
1	RTAHFDY	3660	0	R		TA		H	HFDY	4582
2.1	VGIVGAASY	3661	0	V		GIVGA	3751	AS	Y	
2.2	VGIVGAASY	3661	0	VG		IVGA	3755	AS	Y	
3.1	DRYSGHDLGY	3662	1	DR		YSGYD	4389	LG	Y	
4.1	GIAAADSNWLDP	3663	1	-		GIAAA	4448	D	SNW F DP	4600
4.2	GIAAADSNWLDP	3663	1	-		IAAA	4452	D	SNW F DP	4600
5.1	ERTINWGWGGVYAFDI	3664	3	EGTG	3707	NWG		GGV	YAFDI	4540
5.2	ERTINWGWGGVYAFDI	3664	3	EGTG	3707	NWG		WGT	YAFDI	4540
5.3	ERTINWGWGGVYAFDI	3664	3	ERGG	3719	NWG		GGV	YAFDI	4540

5.4	ERTINWGGVYAEDI	3	3664	ERGG	3719	NWG	WGT	YAFDI	4540
-----	----------------	---	------	------	------	-----	-----	-------	------

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

TN1 Segment Name	Amino Acid Sequence	TN1 Segment Name	Amino Acid Sequence	SEQ ID NO
P000	-	P107	PT	n/a
P001	E	P108	EPT	n/a
P002	D	P109	DPT	n/a
P003	G	P110	GPT	n/a
P004	EG	P111	PV	n/a
P005	DG	P112	EPV	n/a
P006	GG	P113	DPV	n/a
P007	R	P114	GPV	n/a
P008	ER	P115	RP	n/a
P009	DR	P116	ERP	n/a
P010	GR	P117	DRP	n/a
P011	S	P118	GRP	n/a
P012	ES	P119	SP	n/a
P013	DS	P120	ESP	n/a
P014	GS	P121	DSP	n/a
P015	P	P122	GSP	n/a
P016	EP	P123	LP	n/a
P017	DP	P124	ELP	n/a
P018	GP	P125	DLP	n/a
P019	L	P126	GLP	n/a
P020	EL	P127	AP	n/a
P021	DL	P128	EAP	n/a
P022	GL	P129	DAP	n/a
P023	A	P130	GAP	n/a
P024	EA	P131	TP	n/a
P025	DA	P132	ETP	n/a
P026	GA	P133	DTP	n/a
P027	T	P134	GTP	n/a

P028	ET	P135	VP	n/a
P029	DT	P136	EVP	n/a
P030	GT	P137	DVP	n/a
P031	V	P138	GVP	n/a
P032	EV	P139	AGG	n/a
P033	DV	P140	EAGG	3665
P034	GV	P141	DAGG	3666
P035	EGG	P142	GAGG	3667
P036	DGG	P143	EGAG	3668
P037	GGG	P144	DGAG	3669
P038	EGR	P145	GGAG	3670
P039	DGR	P146	EGGA	3671
P040	GGR	P147	DGGA	3672
P041	EGS	P148	GGGA	3673
P042	DGS	P149	EGGG	3674
P043	GS	P150	DGGG	3675
P044	EGP	P151	GGGG	3676
P045	DGP	P152	EGGL	3677
P046	GGP	P153	DGGL	3678
P047	EGL	P154	GGGL	3679
P048	DGL	P155	EGGP	3680
P049	GGL	P156	DGGP	3681
P050	EGA	P157	GGGP	3682
P051	DGA	P158	EGGR	3683
P052	GGA	P159	DGGR	3684
P053	EGT	P160	GGGR	3685
P054	DGT	P161	EGGS	3686
P055	GGT	P162	DGGS	3687
P056	EGV	P163	GGGS	3688
P057	DGV	P164	EGGT	3689
P058	GGV	P165	DGGT	3690
P059	RG	P166	GGGT	3691

P060	ERG	P167	EGGV	3692
P061	DRG	P168	DGGV	3693
P062	GRG	P169	GGGV	3694
P063	SG	P170	EGLG	3695
P064	ESG	P171	DGLG	3696
P065	DSG	P172	GGLG	3697
P066	GSG	P173	EGPG	3698
P067	PG	P174	DGPG	3699
P068	EPG	P175	GGPG	3700
P069	DPG	P176	EGRG	3701
P070	GPG	P177	DGRG	3702
P071	LG	P178	GGRG	3703
P072	ELG	P179	EGSG	3704
P073	DLG	P180	DGSG	3705
P074	GLG	P181	GGSG	3706
P075	AG	P182	EGTG	3707
P076	EAG	P183	DGTG	3708
P077	DAG	P184	GGTG	3709
P078	GAG	P185	EGVG	3710
P079	TG	P186	DGVG	3711
P080	ETG	P187	GGVG	3712
P081	DTG	P188	LGG	n/a
P082	GTG	P189	ELGG	3713
P083	VG	P190	DLGG	3714
P084	EVG	P191	GLGG	3715
P085	DVG	P192	PGG	n/a
P086	GVG	P193	EPGG	3716
P087	PR	P194	DPGG	3717
P088	EPR	P195	GPGG	3718
P089	DPR	P196	RGG	n/a
P090	GPR	P197	ERGG	3719
P091	PS	P198	DRGG	3720

P092	EPS	P199	GRGG	3721
P093	DPS	P200	SGG	n/a
P094	GPS	P201	ESGG	3722
P095	PP	P202	DSGG	3723
P096	EPP	P203	GSGG	3724
P097	DPP	P204	TGG	n/a
P098	GPP	P205	ETGG	3725
P099	PL	P206	DTGG	3726
P100	EPL	P207	GIGG	3727
P101	DPL	P208	VGG	n/a
P102	GPL	P209	EVGG	3728
P103	PA	P210	DVGG	3729
P104	EPA	P211	GVGG	3730
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	3731	DHUNIV_557	YDYV	4107
DHUNIV_002	GTTGT	3732	DHUNIV_558	DYVW	4108
DHUNIV_003	TTGTT	3733	DHUNIV_559	YVWG	4109
DHUNIV_004	GTTG	3734	DHUNIV_560	VWGS	4110
DHUNIV_005	TTGT	3735	DHUNIV_561	WGSY	4111
DHUNIV_006	TGTT	3736	DHUNIV_562	GSYA	4112
DHUNIV_007	GTT	n/a	DHUNIV_563	SYAY	4113
DHUNIV_008	TTG	n/a	DHUNIV_564	YAYT	4114
DHUNIV_009	TGT	n/a	DHUNIV_565	YYD	n/a
DHUNIV_010	GT	n/a	DHUNIV_566	YDY	n/a
DHUNIV_011	TT	n/a	DHUNIV_567	DYV	n/a
DHUNIV_012	TG	n/a	DHUNIV_568	YVW	n/a
DHUNIV_013	VQLER	3737	DHUNIV_569	VWG	n/a
DHUNIV_014	VQLE	3738	DHUNIV_570	WGS	n/a
DHUNIV_015	QLER	3739	DHUNIV_571	SYA	n/a
DHUNIV_016	VQL	n/a	DHUNIV_572	YAY	n/a
DHUNIV_017	QLE	n/a	DHUNIV_573	AYT	n/a
DHUNIV_018	LER	n/a	DHUNIV_574	YD	n/a
DHUNIV_019	VQ	n/a	DHUNIV_575	DY	n/a
DHUNIV_020	QL	n/a	DHUNIV_576	YV	n/a
DHUNIV_021	LE	n/a	DHUNIV_577	VW	n/a
DHUNIV_022	ER	n/a	DHUNIV_578	WG	n/a
DHUNIV_023	YNWND	3740	DHUNIV_579	IMITFGGVMLIP	4115
DHUNIV_024	YNWN	3741	DHUNIV_580	IMITFGGVMLI	4116
DHUNIV_025	NWND	3742	DHUNIV_581	MITFGGVMLIP	4117
DHUNIV_026	YNW	n/a	DHUNIV_582	IMITFGGVML	4118
DHUNIV_027	NWN	n/a	DHUNIV_583	MITFGGVMLI	4119
DHUNIV_028	WND	n/a	DHUNIV_584	ITFGGVMLIP	4120
DHUNIV_029	YN	n/a	DHUNIV_585	IMITFGGVM	4121

DHUNIV_030	NW	n/a	DHUNIV_586	MITFGGVML	4122
DHUNIV_031	WN	n/a	DHUNIV_587	ITFGGVMLI	4123
DHUNIV_032	ND	n/a	DHUNIV_588	TFGGVMLIP	4124
DHUNIV_033	GITGTT	3743	DHUNIV_589	IMITFGGV	4125
DHUNIV_034	GITGT	3744	DHUNIV_590	MITFGGVM	4126
DHUNIV_035	ITGTT	3745	DHUNIV_591	ITFGGVML	4127
DHUNIV_036	GITG	3746	DHUNIV_592	TFGGVMLI	4128
DHUNIV_037	ITGT	3747	DHUNIV_593	FGGVMLIP	4129
DHUNIV_038	GIT	n/a	DHUNIV_594	IMITFGG	4130
DHUNIV_039	ITG	n/a	DHUNIV_595	MITFGGV	4131
DHUNIV_040	GI	n/a	DHUNIV_596	ITFGGVM	4132
DHUNIV_041	IT	n/a	DHUNIV_597	TFGGVML	4133
DHUNIV_042	GIVGATT	3748	DHUNIV_598	FGGVMLI	4134
DHUNIV_043	GIVGAT	3749	DHUNIV_599	GGVMLIP	4135
DHUNIV_044	IVGATT	3750	DHUNIV_600	IMITFG	4136
DHUNIV_045	GIVGA	3751	DHUNIV_601	MITFGG	4137
DHUNIV_046	IVGAT	3752	DHUNIV_602	ITFGGV	4138
DHUNIV_047	VGATT	3753	DHUNIV_603	TFGGVM	4139
DHUNIV_048	GIVG	3754	DHUNIV_604	FGGVML	4140
DHUNIV_049	IVGA	3755	DHUNIV_605	GGVMLI	4141
DHUNIV_050	VGAT	3756	DHUNIV_606	GVMLIP	4142
DHUNIV_051	GATT	3757	DHUNIV_607	IMITF	4143
DHUNIV_052	GIV	n/a	DHUNIV_608	MITFG	4144
DHUNIV_053	IVG	n/a	DHUNIV_609	ITFGG	4145
DHUNIV_054	VGA	n/a	DHUNIV_610	TFGGV	4146
DHUNIV_055	GAT	n/a	DHUNIV_611	FGGVM	4147
DHUNIV_056	ATT	n/a	DHUNIV_612	GGVML	4148
DHUNIV_057	IV	n/a	DHUNIV_613	GVMLI	4149
DHUNIV_058	VG	n/a	DHUNIV_614	VMLIP	4150
DHUNIV_059	GA	n/a	DHUNIV_615	IMIT	4151
DHUNIV_060	AT	n/a	DHUNIV_616	MITF	4152
DHUNIV_061	WELL	3758	DHUNIV_617	ITFG	4153

DHUNIV_062	WEL	n/a	DEUNIV_618	TFGG	4154
DHUNIV_063	ELL	n/a	DEUNIV_619	FGGV	4155
DHUNIV_064	WE	n/a	DEUNIV_620	GGVM	4156
DHUNIV_065	EL	n/a	DEUNIV_621	GVML	4157
DHUNIV_066	LL	n/a	DEUNIV_622	VMLI	4158
DHUNIV_067	YSGSYY	3759	DEUNIV_623	MLIP	4159
DHUNIV_068	YSGSY	3760	DEUNIV_624	IMI	n/a
DHUNIV_069	SGSYY	3761	DEUNIV_625	MIT	n/a
DHUNIV_070	YSGS	3762	DEUNIV_626	ITF	n/a
DHUNIV_071	SGSY	3763	DEUNIV_627	TFG	n/a
DHUNIV_072	GSYY	3764	DEUNIV_628	FGG	n/a
DHUNIV_073	YSG	n/a	DEUNIV_629	GGV	n/a
DHUNIV_074	SGS	n/a	DEUNIV_630	GVM	n/a
DHUNIV_075	GSY	n/a	DEUNIV_631	VML	n/a
DHUNIV_076	SY	n/a	DEUNIV_632	MLI	n/a
DHUNIV_077	YS	n/a	DEUNIV_633	LIP	n/a
DHUNIV_078	SG	n/a	DEUNIV_634	IM	n/a
DHUNIV_079	GS	n/a	DEUNIV_635	MI	n/a
DHUNIV_080	SY	n/a	DEUNIV_636	TF	n/a
DHUNIV_081	YY	n/a	DEUNIV_637	VM	n/a
DHUNIV_082	LEL	n/a	DEUNIV_638	LI	n/a
DHUNIV_083	YNWNY	3765	DEUNIV_639	WLLL	4160
DHUNIV_084	NWNY	3766	DEUNIV_640	WLL	n/a
DHUNIV_085	WNY	n/a	DEUNIV_641	WL	n/a
DHUNIV_086	NY	n/a	DEUNIV_642	YYYDSSGYYY	4161
DHUNIV_087	RIL	n/a	DEUNIV_643	YYYDSSGYY	4162
DHUNIV_088	LLL	n/a	DEUNIV_644	YYDSSGYYY	4163
DHUNIV_089	RI	n/a	DEUNIV_645	YYYDSSGY	4164
DHUNIV_090	IL	n/a	DEUNIV_646	YYDSSGYY	4165
DHUNIV_091	WW	n/a	DEUNIV_647	YDSSGYYY	4166
DHUNIV_092	GYCSGGSCYS	3767	DEUNIV_648	YYYDSSG	4167
DHUNIV_093	GYCSGGSCY	3768	DEUNIV_649	YYDSSGY	4168

DHUNIV_094	YCSGGSCYS	3769	DHUNIV_650	YDSSGYY	4169
DHUNIV_095	GYCSGGSC	3770	DHUNIV_651	DSSGYYY	4170
DHUNIV_096	YCSGGSCY	3771	DHUNIV_652	YYDSS	4171
DHUNIV_097	CSGGSCYS	3772	DHUNIV_653	YYDSSG	4172
DHUNIV_098	YCSGGSC	3773	DHUNIV_654	YDSSGY	4173
DHUNIV_099	CSGGSCY	3774	DHUNIV_655	DSSGYY	4174
DHUNIV_100	CSGGSC	3775	DHUNIV_656	SSGYYY	4175
DHUNIV_101	SGGS	3776	DHUNIV_657	YYYDS	4176
DHUNIV_102	SGG	n/a	DHUNIV_658	YYDSS	4177
DHUNIV_103	GGG	n/a	DHUNIV_659	YDSSG	4178
DHUNIV_104	GY	n/a	DHUNIV_660	DSSGY	4179
DHUNIV_105	GG	n/a	DHUNIV_661	SSGYY	4180
DHUNIV_106	DIVVVVAATP	3777	DHUNIV_662	SGYYY	4181
DHUNIV_107	DIVVVVAAT	3778	DHUNIV_663	YYD	4182
DHUNIV_108	IVVVVAATP	3779	DHUNIV_664	YYDS	4183
DHUNIV_109	DIVVVVAA	3780	DHUNIV_665	YDSS	4184
DHUNIV_110	IVVVVAAT	3781	DHUNIV_666	DSSG	4185
DHUNIV_111	VVVVAATP	3782	DHUNIV_667	SSGY	4186
DHUNIV_112	DIVVVVA	3783	DHUNIV_668	SGYY	4187
DHUNIV_113	IVVVVAA	3784	DHUNIV_669	GYYY	4188
DHUNIV_114	VVVVAAT	3785	DHUNIV_670	YDS	n/a
DHUNIV_115	VVVAATP	3786	DHUNIV_671	DSS	n/a
DHUNIV_116	DIVVVV	3787	DHUNIV_672	SSG	n/a
DHUNIV_117	IVVVVA	3788	DHUNIV_673	SGY	n/a
DHUNIV_118	VVVVAA	3789	DHUNIV_674	GYY	n/a
DHUNIV_119	VVVAAT	3790	DHUNIV_675	DS	n/a
DHUNIV_120	VVAATP	3791	DHUNIV_676	ITMIVVVITT	4189
DHUNIV_121	DIVVV	3792	DHUNIV_677	ITMIVVVIT	4190
DHUNIV_122	IVVVV	3793	DHUNIV_678	TMIVVVITT	4191
DHUNIV_123	VVVVA	3794	DHUNIV_679	ITMIVVVI	4192
DHUNIV_124	VVVAA	3795	DHUNIV_680	TMIVVVIT	4193
DHUNIV_125	VVAAT	3796	DHUNIV_681	MIVVVITT	4194

DHUNIV_126	VAATP	3797	DEUNIV_682	ITMIVVV	4195
DHUNIV_127	DIVV	3798	DEUNIV_683	TMIVVVI	4196
DHUNIV_128	IVVV	3799	DEUNIV_684	MIVVVIT	4197
DHUNIV_129	VVVV	3800	DEUNIV_685	IVVVITT	4198
DHUNIV_130	VVVA	3801	DEUNIV_686	ITMIVV	4199
DHUNIV_131	VVAA	3802	DEUNIV_687	TMIVVV	4200
DHUNIV_132	VAAT	3803	DEUNIV_688	MIVVVI	4201
DHUNIV_133	AATP	3804	DEUNIV_689	IVVVIT	4202
DHUNIV_134	DIV	n/a	DEUNIV_690	VVVITT	4203
DHUNIV_135	IVV	n/a	DEUNIV_691	ITMIV	4204
DHUNIV_136	VVV	n/a	DEUNIV_692	TMIVV	4205
DHUNIV_137	VVA	n/a	DEUNIV_693	MIVVV	4206
DHUNIV_138	VAA	n/a	DEUNIV_694	VVVIT	4207
DHUNIV_139	AAT	n/a	DEUNIV_695	VVITT	4208
DHUNIV_140	ATP	n/a	DEUNIV_696	ITMI	4209
DHUNIV_141	DI	n/a	DEUNIV_697	TMIV	4210
DHUNIV_142	VV	n/a	DEUNIV_698	MIVV	4211
DHUNIV_143	VA	n/a	DEUNIV_699	VVIT	4212
DHUNIV_144	AA	n/a	DEUNIV_700	VITT	4213
DHUNIV_145	TP	n/a	DEUNIV_701	TMI	n/a
DHUNIV_146	YQLL	3805	DEUNIV_702	MIV	n/a
DHUNIV_147	YQL	n/a	DEUNIV_703	VIT	n/a
DHUNIV_148	QLL	n/a	DEUNIV_704	ITT	n/a
DHUNIV_149	YQ	n/a	DEUNIV_705	VLRFLEWLLY	4214
DHUNIV_150	GYCSSTSCYA	3806	DEUNIV_706	VLRFLEWLL	4215
DHUNIV_151	GYCSSTSCY	3807	DEUNIV_707	LRFLEWLLY	4216
DHUNIV_152	YCSSTSCYA	3808	DEUNIV_708	VLRFLEWL	4217
DHUNIV_153	GYCSSTSC	3809	DEUNIV_709	LRFLEWLL	4218
DHUNIV_154	YCSSTSCY	3810	DEUNIV_710	RFLEWLLY	4219
DHUNIV_155	CSSTSCYA	3811	DEUNIV_711	VLRFLEW	4220
DHUNIV_156	YCSSTSC	3812	DEUNIV_712	LRFLEWL	4221
DHUNIV_157	CSSTSCY	3813	DEUNIV_713	RFLEWLL	4222

DHUNIV_158	CSSTSC	3814	DHUNIV_714	FLEWLLY	4223
DHUNIV_159	SSTS	3815	DHUNIV_715	VLRFL	4224
DHUNIV_160	SSI	n/a	DHUNIV_716	LRFL	4225
DHUNIV_161	SIS	n/a	DHUNIV_717	RFLEW	4226
DHUNIV_162	SS	n/a	DHUNIV_718	FLEWLL	4227
DHUNIV_163	ST	n/a	DHUNIV_719	LEWLLY	4228
DHUNIV_164	TS	n/a	DHUNIV_720	VLRFL	4229
DHUNIV_165	YA	n/a	DHUNIV_721	LRFL	4230
DHUNIV_166	DIVVVPAAMP	3816	DHUNIV_722	RFLEW	4231
DHUNIV_167	DIVVVPAAM	3817	DHUNIV_723	FLEWL	4232
DHUNIV_168	IVVVPAAMP	3818	DHUNIV_724	LEWLL	4233
DHUNIV_169	DIVVVPA	3819	DHUNIV_725	EWLLY	4234
DHUNIV_170	IVVVPAAM	3820	DHUNIV_726	VLR	4235
DHUNIV_171	VVVPAAMP	3821	DHUNIV_727	LRFL	4236
DHUNIV_172	DIVVVPA	3822	DHUNIV_728	RFLE	4237
DHUNIV_173	IVVVPA	3823	DHUNIV_729	FLEW	4238
DHUNIV_174	VVVPAAM	3824	DHUNIV_730	LEWL	4239
DHUNIV_175	VVPAAAMP	3825	DHUNIV_731	EWLL	4240
DHUNIV_176	DIVVVP	3826	DHUNIV_732	WLLY	4241
DHUNIV_177	IVVVPA	3827	DHUNIV_733	VLR	n/a
DHUNIV_178	VVVPA	3828	DHUNIV_734	LR	n/a
DHUNIV_179	VVPAA	3829	DHUNIV_735	RFL	n/a
DHUNIV_180	VPAAMP	3830	DHUNIV_736	FLE	n/a
DHUNIV_181	IVVVP	3831	DHUNIV_737	LEW	n/a
DHUNIV_182	VVPA	3832	DHUNIV_738	EWL	n/a
DHUNIV_183	VVPA	3833	DHUNIV_739	RF	n/a
DHUNIV_184	VPAAM	3834	DHUNIV_740	FL	n/a
DHUNIV_185	PAAMP	3835	DHUNIV_741	EW	n/a
DHUNIV_186	VVVP	3836	DHUNIV_742	YDFWSGYYT	4242
DHUNIV_187	VVPA	3837	DHUNIV_743	YDFWSGYY	4243
DHUNIV_188	VPA	3838	DHUNIV_744	YDFWSGYYT	4244
DHUNIV_189	PAAM	3839	DHUNIV_745	YDFWSGY	4245

DEMANDES OU BREVETS VOLUMINEUX

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

CECI EST LE TOME __1__ DE __2__

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__

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

WHAT IS CLAIMED IS:

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

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

TN1 is a polypeptide corresponding to any of the TN1 polypeptides of Tables 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 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 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 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 the polynucleotides encode CDRH3 polypeptides produced by assembling combinations of members of TN1, DH, N2, and H3-JH segment pools provided in any one of Tables 23-25.
3. The library of claim 1, wherein the polynucleotides encode CDRH3 polypeptides produced by assembling combinations of members of the TN1 segment pool provided in Table 26, members of the DH segment pool provided in Table 28, members of the N2 segment pool provided in Table 30 and members of the H3-JH segment pool provided in Table 32.

4. 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.
5. 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.
6. The library of claim 1, further comprising a polynucleotide encoding one or more light chain variable domain polypeptides.
7. The library of claim 6, wherein the polypeptides are expressed as full-length IgGs.
8. A method of making a library of synthetic polynucleotides encoding at least 10^4 polypeptides comprising CDRH3 sequences represented by the formula [TN1]-[DH]-[N2]-[H3-JH], the method comprising:
 - (a) providing a theoretical segment pool containing TN1, DH, N2, and H3-JH segments;
 - (b) providing a reference set of preimmune CDRH3 sequences having sequence diversities and length diversities similar to naturally occurring human antibody sequences before these sequences have undergone negative selection and/or hypermutation;
 - (c) progressively linking the TN1, DH, N2, and H3-JH segments contained in the theoretical segment pool of (a) and aligning the segment combinations with the CDRH3 sequences in the reference set of (b) to identify the closest match(es);
 - (d) selecting segments from the closest match(es) identified in step (c) for inclusion in a synthetic library; and
 - (e) synthesizing the synthetic CDRH3 library.

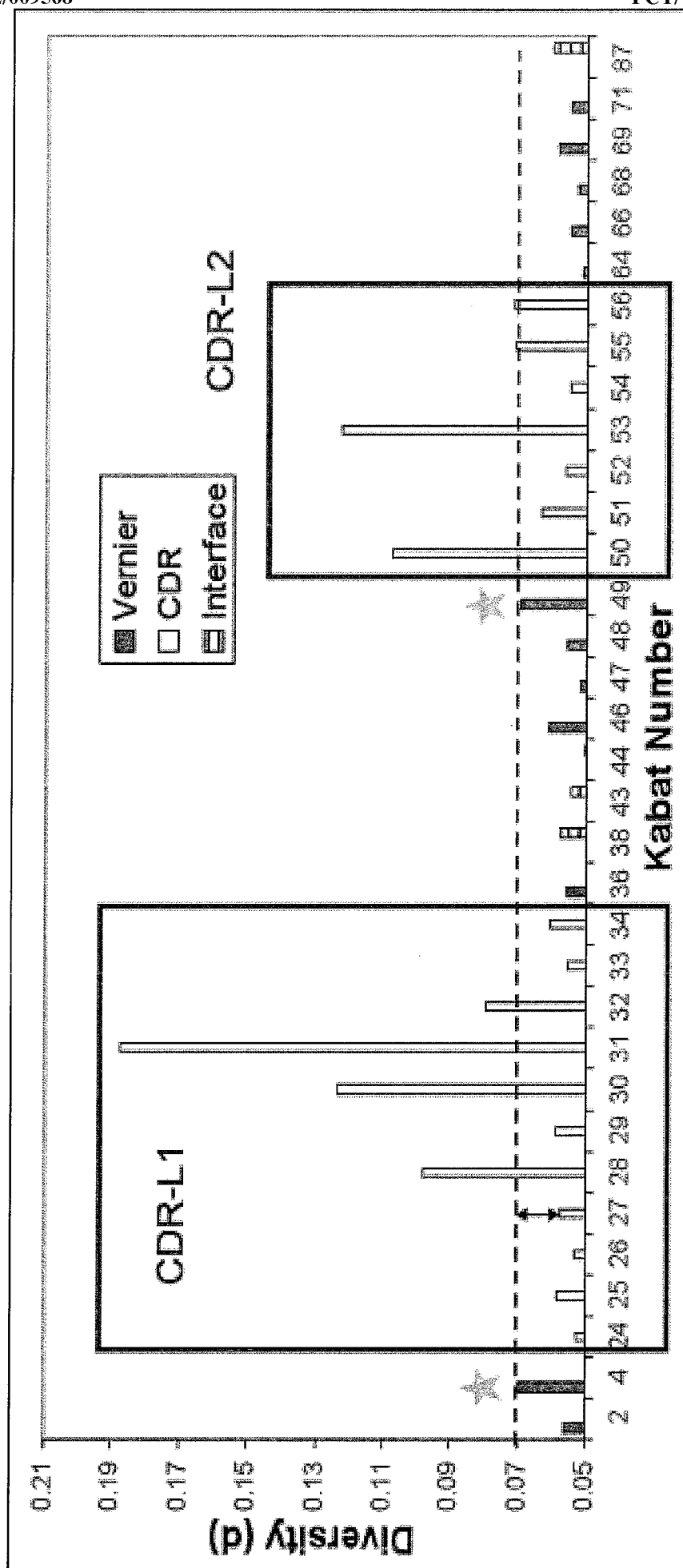
9. The method of claim 8, wherein the segments selected for inclusion in the synthetic library are selected according to their segment usage weight in the reference set of CDRH3 sequences.
10. The method of claim 8, wherein the segments selected for inclusion in the synthetic library are selected according to one or more physicochemical properties.
11. The method of claim 8, further comprising selecting additional TN1 and N2 segments occurring in the reference set but not in the theoretical segment pool.
12. The method of claim 8, wherein stop codons are reduced or eliminated from the library.
13. The method of claim 8, wherein unpaired Cys residues, N-linked glycosylation motifs, and deamidation motifs are reduced or eliminated in translation products of the library.
14. The method of claim 8, wherein the DH segments and H3-JH segments are progressively truncated prior to matching with the CDRH3 sequences in the reference set.
15. The method of claim 8, further comprising introducing one or two degenerate codons in a segment selected from the group consisting of DH, N2, and combinations thereof.
16. The method of claim 8, further comprising introducing one degenerate codon in the H3-JH segment.
17. The method of claim 8, wherein the DH segments are provided according to the following steps:
 - (i) obtaining a set of polynucleotide sequences of human IGHD genes and alleles;
 - (ii) translating the polynucleotide sequences of (i), wherein translation of a partial codon containing two untranslated bases is completed to produce a full codon if the two bases encode a single amino acid regardless of the third position;
 - (iii) progressively deleting amino acid sequences of (ii) at their N- and/or C-termini until at least two amino acids remain; and

- (iv) collecting unique sequences from (iii) that contain 2 or more amino acids to thereby provide the DH theoretical segment pool.
18. The method of claim 8, wherein the DH segments are provided according to the following steps:
- (i) obtaining a set of polynucleotide sequences of human IGHD genes and alleles;
 - (ii) progressively deleting one or more nucleotide bases from the polynucleotide sequences of (i) on either or both of the 5' and 3' ends;
 - (iii) systematically adding 0, 1 and/or 2 or more nucleotides to the 5' and/or 3' ends of the polynucleotide sequences from (ii); and
 - (iv) translating the polynucleotide sequences of (iii) and selecting unique amino acid sequences to thereby provide the DH theoretical segment pool.
19. The method of claim 18, wherein step (ii) includes progressively deleting nucleotide bases until 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 remain.
20. The method of claim 18, wherein step (iii) includes:
- A) addition of 0, 1 or 2 nucleotides to the 5' and/or 3' ends of the polynucleotide sequences from (ii);
 - B) addition of 1 or 2 nucleotides at both the 5' and 3' ends of the polynucleotide sequences from (ii);
 - C) addition of 1 nucleotide on the 5' end combined with 1 or 2 nucleotides on the 3' end of the polynucleotide sequences from (ii);
 - D) addition of 1 or 2 nucleotides to the 3' end only of the polynucleotide sequences from (ii); and/or

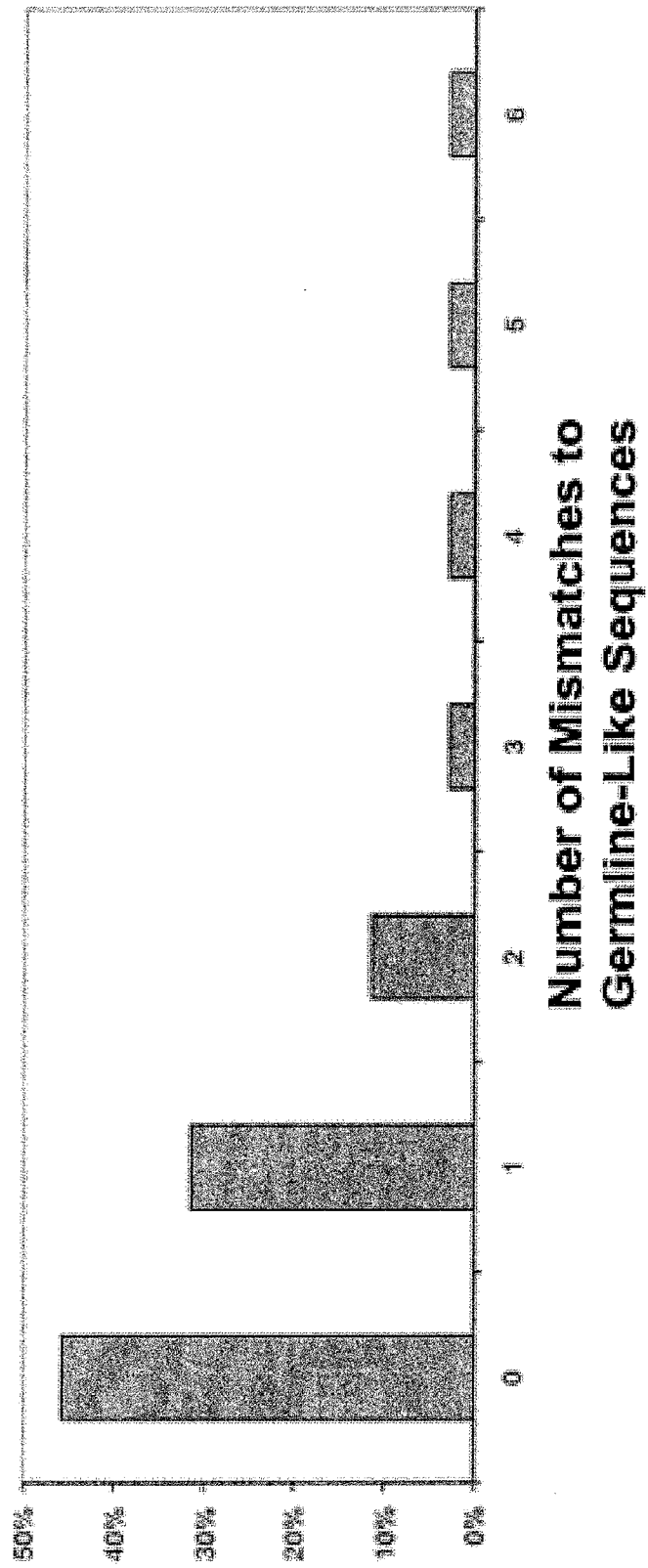
- E) addition of 2 nucleotides on the 5' end combined with 1 or 2 nucleotides on the 3' end of the polynucleotide sequences from (ii).
21. The method of claim 17 or 18, wherein the unique sequences of (iv) do not have stop codons, unpaired Cys residues, deamidation motifs, or Asn in the last or next to last position that can lead to N-linked glycosylation motifs.
 22. The method of claim 8, wherein the theoretical segment pool contains TN1 segments 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.
 23. The method of claim 8, wherein the theoretical segment pool contains DH segments 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.
 24. The method of claim 8, wherein the theoretical segment pool contains N2 segments 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.
 25. The method of claim 8, wherein the theoretical segment pool contains H3-JH segments 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.

1/15

FIGURE 1

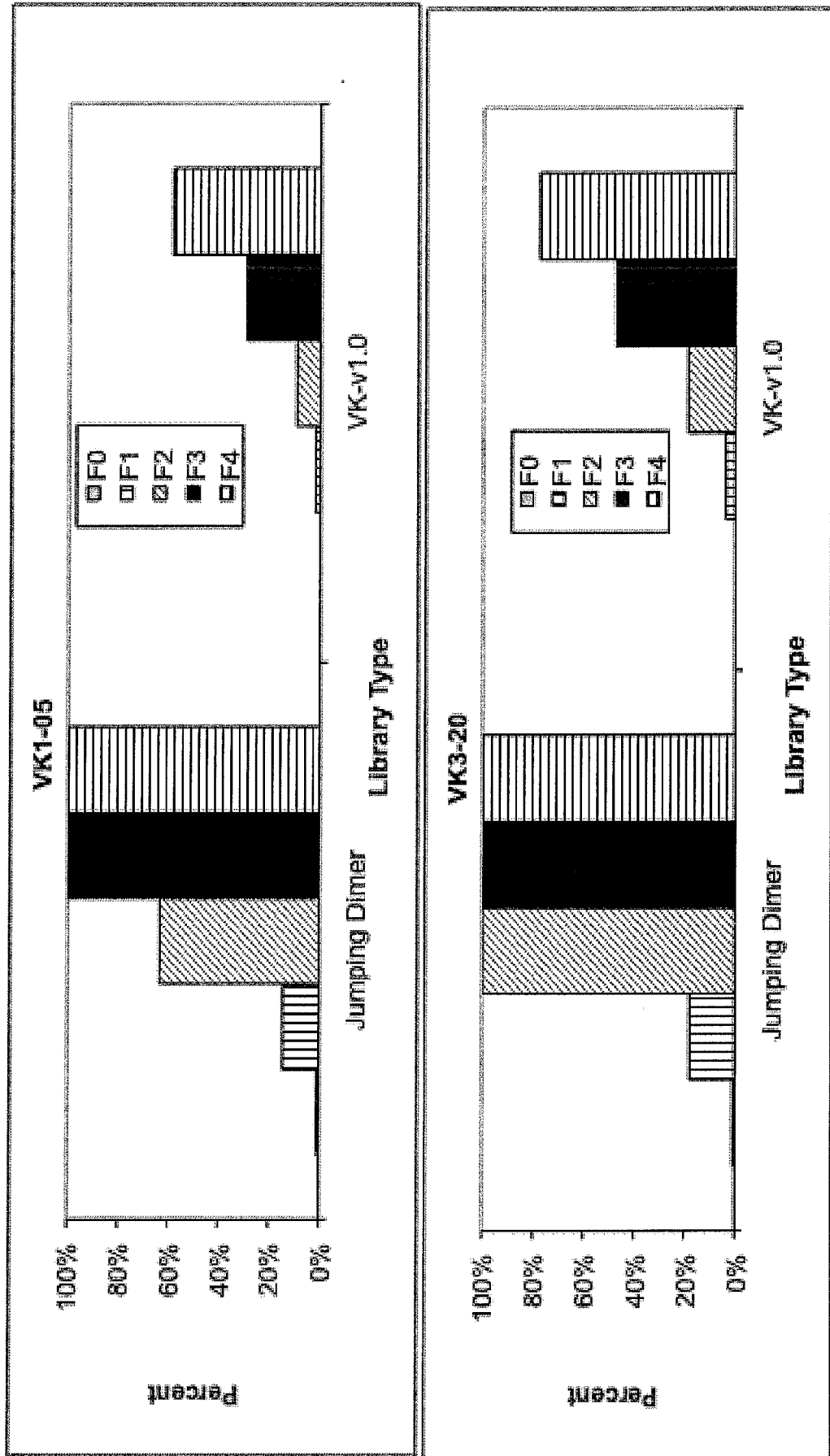


2/15

FIGURE 2

3/15

FIGURE 3



4/15

FIGURE 4

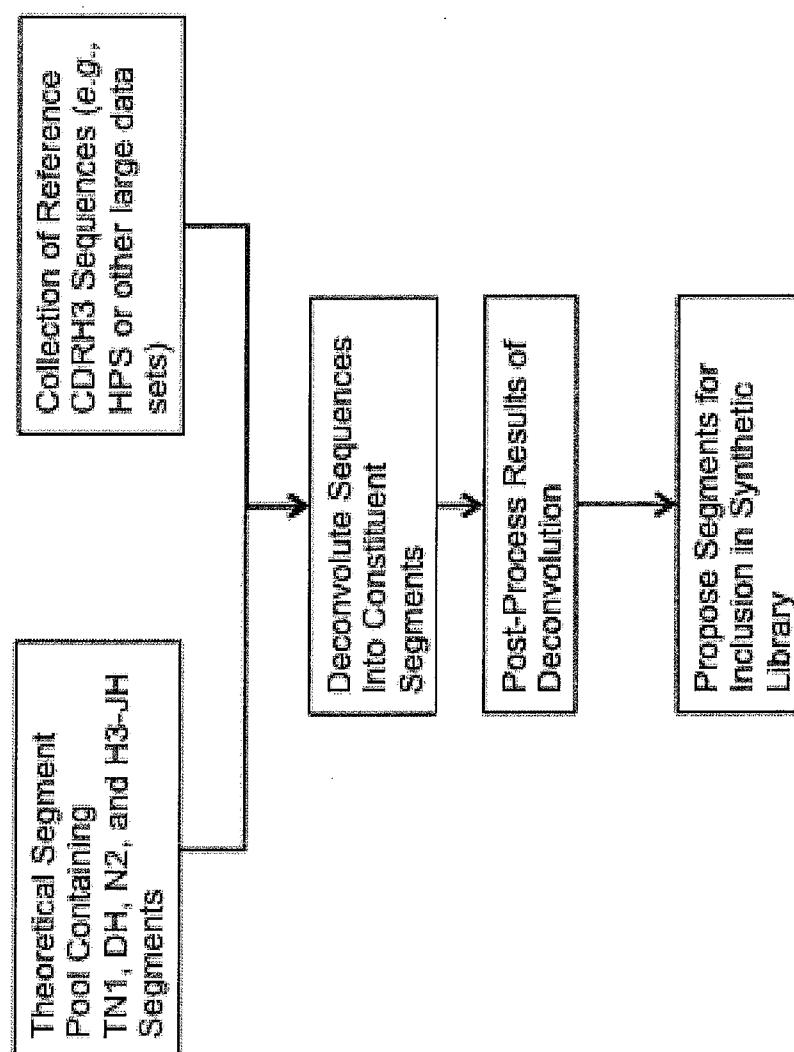
IGHJ3-02	-	ATGCTTTTGATATCT...TCCTCAG
IGHJ3-02-1	Mutate 5' doublet	NNTGCTTTTGATATCT...TCCTCAG
IGHJ3-02-2	Delete 5' codon	GCTTTTGATATCT...TCCTCAG
IGHJ3-02-3	Mutate 5' doublet	NNTTTGATATCT...TCCTCAG
IGHJ3-02-4	Delete 5' codon	TTTGATATCT...TCCTCAG
IGHJ3-02-5	Mutate 5' doublet	NNTGATATCT...TCCTCAG

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

IGHJ2-01	-	CTACTGGTACTTCGA...TCCTCAG
IGHJ2-01-1	Mutate 5' doublet	NNACTGGTACTTCGA...TCCTCAG
IGHJ2-01-2	Delete 5' codon	CTGGTACTTCGA...TCCTCAG
IGHJ2-01-3	Mutate 5' doublet	NNGGTACTTCGA...TCCTCAG
IGHJ2-01-4	Delete 5' codon	GTACTTCGA...TCCTCAG
IGHJ2-01-5	Mutate 5' doublet	NNACTTCGA...TCCTCAG

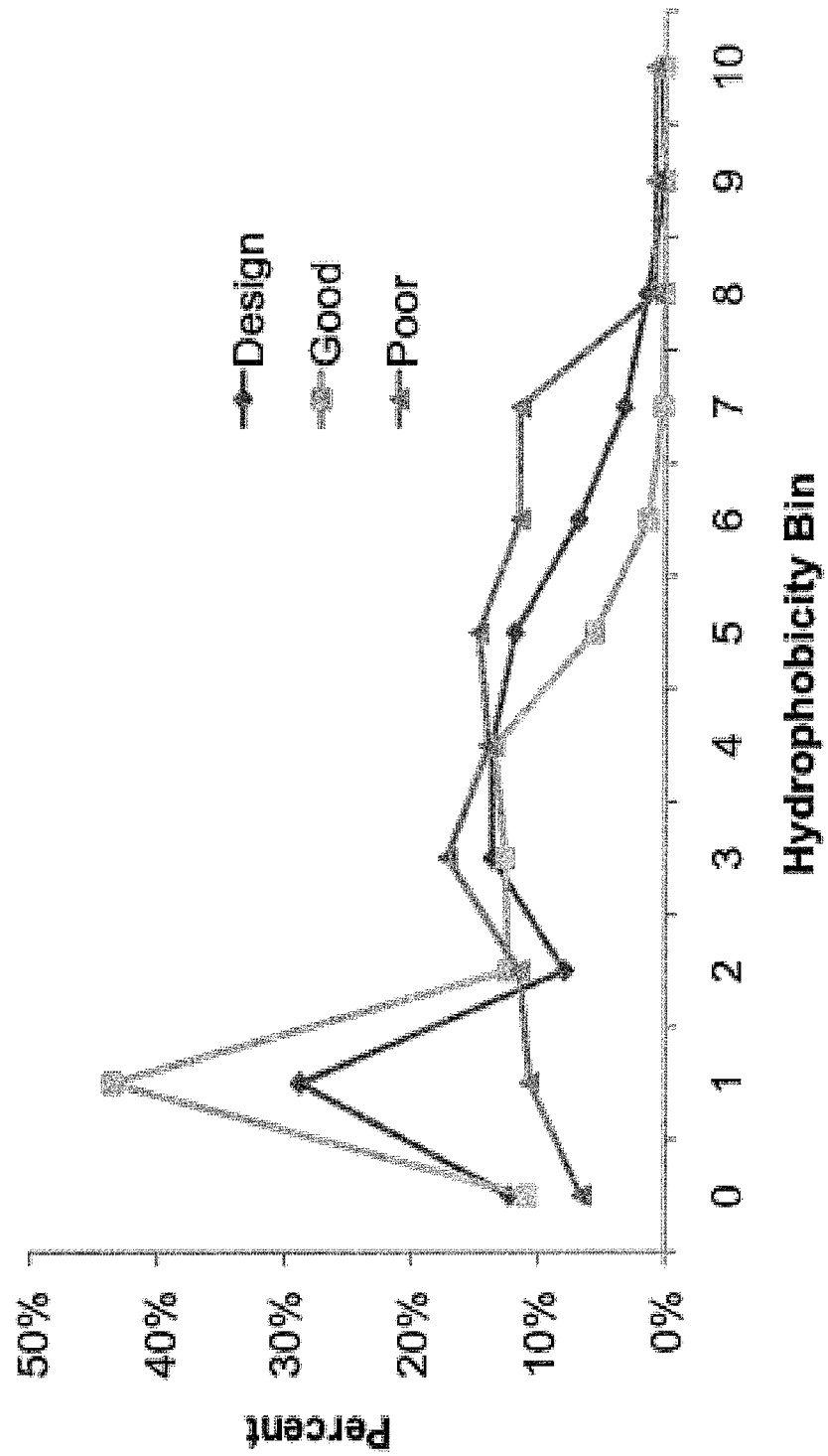
5/15

FIGURE 5



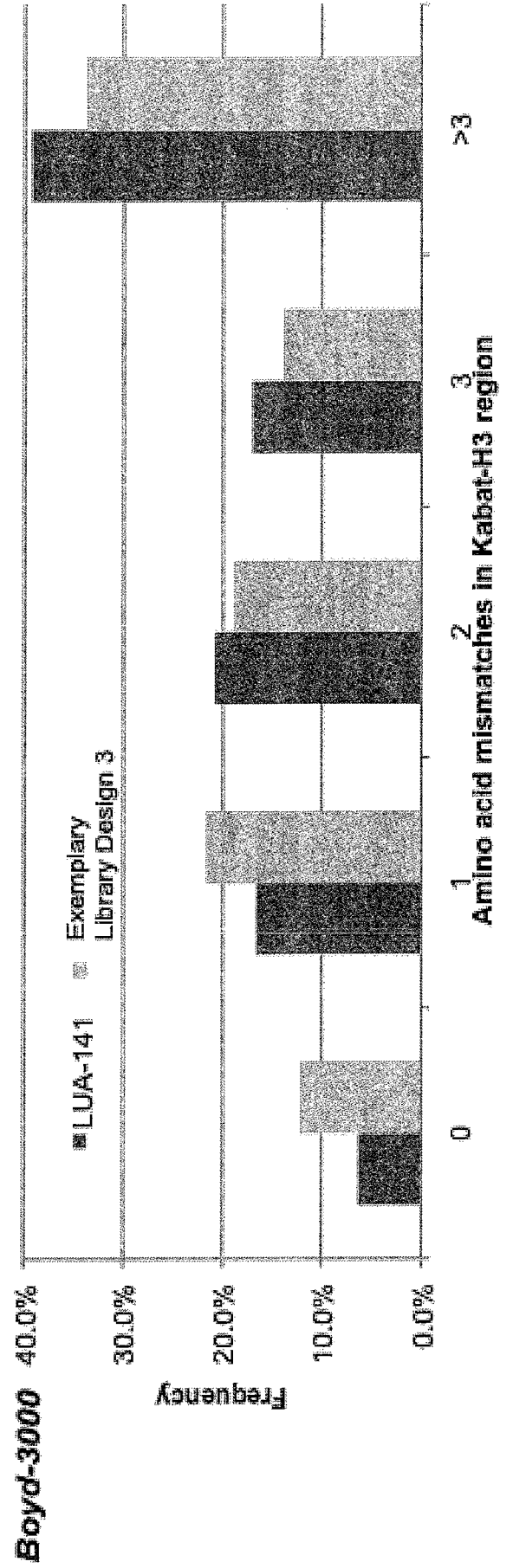
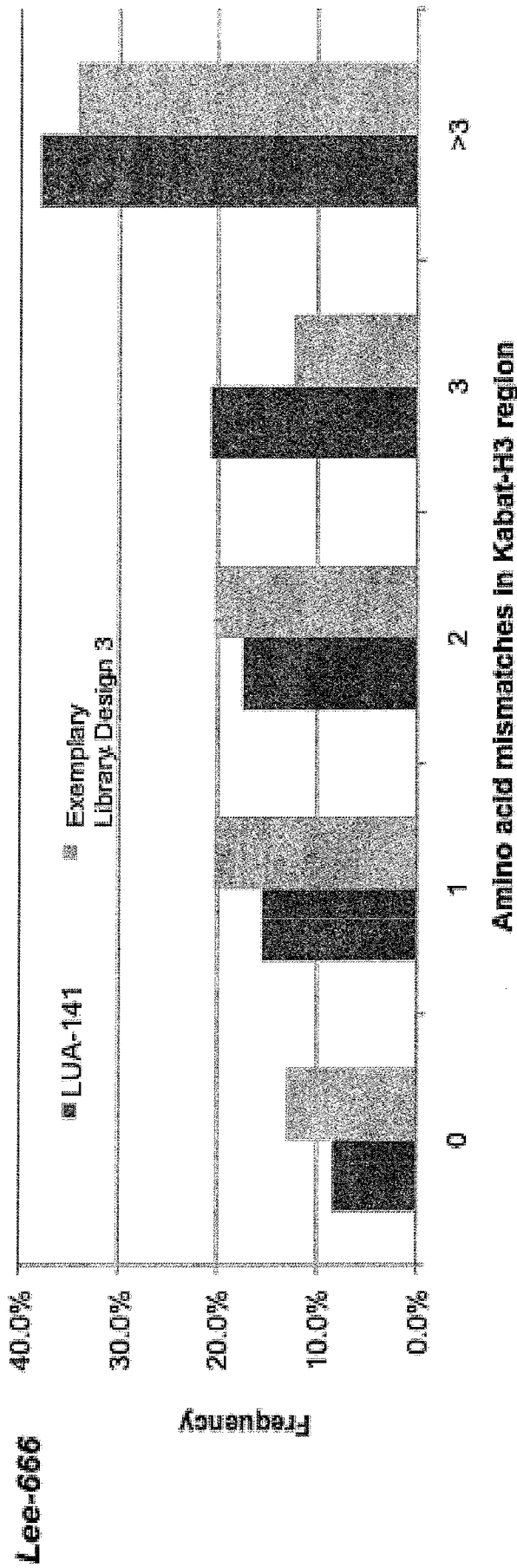
6/15

FIGURE 6



7/15

FIGURE 7



8/15

FIGURE 8

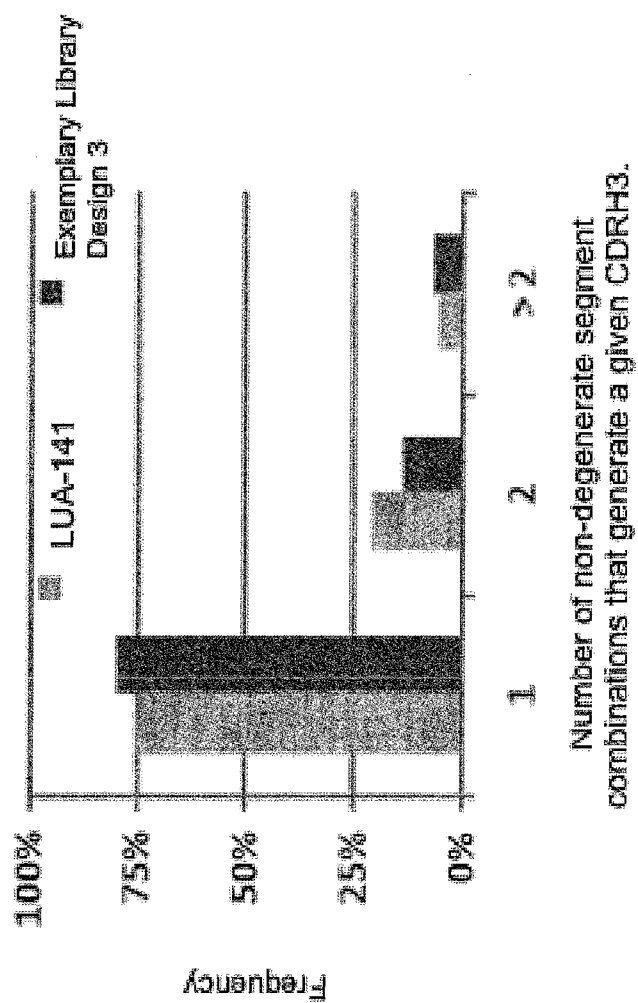
Based on a set of CDRH3 sequences derived from 55
clinically relevant human mAbs

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

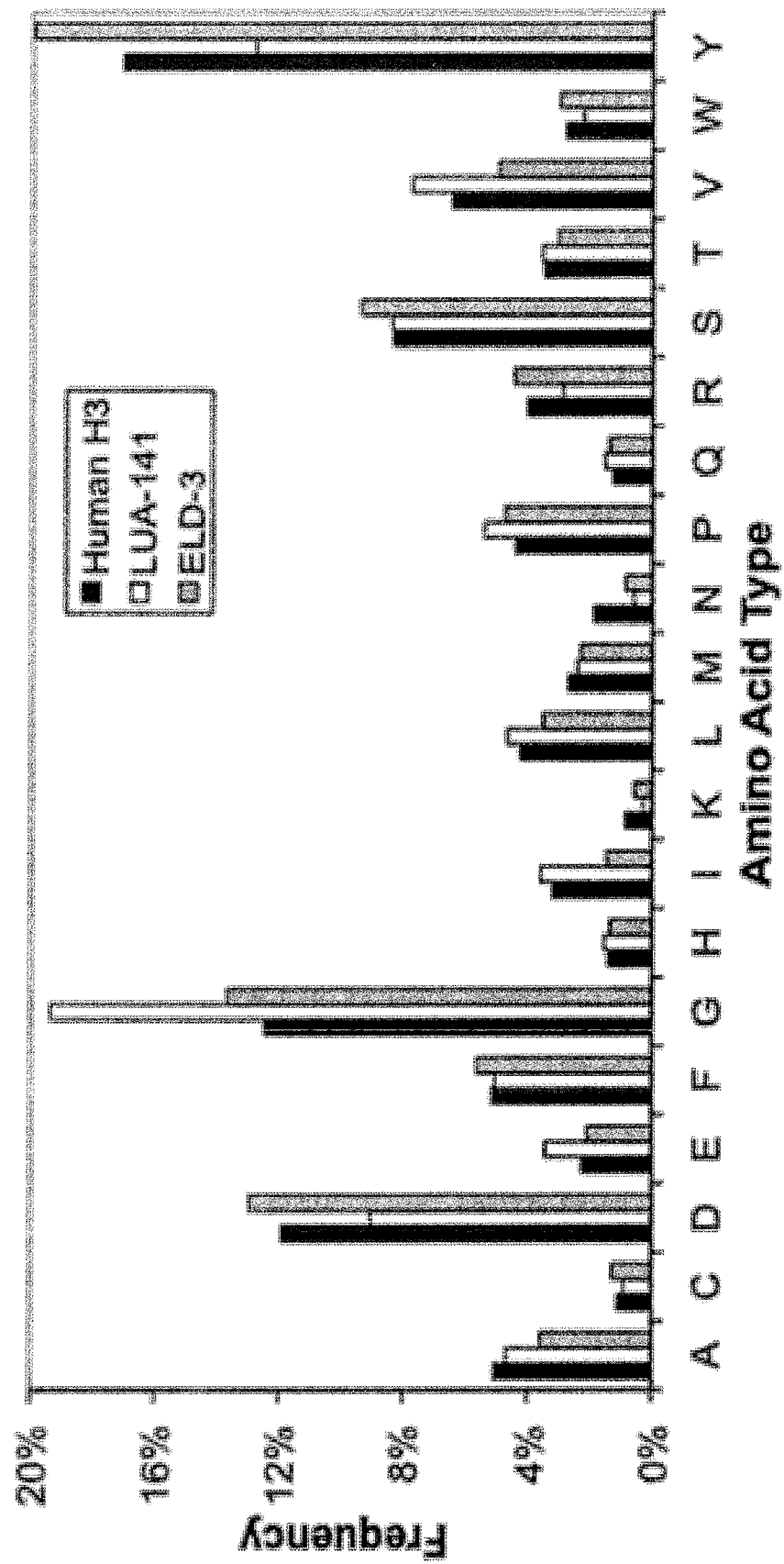
FIGURE 9

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



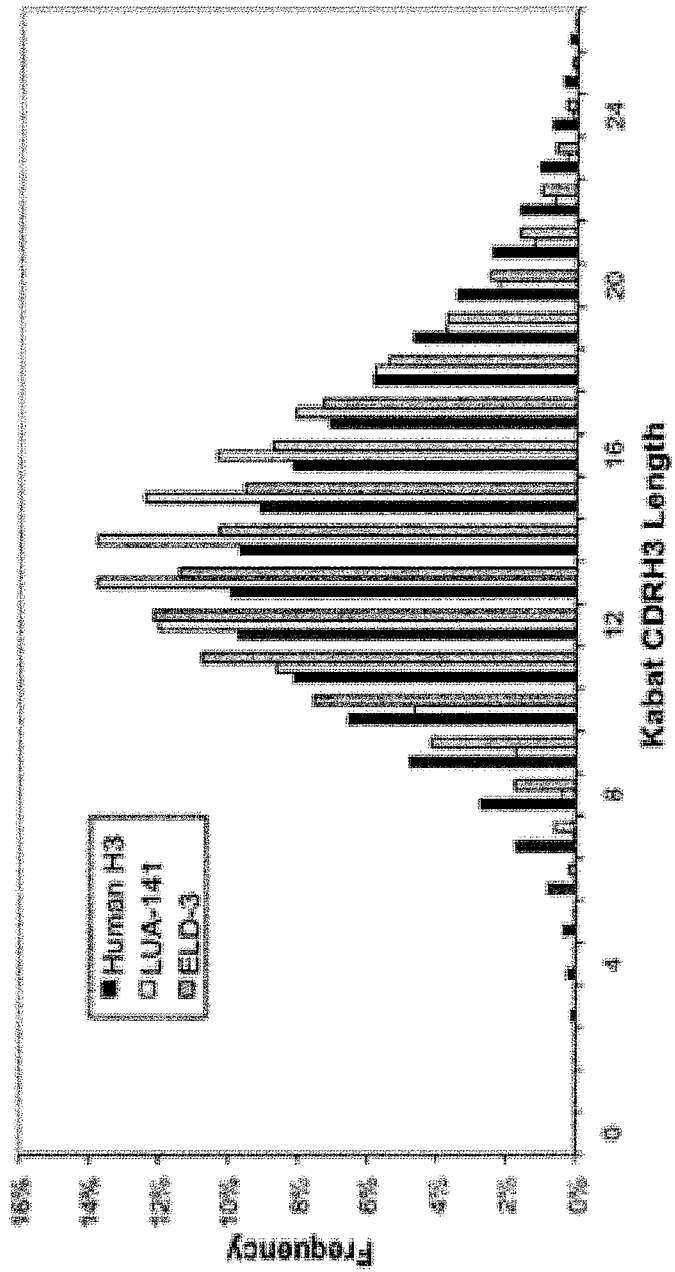
10/15

FIGURE 10



11/15

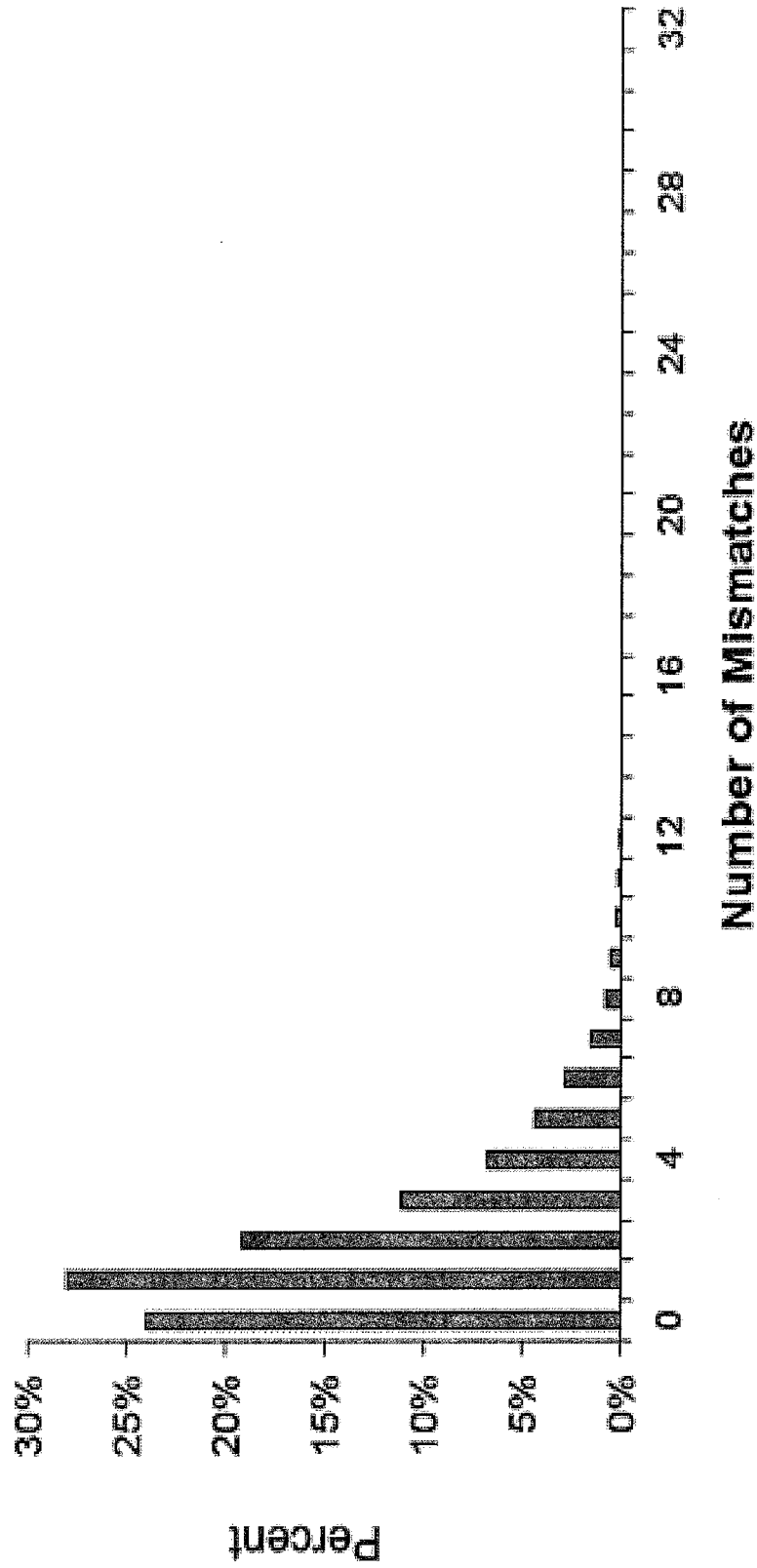
FIGURE 11



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

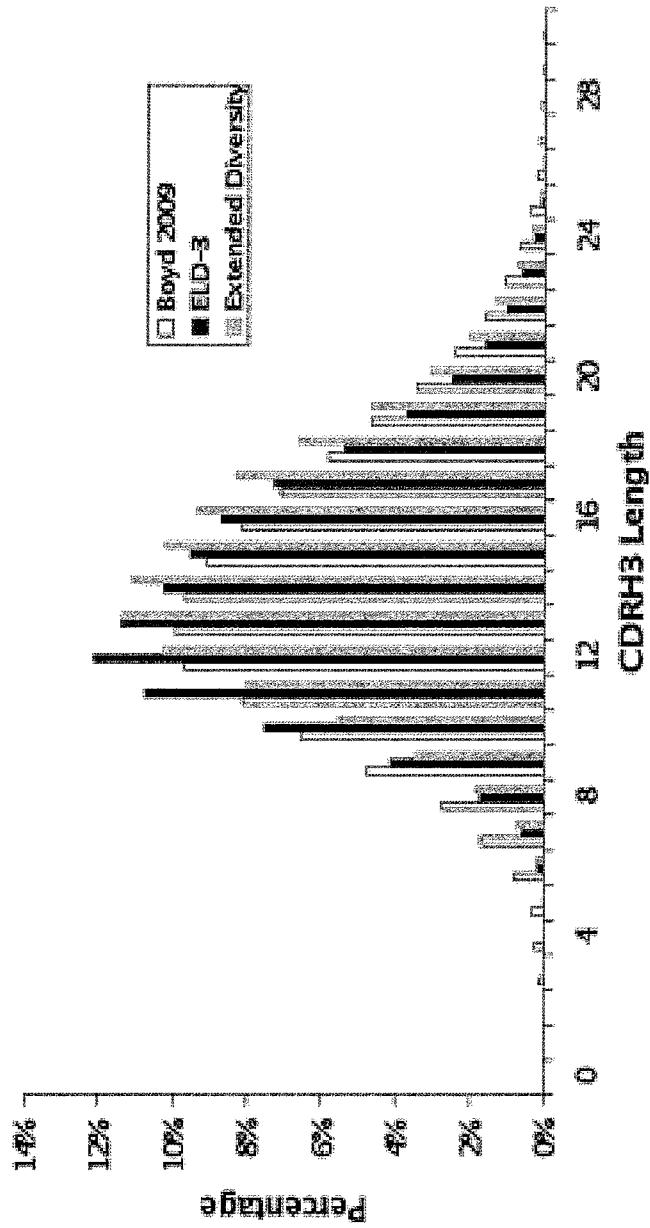
12/15

FIGURE 12



13/15

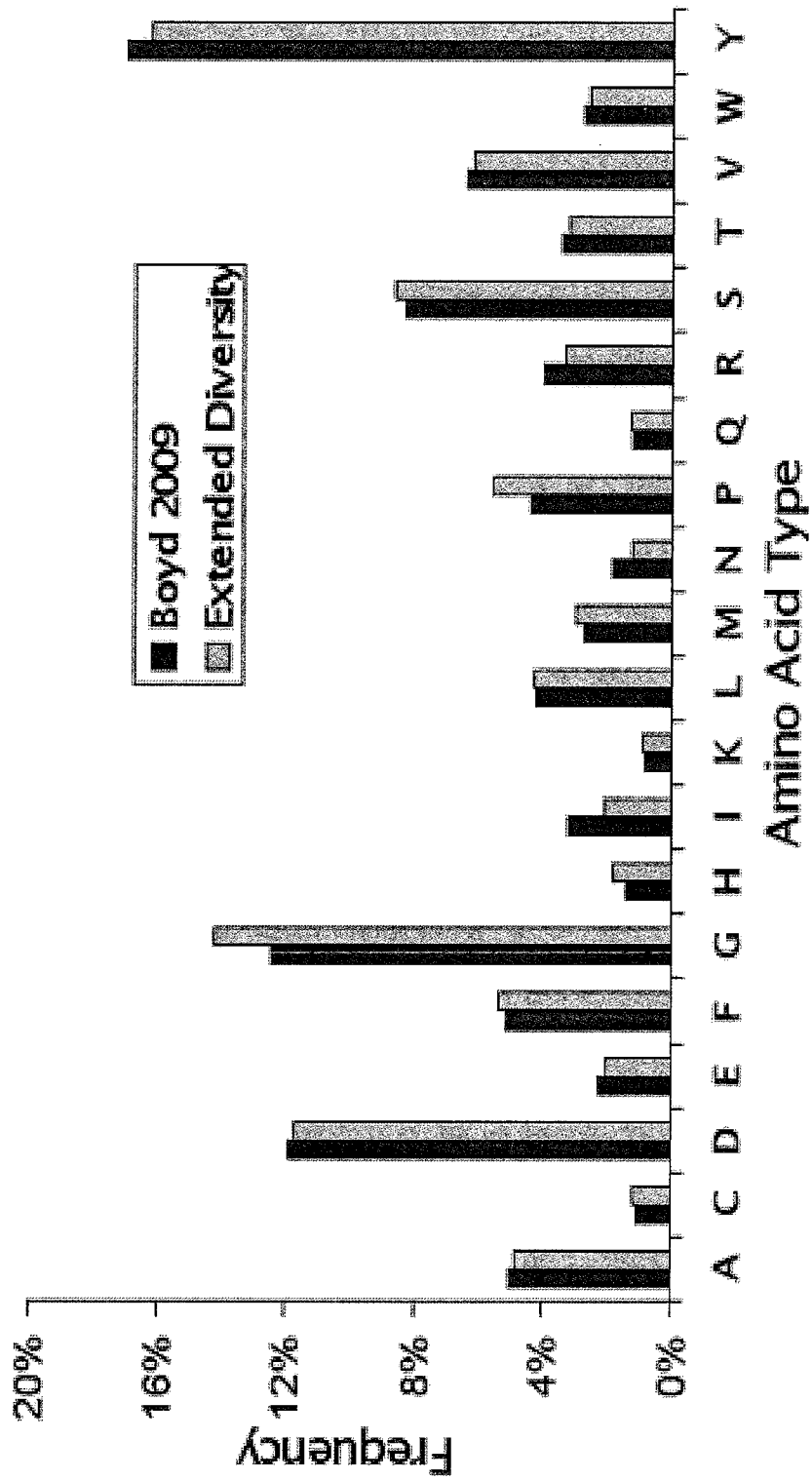
FIGURE 13



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	39	28	30

14/15

FIGURE 14



15/15

FIGURE 15

