Immunogenetics 30: 475–493, 1989

Immuno-genetics

Mouse V_k gene classification by nucleic acid sequence similarity

Robert Strohal, Arno Helmberg, Guido Kroemer, and Reinhard Kofler

Institute for General and Experimental Pathology, University of Innsbruck Medical School, Innsbruck, Austria

Abstract. Analyses of immunoglobulin (*Ig*) variable (*V*) region gene usage in the immune response, estimates of *V* gene germline complexity, and other nucleic acid hybridization-based studies depend on the extent to which such genes are related (i. e., sequence similarity) and their organization in gene families. While mouse *Igh* heavy chain *V* region (V_H) gene families are relatively well-established, a corresponding systematic classification of *Igk* light chain *V* region (V_k) genes has not been reported. The present analysis, in the course of which we reviewed the known extent of the V_k germline gene repertoire and V_k gene usage in a variety of responses to foreign and self antigens, provides a classification of mouse V_k genes in gene families composed of members with $>80\%$ overall nucleic acid sequence similarity. This classification differed in several aspects from that of V_H genes: only some V_k gene families were as clearly separated (by $>25\%$ sequence dissimilarity) as typical V_H gene families; most V_k gene families were closely related and, in several instances, members from different families were very similar ($>80\%$) over large sequence portions; frequently, classification by nucleic acid sequence similarity diverged from existing classifications based on amino-terminal protein sequence similarity. Our data have implications for V_k gene analyses by nucleic acid hybridization and describe potentially important differences in sequence organization between V_H and V_k genes.

Introduction

The ability of the immune system to recognize virtually any antigen is mediated by the enormous sequence variability in the amino-terminal region of immunoglobulin (Ig) heavy and light chains. Among other mechanisms, this diversity is generated by somatic juxtaposition of gene segments that are separated in the germline, termed variable (*V*), diversity (heavy chain only), and joining (*J*) gene segments (reviewed by Tonegawa 1983, Alt et al. 1986). *V* genes contribute all residues of the first and second complementarity determining region (CDR) of both heavy and light chains, as well as part of the light chain CDR-3, and hence contribute the majority of antigen contact residues (Kabat et al. 1987). In mice, several hundred V_H and V_k (over 90% of all serum Ig is of the Igk isotype) gene segments exist in the germ line (Brodeur and Riblet 1984, Livant et al. 1986, Cory et al. 1981, Kofler et al. 1989). These genes can be very similar or may differ by over 40% nucleotides, and *V* region classifications based on nucleic and/or amino acid sequence similarity have been proposed (Brodeur and Riblet 1984, Dildrop 1984, Potter et al. 1982). Thus, mouse V_H genes have been grouped in 11 V_H gene families in which members generally share $>80\%$ of their nucleic acid sequence within, and <70–75% between, families (Brodeur and Riblet 1984, Winter et al. 1985, Kofler 1988, Reininger et al. 1988). Individual members of a given family cross-hybridize in nucleic acid hybridization assays only with members of their own family. These V_H gene families correspond well with a V_H region classification based on similarities at the protein level (Dildrop 1984). Understanding V_H gene relatedness on the nucleic acid sequence level has greatly facilitated studies regarding the expression of different V_H gene families during ontogeny (Yancopoulos et al. 1984, Perlmutter et al. 1985) and in response to foreign and self antigens (Manser et al. 1987b, Kofler et al. 1987a). These studies have thus provided an important insight into B-cell repertoire generation.

V_k classifications reported to date are confined to the protein level. One attempt to systematically classify V_k proteins was based on the partial amino acid sequence up to the invariant cysteine in position 23 (Cys23), leading to 26 V_k subgroups, designated V_kCys (Potter 1977). A

Address correspondence and offprint requests to: Dr. Reinhard Kofler, Institute for General and Experimental Pathology, Fritz-Pregl-Straße 3, A-6020 Innsbruck, Austria.

modified classification, based on the length and similarity of the amino termini up to the invariant tryptophan 35 (Trp35) of 79 V_k proteins, was introduced in 1982 (V_kTrp subgroups; Potter et al. 1982). Four of the V_kCys subgroups were condensed and two new groups were added, resulting in a total of 24 V_k subgroups, six of which are still defined only by sequences up to Cys23. This classification has now been generally accepted and, although an extended comparison at the nucleic acid level has never been reported, the corresponding V_k protein subgroups have been widely used synonymously with V_k gene families. More recently, we have performed a detailed restriction fragment length polymorphism (RFLP) analysis with DNA probes corresponding to 16 V_k protein subgroups, and obtained evidence that such protein groups may not necessarily correspond to gene families analogous to those described for V_H genes (Kofler et al. 1989). Since a large number of full-length V_k nucleic acid sequences has been reported, it is now possible to address, by direct sequence comparison, the matter of whether V_k genes can be organized into gene families, as has been accomplished with V_H genes, and how such V_k gene families relate to the existing V_k protein groups. This issue is of considerable interest for V_k gene usage determinations, repertoire estimates, genomic mapping, and similar studies using nucleic acid hybridization, since such procedures depend on relatedness between V_k groups, gene families, and corresponding DNA probes.

We compiled 248 full-length V_k nucleic acid sequences from the literature and several databases, and assigned them to existing V_k protein classifications with subsequent grouping into gene families comprised of members with >80% overall nucleic acid sequence similarity. Our analysis revealed that the current classification in V_k protein groups or subgroups frequently did not reflect relatedness on the nucleic acid sequence level. Furthermore, V_k gene family organization differed in important aspects from that of V_H gene families; only some of the V_k gene families were clearly separated by sequence dissimilarity of >25%, as is usually observed in V_H gene families. The remaining families were more similar to each other and, in several instances, large portions of genes from different families shared >80% of their sequences, leading to cross-hybridization between those families in hybridization analyses. In addition, although ancillary to the primary aim of this study, we reviewed the specificities encoded by the various V_k gene families and estimated their germline gene complexity.

Methods and nomenclature

V_k nucleic acid sequence bank. A database was constructed consisting of V_k nucleic acid sequences from the Genetic Sequence Data Bank (GenBank, Los Alamos, New Mexico), E. A. Kabat's collection (Kabat et al. 1987), and other publications. Only sequences encoding the entire mature V_k protein were included in the database. If applicable, sequence portions encoding untranslated region, leader sequence, introns, or *J* segments were removed prior to comparisons. This primary database of 248 full-length V_k sequences was then condensed to a final database of 109 (Fig. 1) by deleting duplicate sequences and those differing by only 1 to 4 base pairs (bp).

V_k protein groups and subgroups. All nucleic acid sequences were translated into amino acids and organized into V_k protein groups and subgroups. Assignment to V_k protein groups (labeled I to VII) was based on the length of the amino-terminal sequence up to the invariant Trp35 (41, 40, 39, 36, 35, 34, and 33 residues, respectively; Kabat et al. 1987). Organization into V_k protein subgroups was based on < 13 substitutions up to Trp35 (V_kTrp subgroups; Potter et al. 1982). Sequences meeting assignment criteria for more than one subgroup were assigned to the subgroup with the best match.

V_k gene families. Analogous to V_H gene families, we defined a "V_k gene family" as a group of nucleic acid sequences that exhibit >80% overall sequence similarity with every member of this family, and <80% with V_k genes from other families. In nucleic acid hybridization analyses under defined stringency conditions (Brodeur and Riblet 1984), all members of a gene family can be expected to cross-hybridize with each other. The V_k gene family nomenclature proposed in this study was adjusted as far as possible to that used for V_k protein subgroups, in order to minimize confusion in the literature; when V_k protein subgroups and V_k gene familes corresponded to each other (e.g., V_k21), the V_k subgroup designation was used for the V_k gene family as well. V_k gene families comprising two or more V_k protein subgroups were given the designation of the respective subgroups (e.g., the *$V_k4/5$* gene family comprised V_k4 and V_k5 protein subgroups). Addition of capital letters to the designation indicates that a V_k protein subgroup included members from two distinct V_k gene families (e.g., the V_k9 protein subgroup comprised members from two distinct V_k gene families, termed *V_k9A* and *V_k9B*, respectively). *V_kRF* and (tentatively) *V_k38C* were two new gene families that could not be related unambiguously to any V_k protein subgroup and, hence, were named after a prototypic sequence.

Organization of mouse V_k sequences on the protein and nucleic acid level

The major goal of this study was to investigate the organization of mouse V_k genes in terms of nucleic acid sequence similarity, and to determine the relationship of such organization to existing V_k protein classifications. To this end, we first compiled 109 distinct (i.e., >4 bp different), full-length V_k nucleic acid sequences that were used as a database for subsequent analyses (Fig. 1). The sequences were translated into amino acids (Fig. 2) and assigned to protein groups and subgroups (Table 1).

Classification into protein groups was based on the number of residues up to the invariant Trp35 and, hence, was unambiguous in all instances. However, this classification was of limited practical value, since it frequently did not reflect structural relatedness (i.e., sequence similarity) between V_k sequences. For example, group V included members of several, sometimes quite dissimilar, V_k gene families (*V_k23, $V_k12/13$, V_kRF, V_k11, V_k9A, V_k9B, V_k10, V_k38C, $V_k19/28$*). On the other hand,

```
A   |--------------------------------FR1--------------------------------| |--------------CDR1----------------
            10        20        30        40        50        60          70        80        90      100
001 AACATTGTGCTGACCCAATCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGGGCCACCATATCCTGC AGAGCCAGTGAAAGTGTTGATAGT......TAT
002 GACATTGTGCTGACCCAATCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGGGCCACCATATCCTGC AGAGCCAGTGAAAGTGTTGATAGT......TAT
003 GACATTGTGCTGACACAGTCTCCTGCTTCCTTAGCTGTATCTCTGGGGCAGAGGGCCACCATCTCATGC AGGGCCAGCAAAAGTGTCAGTACA......TCT
004 GACATTGTGCTAACACAGTCTCCTGCTTCCTTAGCTGTATCTCTGGGGCAGAGGGCCACCATCTCATGC AGGGCCAGCCAAAGTGTCAGTACA......TCT
005 GACATTGTGCTGACCCAATCTCCAGCTTCTTTGGCTGTGTCTCTAGGACAGAGAGCCACTATCTTCTGC AGAGCCAGCCAGAGTGTCGATTAT......AAT
006 GACATTGTGCTGACCCAATCTCCAGGATCTTTGGCTGTGTCTCTAGGGCAGAGGGCCACCATATCCTGC AGAGCCAGTGAAAGTGTTGAAAGT......TCT
007 AAAATTGTGCTGACCCAATTTCCAGCTTCTTTGGCTGTGTCTCTAAGGCAGAGGGCCACCATATCCTGC AGAGCCAGTGAAAGTGTTGATAGT......TAT
008 GACATTGTGCTCACCCAATCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGTGTCACCATCTCCTGC AGAGCCAGTGAAAGTGTTGAATAT......TAT
009 GACATTGTGCTGACACAGTTTCCTGCTTCCTTAGCTGTATCTCTGGGGCAGAGGGCCACCATCTCATAC AGGGCCAGCAAAAGTGTCAGTACA......TCT
010 GACATCTTGCTGACTCAGTCTCCAGCCATCCTGTCTGTGAGTCCAGGAGAAAGAGTCAGTTTCTCCTGC AGGGCCAGTCAG..................AGC
011 GACATCTTGCTGACTCAGTCTCCAGCCATCCTGTCTGTGAGTCCAGGAGAAAGAGTCAGTTTCTCCTGC AGGGCCAGTCAG..................AGC
012 GACATCTTGCTGACTCAGTCTCCAGCCATCCTGTCTGTGAGTCCAGGAGAAAGAGTCAGTTTCTCCTGC AGGGCCAGTCAG..................AGC
013 GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGATAGCGTCAGTCTTTCCTGC AGGGCCAGCCAA..................AGT
014 GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGATAGCGTCAGTCTTTCCTGC AGGGCCAGCCAA..................AGT
015 GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGATAGCGTCAGTCTTTCCTGC AGGGCCAGCCAA..................AGT
016 GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAAGAGATAGCGTCAGTCTTTCCTGC AGGGCCAGCCAA..................AGT
017 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCTGGGGAGAAGGTCACCATGACCTGC AGTGCCAGATCAAGT...............GTA
018 GAAAATGTGCTGACCCAGTCTCCAGCAATCATGGCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCAAGT...............GTA
019 GAAAATGTGCTCACCCAGTCTCCAGCAATAATGGCTGCCTCTCTGGGGCAGAAGGTCACCATGACCTGC AGTGCCAGCTCAAGT...............GTA
020 GAAAATGTGCTCACCCAGTCTCCAGCAATAATGGCTGCCTCTCTGGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCAAGT...............GTA
021 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCCTCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCAAGT...............GTA
022 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCTGGGGAACGGGTCACCATGACCTGC AGTGCCAGCTCAAGT...............GTA
023 GAAATTGTGCTCACCCAGTCTCCAACCACCATGGCTNNATCTCCCGGGGAGAAGATCACTATCACCTGC AGTGCCAACTCAAGT...............ATA
024 GAAAATGTGCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGNAAAAGGTCACCATGACCTGT AGGGCCAGCTCAAGT...............GTA
025 GAAATTGTGCTCACCCAGTCTCCAACCACCATGGCTGCATCTCCCGGGGAGAAGATCACTATCACCTGC AGTGCCAGCTCAAGT...............ATA
026 GAAATTGTGCTCACCCAGTCTCCAGCACTCATGGCTGCATCTCCAGGGGAGAAGGTCACCATTACCTGC AGTGTCAGCTCAAGT...............ATA
027 CAAATTGTTCTCACCCAGTCTCCAGCATTCATGTCTGCATCTCTAGGGGAACGGGTCACCATGACCTGC ACTGCCAGGTCAAGT...............GTA
028 GAAATTTTGCTCACCCAGTCTCCAGCAATCATAGCTGCATCTCCTGGGGAGAAGGTCACCATCACCTGC AGTGCCAGCTCA.....................
029 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
030 GAAAATGTGCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCTAGGGGAGAAGGTCACCATGAGCTGC AGGGCCAGCTCA.....................
031 GGAATTGTGCTCACCCAATCTCCAACAACCATGACTGCATTTCCAGGGGAGAATGTCACCATCACCTGC AGTGCCAGCTCA.....................
032 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
033 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATATCCTGC AGTGCCAGCTCA.....................
034 GAAATTGTGCTCACTCAGTCTCCAGCCATCACAGCTGCATCTCTGGGGCAAAAGGTCACCATCACCTGC AGTGCCAGCTCA.....................
035 CAAATTGTTCTCACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
036 CAAATTCTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
037 CAAATTGTTCTCACCCAGTCTCCAGCACTCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
038 CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACATTGACTTGC AGGGCCAGCTCA.....................
039 CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGACTTGC AGGGCCAGCTCA.....................
040 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGCAGAAAGTCACCATAACCTGC AGTGCCAGCTCA.....................
041 CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGATTTGC AGGGCCAGCTCA.....................
042 CAAATTCTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGCCAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
043 CAAATTCTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGCCAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
044 CAAAATGTTCTCAACCAGTCTCCAGNAATCATGTCTGNATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
045 CAAATTGTTCTCACCCAGTCTCCAGNAATCATGTCTGNATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGTTCA.....................
046 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
047 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
048 CAAATTGTTCTCATACAGTCTCCAGNAATCATGTCTGCATCTCCAGGGGNGAAGGNCACCATGACCTGC AGTGCCAGCTCA.....................
049 CAAATTGTTCTCACCCAGTCTCCACCAATCATGTCTGCCTCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
050 CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGTTCA.....................
051 CAAGTTGTTCTCACCCAGTCTCCAGNAATCATGTCTGCATCTCCAGGGNAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
052 CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAGGGTCACATTGACTTGC AGGGCCAGCTCA.....................
053 CAAATTGTTCTCACCCAGTCTCCAGNAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC AGTGCCAGCTCA.....................
054 CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGCAGGTCACAATGACTTGC AGGGCCAGCTCA.....................
```

B	CDR1	FR2	CDR2	FR3
	110 120	130 140 150 160	170 180	190 200
001	GGCAATAGTTTTATGCAC	TGGTACCAGCAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	CTTGCATCCAACCTAGAATCT	GGGGTCCCTGCCAGG
002	GGCAATAGTTTTATGCAC	TGGTACCAGCAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	CGTGCATCCAACCTAGAATCT	GGGATCCCTGCCAGG
003	GGCTATAGTTATATGCAC	TGGTACCAACAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	CTTGCATCCAACCTAGAATCT	GGGGTCCCTGCCAGG
004	AGCTATAGTTATATGCAC	TGGTACCAACAGAAACCAGGACAGCCACCCAAACTCCTCATCAAG	TATGCATCCAACCTAGAATCT	GGGGTCCCTGCCAGG
005	GGAATTAGTTATATGCAC	TGGTTCCAACAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	GCTGCATCCAACCTAGAATCT	GGGATCCCTGCCAGG
006	GGCAATAATTTTATCCAC	TGGCACCAGCAGAAACCAGGACAGCCACCCNAACTCCTCATCTAT	CGTGCATCCAACCTAGCATCT	GGGATCCCTGCCAGG
007	GGCAATAGTTTTATGTAC	TGGTACCAGCAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	CGTGCATCCAACCTAGAATCT	GGGGTCCCTGCCAGG
008	GGCAGTAGTTTAATGCAG	TGGTACCAACAGAAACCAGGACAGCCACCCAAACTCCTCATCTAT	GGTGCATCCAACGTAGAATCT	GGGGTCCCTGCCAGG
009	GGCTATAGTTATATGCAC	TGGAACCAACAGAAACCAGGACAGCCACCCAGACTCCTCATCTAT	CTTGTATCCAACCTAGAATCT	GGGGTCCCTGCCAGG
010	ATTGGCACAAGCATACAC	TGGTATCAGCAAAGAACAAATGGTTCTCCAAGGCTTCTCATAAAG	TATGCTTCTGAGTCTATCTCT	GGGATCCCTTCCAGG
011	ATTGGCACAAGCATACAC	TGGTATCAGCAAAGAACAAATGGTTCTCCAAGGCTTCTCATAAAG	AATGCTTCTGAGTCCATCTCT	GGGATCCCTTCCAGG
012	ATTGGCACAAGTCTTCAC	TGGTATCAACAAAGAACAAATGGTTCTCCAAGGCTTCTCATAAAG	TATGCTTCTGAGTCTATCTCT	GGGATCCCTTCCAGG
013	ATTATCAACAACCTACAC	TTATATCGATAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAA	TATGCTTCCCAGTCCATCTCT	GGGATCCCCTCTAGG
014	ATTAGCAACAACCTACAC	TGGTATCAACAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAT	TATGCCTCCCAGTCCATGTCT	GGGATCCCCTCCAGG
015	ATTAGCAACAACCTACAC	TGGTATCAACAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAG	TATGCTTCCCAGTCCATCTCT	GGGATCCCCTCCAGG
016	ATTAGCAACAACCTACAC	TGGTATCAACAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAA	TATGCTTCCCAGTCCATCTCT	GGGATCCCCTCTAGG
017	AGTTCCAGCTACTTGTAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACTCTGGATTTAT	AGCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
018	AGTTCTAGTAACTTGCAC	TGGTACCAGCAGAAGTCAGGCACTTCTACCAAATTCTGGATTTAT	AGGACATCCAACCTGGCTTCA	GAAGTCCCAGCTCCC
019	AGTTCCAGTTACTTGCAC	TGGTACCAGCAGAAGTCAGGCGCTTCCCCCAAACCCTTGATTCAT	AGGACATCCAACCTGGCTTCT	GGAGTCCCAGCTCGC
020	AGTTCCAGCTACTTGCAC	TGGTACCAGCAGAAGTCAGGCACTTCCCCCAAACTCTGGATTTAT	GGCACATCCAACCTGGCTTCT	GGAGTCCCAGCTCGC
021	AGTTCCAAATACTTGAAC	TGGTACCAGCAGAGGTCAGGAGCCTCCCCCAAACTCTGGATTTAT	GGCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
022	AGTTCCAGCTACTTGTAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACTATGGATTTAT	AGCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
023	AGTTCCAATTACTTGCAT	TGGTATCAGCAGAAGCCAGGATTCTCCCCTAAACTCTTGATTTAT	AGGACATCCAATCTAGCTTCT	GGAGTCCAAGCTCGC
024	AGTTCCAGTTACTTGCAC	TGGTACCAGCAGAAGTCAGGTGCCTCCCCCAAACTCTGGATTTAT	AGCACATCCAACCTGGCTTCT	GGNGTCCCTGCTCGC
025	AGTTCCAATTACTTGAAT	TGGTTTCAGCAGAAGCCAGGATTCTCCCCTAAACTCTTGATTTAT	AGGACATCCAATCTGGCTTCT	GGAGTCCCAGATCGC
026	AGTTCCAGCAACTTGCAC	TGGTACCAGCAGAAGTCAGAAACCTCCCCCAAATCTTGGATTTAT	GGCACATCCAACCTGGCTTCT	GGAGTCCCTGTTCGC
027	AGTTCCAGTTACTTCCAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACTCTGGATTTAT	AGCACATCCAACCTGGCTTCT	GGAGTCCCAACTCGC
028	AGTGTAAGTTACATGAAC	TGGTACCAGCAGAAACCAGGATCCTCCCCCAAAATATGGATTTAT	GGTATATCCAACCTGGCTTCT	GGAGTTCCTGCTCGC
029	AGTATAAGTTACATGCAC	TGGTACCAGCAGAAGCCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGGCTTCT	GGAGTCCCTGCTCGC
030	AGTGTAAATTACATGTAC	TGGTACCAGCAGAAGTCAGATGCCTCCCCCAAACTATGGATTTAT	TACACATCCAACCTGGCTCCT	GGAGTCCCAGCTCGC
031	AGTATAAATTACATTCAC	TGGTACCAGCAGAAGTCAGGAAATACCCCCAAACAATGAATTTAT	AAGACATCCGACCTGCCTTCT	GGAGTCCCAACTCTC
032	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGGCTTCT	GGAGTCCCTGCTCGC
033	AGTGTAAGTTACATGTAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTAT	CGCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
034	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGTCAGGCACCTCCCCCAAACCATGGATTTAT	GAAATATCCAAACTGGCTTCT	GGAGTCCCAGCTCGC
035	AGTGTAAGTTACATGTAA	TGGTTCCAGCAGAAGCCAGGATCCTCCCCCAAACTCTGGATTTAT	AGCATATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
036	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGCCAGGATCCTCGCCCAAACCCTGGATTTAT	GACACATCCAACCTGGCTTCT	GGATTCCCTGCTCGC
037	AGTGTAAGTTACATGTAC	TGGTACCAGCAGAAGCCAAGATCCTCCCCCAAACCCTGGATTTAT	CTCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
038	AGTGTAAGTTTCATGAAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTAT	GCCACATCCAATTTGGCTTCT	GAGTTCCCTGGTCGC
039	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGCTTGGATCCTCCCCCAAACCATGGATTTAT	GCCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
040	AGTGTAAATTACATGCAC	TGGTACCAGCAGAAGCTAGGATCCTCCCCCAAACTCTGGATTTAT	GACACATCCAAACTGGCTCCT	GGAGTCCCTGCTCGC
041	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTAT	GCCCCATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
042	AGTGTAAGTTACATGTAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAGACTCCTGATTTAT	GACACATCCAACCTGGCTTCT	GGAGTCCCTGTTCGC
043	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGGCTTCT	GGAGTCCCTGCTCGC
044	AGTGTAAGTTACATGCAG	TGGTTCCAGCAGNAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAGCTGGNTTCT	GGNGTCCCTACTCGC
045	AGTGTAAGGTACATGAAC	TGGTTCCAACAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGTCTTCT	GGAGTCCCTGCTCGC
046	AGTGTAAGTTACATGAAC	TGGTTCCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGTCTTCT	GGAGTCCCTCCTCGC
047	ATTGTAAGTTACGTGCAG	TGGTTCCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTCT	GACACATCCAAACTGCCTTCT	GGAGTCCCTGCTCGC
048	AGTGTAAGTTATATGAAC	TGGTACCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGGCTTCT	GGAGTCCCTGCTCGC
049	AGTGTAAGTTACTTGCAG	TGGTTCCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	GACACATCCAAACTGGATTCT	GGNGTCCCTGCTCGC
050	AGTGTTAGTTACATGAAC	TGGTTCCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGGTTTTT	GCCACATCCAAACTGGNTTCT	GGAGTCCCTGCTCGC
051	AGTGTAAGTTACATGCAG	TGGTTCCAGCAGAAGTCAGGCACCTCCCCCAAAAGATTGATTTTT	TACACATCCAAACTGACTTCT	GGAGTCCCTGCTCGC
052	AGTGTAAGTTACATTCAG	TGGTTCCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTCAT	GCCACATCCAAGNTGGCTTCT	GGAGTCCCTGCTCGC
053	AGTGTAAGTTTCATGCAG	TGGTACCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTAT	CACACATCCAAACTGGCTTCT	GGAGTCCCTGCTCGC
054	AGTGTAAGTTACATGCAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTAT	GCCACATCCAACNTGGCTTCT	GGAGTCCCNGCTCGC

```
C   ────────────────────────────────────────FR3────────────────────────────────────────┐ ┌─────CDR3──────────
           210       220       230       240       250       260       270       280        290       300
001 TTCAGTGGCAGTGGGTCTAGGACAGACTTCACCCTCACCATTGATCCTGTGGAGGCTGATGATGCTGCAACCTATTACTGT CAGCAAAATAATGAGGATCCT
002 TTCAGTGGCAGTGGGTCTAGGACAGACTTCACCCTCACCATTAATCCTGTGGAGGCTGATGATGTTGCAACCTATTACTGT CAGCAAAGTAATGAGGATCCT
003 TTCAGTGGCAGTGGGTCTGGGACAGACTTCACCCTCAACATCCATCCTGTGGAGGAGGAGGATGCTGCAACCTATTACTGT CAGCACAGTAGGGAGCTTCCT
004 TTCAGTGGCAGTGGGTCTGGGACAGACTTCACCCTCAACATCCATCCTGTGGAGGAGGAGGATACTGCAACATATTACTGT CAGCACAGTTGGGAGATTCCT
005 TTCAGTGGCAGTGGGTCTGGGACAGACTTCACCCTCAACATCCATCCTGTGGAGGAGGAAGATGCTGCAACCTATTACTGT CAGCAAAGTATTGAGGATCCT
006 TTCAGTGGCAGTGGGTCTATGACAGACTTCACCCTCACCATTAATCCTGTGGAGGCTGATGATGTTGCAACATATTACTGT CAGCAAAGTAATGAGGATCCA
007 TTCAGTGGCAGTGGGTCTAGGACAGACTTCACCCTCACCATTGATCCTGTGGAGGCTGATGATGGTGCAACCTATTACTGT CAGCAAAATAATGAGGATCCG
008 TTTAGTGGCAGTGGGTCTGGGACAGACTTCAGCCTCAACATCCATCCTGTGGAGGAGGATGATATTGCAGTGTATTTCTGT CAGCAAAGTAGGAAGGTTCCT
009 TTCAGTGGCAGTGGGTCTGGGACAGACTTCACCCTCAACATCCATCCTGTGGAGGAGGAGGATGCTGCAACCTATTACTGT CAGCACATTAGGGAGCT
010 TTTAGTGGCAGTGGATCAGGGACAGATTTTACTCTTAGCATCAACAGTGTGGAGTCTGAAGATATTGCAGATTATTACTGT CAACAAAGTAATAGCTGGCCA
011 TTTAGTGGCAGTGGATCAGGGACAGATTTTACTCCTAGCATCAACAGTGTGGAGTCTGAAGATATTGCAGAATATTACTGT CAACAAAGTTATAGGTGGCCA
012 TTTAGTGGCAGTGGATCAGGGACAGACTTTACTCTTAGCATCAACAGTGTGGAGTCTGAAGATGTTGCAGATTATTACTGT CAACAAACTAATAGCTGGCCG
013 TTCAGTGGCAGTGGATCAGGGACAGATTTCACTCTCAGTATCAACAGTGTGGAGACTGAAGATTTTGGAATGTATTTCTGT CAACAGAGTAACAGCTGGCCT
014 TTCAGTGGCAGTGGATCAGGGACAGATTTCACTCTCAGTATCAACAGTGTGGAGACTGAAGATTTTGGAATGTATTTCTGT CAACAGAGTAACAACTGGCCT
015 TTCAGTGGCAGTGGATCAGGGACAGATTTCNCTCTCATTATCAACAATGTGGAGACTGAAGATTTTGGAATGTATTTCTGT CAACAGAGTAACAGCTGGCCT
016 TTCAGTGGCAGTGGATCAGGGACAGATTTCACTCTCAGTATCAACAGTGTGGAGACTGAAGATTTTGGAATGTATTTCTGT CAACAGAGTAACAGCTGGCCT
017 TTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTTTTACTGC CAGCAGTACAGTGGTTACCCA
018 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTTACAATCAGCAGCGTGGAGGCCGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTGGTTACCCA
019 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCGTGGAGGCTGAAGATGATGCAACTTATTACTGC CAGCAGTGGAGTGGTTACCCA
020 TTCAGTGGCAGTGGGGCTGGGATCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAAATGATGCAACTTATTACTGC CAGCAGTGGAGTGGTTACCCA
021 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTATCATAGTGACCCA
022 TTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTACAGTGGTTACCCA
023 TTCAGTGGCAGTGGGNNTGTGACCTCTTACTCTCTCACAATTGGCACCATGGAGGCTNAAGATNTTGCCACTTACTACTGC CAGCAGGGTAGTAGTATACCG
024 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTACAGTGGTTACCCA
025 TTCAGTGGNAGTGGGTCTGGNACCTCTTACTCTCTCACAATTGGCACCATGGAGGCTGAAGATGTTGCCACTTACTACTGC CAGCAGGGTAGTAGTATACCG
026 TTCAGTGGCAGTGGATCTGGGACCTCTTATTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGT CAACAGTGGAGTAGTTACCCA
027 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CACCAGTATCATCGTTCCCCA
028 TTCAGTGGCAGTGGGTCTGGGACATCTTTCTCTTTCACAATCAACAGCATGGAGGCTGAAGATGTTGCCACTTATTACTGT CAGCAAAGGAGTAGTTACCCA
029 TTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CATCAGCGGAGTAGTTACCCA
030 TTCAGTGGCAGTGGGTCTGGGAACTCTTATTCTCTCACAATCAGCAGCATGGAGGGTGAAGATGCTGCCACTTATTACTGC CAGCAGTTTACTAGTTCCCCA
031 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTGGTTACCAA
032 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
033 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTATCATAGTTACCCA
034 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCATTTATTACTGC CAGCAGTGGAATTATCCTCTT
035 TTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTCTCACAATCAGCAGCGTGAAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTTCCCCA
036 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCATAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CATCAGCGGAGTAGTTACCCA
037 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
038 TTCAGTGGCGAGTGGTCTGGGACCTCTTACTCTCTCGCAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAATAGTAACCCA
039 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
040 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCTCTTATTTCTGC CATCAGTGGAGTAGTTACCCG
041 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTTTTAACCCA
042 TTCAGTGGCAGTGGGTCTGCGACCTCTTACTCTCTCACAATCACCCGAATGCAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTTACCCA
043 TTCAGTGGCAGTGGGTCTGCGACCTCTTACTCTCTCACAATCACCAGCATGCAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCG
044 TTCAGTGGNAGTGGGTCTGGGACCNCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAGGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAATCCA
045 TTNAGTGGCAGTGGGTCTGGGACCTCTTNCTCTCTCACAATCAGCAGCATGGAGGNNGAAGATGNNGCCACTTATTACTGC CAGCAGTGGAGTAGTAATCCA
046 TTCAGTGGCAGTNGGNCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGT CAGCAGTGGAATAGTAACCCA
047 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGACTAGTAACCCA
048 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAATAGTAACCCG
049 TTCAGTGGNAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGACTAGTAACCCG
050 TTCAGTGGCAGTGGGNCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
051 TTCAGTGGNAGTGGGNNTGGGACCTCTTACTCTCTCACAATCAGTAGCATGGAGGCTGAGGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAATCCA
052 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTTTTACTGC CAGCAGTGGAGTAGTAACCCG
053 TTCAGTGGNAGTGGGTNTGGGACCTCTTACTCTCTCACAATCACCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTNGAGTGGNAATCCA
054 TTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
```

```
D    ┌─────────────────────────────── FR1 ───────────────────────────────┐ ┌──────────── CDR1 ──────────────
              10        20        30        40        50        60        70        80        90       100
055  GAAAATGTTCTCACCCAGTCTCCAGNAATCATGTCTGNATCTCCNGGGGNAAAGGTCACCATGACCTGC AGTGCCAGGTCA.....................
056  CAAATTGTTCTCTCCCAGTCTCCGGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGACTTGC AGGGCCAGCTCA.......... ..........
057  CAAATTGTTCTCTCCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGACTTGC AGGGCCAGCTCA.....................
058  GACATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGCATCTGTGGGTGAAACTGTCACCATCACATGT CGAGCAAGTGGG..................AAT
059  GACATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGTATCCGTGGGAGAAACTGTCACCATCACATGT CGAGCAAGTGAG..................AAT
060  GACATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGNAACTGTCACCATCACATGT CGAGCAAGTGAG..................AAT
061  GACATCCAGATGACGCAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGNAACTGTCACCATCACATGT CGAGCAAGTGAA..................AAT
062  GATGTCCAGATAACCCAGTCTCCATCTTATCTTGCTGCATCTCCTGGAGAAACCATTACTATTAATTGC AGGGCAAGTAAG..................AGC
063  GATGTTCAAATGACCCAGTCTCCATCCTCCCTGTCTGCATCTTTGGGAGAAAGAGTCTCCCTGACCTGC CAGGCAAGTCAG..................AGC
064  GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGT CGGGCAAGTCAG..................GAC
065  GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGC CGGGCAAGTCAG..................GAC
066  GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGT CGGGCAAGTCAG..................GAA
067  GACATCCAGATGATTCAGTCTCCATCGTCCATGTTTGGCTCTCTGGGAGACAGAGTCAGTCTCTCTTGC CGGGCTAGTCAG..................GGC
068  GACATCAAGATGACCCAGTCTCCATCTTCCATGTATGCATCTCTAGGAGAGAGAGTCACTATCACTTGC AAGGCGAGTCAG..................GAC
069  GACATCAAGATGACCCAGTCTCCATCTTCCATGTATGCATCTCTAGGAGAGAGAGTCACTATCTCTTGC AAGGCGAGTCAG..................GAC
070  GACATCAAGATGACCCAGTCTCCATCCTCCATGTATGCATCGCTGGGAGAGAGAGTCACTATCACTTGC AAGGCGAGTCAG..................GAC
071  GATATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCTGGGAGACAGAGTCACCATCAGTTGC AGGGCAAGTCAG..................GAC
072  GATATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCCGGGAGACAGAGTCACCATCAGTTGC AGGACAAGTCAG..................GAC
073  GACATCCAGATGACACAGTCTCCATCCTCACTGTCTGCATCTCTGGGAGGCAAAGTCACCATCACTTGC AAGGCAAGCCAA..................GAC
074  GACATCCAGATGACACAGTCTCCATCCTCACTGTCTGCATCTCTGGGAGGCAAAGTCACCATCACTTGC AAGGCAAGCCAG..................GAC
075  GATATTGTGATAACCCAGGATGAACTCTCCAATCCTGTCACTTCTGGAGAATCAGTTTCCATCTCCTGC AGGTCTAGTAAGAGTCTCCTATATAAG...GAT
076  GATATTGTGATGACGCAGGCTGCATTCTCCAATCCAGTCACTCTTGGAACATCAGCTTCCATCTCCTGC AGGTCTAGTAAGAGTCTCCTGCACAGT...AGT
077  GATATTGTGATGACGCAGGCTGCATTCTCCAATCCAGTCACTCTTGGAACATCAGCTTCCATCTCCTGC AGGTCTAGTAAGAGTCTCCTACATAGT...AAT
078  GATATTGTGATGACTCAGGCTGCACCCTCTGTATCTGTCACTCCTGGAGAGTCAGTATTCATCTCCTGC AGGTCTAGTAAGAGTCTCCTGCATAGT...AAT
079  GATATTGTGATGACTCAGGCTACACCCTCTGTATCTGTCACTCCTGGAGAGTCAGTATTCATCTCCTGC AGGTCTAGTAAGAGTCTCCTGTATATT...AAT
080  GATATTGTGATGACTCAGGCTGCACCCTCTGTACCTGTCACTCCTGGAGAGTCAGTATCCGTCTCCTGC AGGTCTAGTAAGAGTCTCCTGCATAGT...AAT
081  GATATTGTGATGACGCAGGCTGCCTTCTCCAATCCAGTCACTCTTGGAACATCAGCTTCCATCTCCTGC AGGTCTAGTAAGAATCTCCTACATAGT...AAT
082  GATGTTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGC AGATCTAGTCAGAGCCTTGTACACAGT...AAT
083  GATGTTTTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGC AGATCTAGTCAGAGCATTGTACATAGT...AAT
084  GATGCTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGC AGGTCTAGTCAGAGCCTTGAAAACAGT...AAT
085  GATGTTGTGGTGACTCAAACTCCACTCTCCCTGCCTGTCAGCTTTGGAGATCAAGTTTCTATCTCTTGC AGGTCTAGTCAGAGTCTTGCGACCAGT...CAT
086  GATGTTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGCCTGGGAGATCAAGCCTCCATCTCTTGC AGATCTAGTCAGAGCATTGTACACAGT...AAT
087  GATGCTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGC AGGTCTAGTCAGAGCATTGAAAACAGT...AAT
088  GATATTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGC AGATCTAGTCAGAGCATTGTAATCAGT...AAT
089  GATGTTGTGATGACCCAGACTCCACTCACTTTGTCGGTTACCATTGGACAACCAGCCTCCATCTCTTGC AAGTCAAGTCAGAGCCTCTTAGATAGT...GAT
090  GACATTGTGATGACACAGTCTCCATCCTCCCTGGCTATGTCAGTAGGACAGAAGGTCACTATGAGCTGC AAGTCCAGTCAGAGCCTTTTAAATAGTAGCAAT
091  GGCATTGTGATGTCACAGTCTCCATCCTCCCTAGCTGTGTCAGTTGGAGAGAAGGTTACTATGAGCTGC AAGTCCAGTCAGAGCCTTTTCTATAGTAGCAAT
092  GACATTGTGATGACACAGTCTCCATCCTCCCTGACTGTGACAGCAGGAGAGAAGGTCACTATGAGCTGC AAGTCCAGTCAGAGTCTGTTAAACAGTGGAAAT
093  GACATTGTGATGACACAGTCTCCATCCTCCCTGAGTGTGTCAGCAGGAGAGAAGGTCACTATGAGCTGC AAGTCCAGTCAGAGTCTGTTAAACAGTGGAAAT
094  AACATTATGATGACACAGTCGCCATCATCTCTGGCTGTGTCTGCAGGAGAAAAGGTCACTATGAGCTGT AAGTCCAGTCAAAGTGTTTTATACAGTTCAAAT
095  GACGTTGTGATGTCACAGTCTCCATCCTCCCTGGCTGTGTCAGCAGGAGAGAAGGTCGCTGTGAGCTGC AAATCCAGTCAGAGTCTGTCTACAGTAGAACCC
096  GNCATTGTGATGACACAGTCTCCNCNCTCCCTGAGTGTGTCAGCAGGANNNAAGGTCACTATGAGCTGT AAGTCCAGTCAGNGTCTGNTTAACAGTGNAGTC
097  GACATTGTGATGACTCAGTCTCCAACTTTCCTTGCTGTGACAGCAAGTAAGAAGGTCACCATTAGTTGC ACTGCNTCTGAGAGCCTTTATTCAAGCAAACAC
098  AGTATTGTGATGACCCAGACTCCCAAATTCCTGCTTGTATCAGCAGGAGAGAGGGTTACCATAACCTGC AAGGCCAGTCAG..................AGT
099  AGTATTGTGATGACCCAGACTCCCAAATTCCTGCCTGTATCAGCAGGAGACAGGGTTACCATGACCTGC AAGGCCAGTCAG..................AGT
100  GACATTGTGATGACCCAGTCTCACAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGC AAGGCCAGTCAG..................GAT
101  GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGC AAGGCCAGTCAG..................AAT
102  GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGTGGGAGACAGGGTCAGCGTCACCTGC AAGGCCAGTCAG..................AAT
103  GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGC AAGGCCAGTCAG..................AAT
104  GACATTGTGATGACCCAGTCTCACAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCGTCACCTGC AAGGCCAGTCAG..................AAT
105  GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCGTCACCTGC AAGGCCAGTCAG..................AAT
106  GACATTGTGATGACCCAGTCTCACAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGC AAGGCCAGTCAG..................GAT
107  GACATTGTGATGACCCAGTCTCACAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGC AAGGCCAGTCAG..................GAT
108  AACATTGTAATGACCCAATCTCCCAAATCCATGTCCGTGTCAGTAGGAGAGAGGGTCACCTTGACCTGC AAGGCCAGTGAG..................AAT
109  AACATTGTAATGACCCAATCACCCAAATCCATGTCCATGTCAGTAGGAGAGAGGGTCACCTTGAGCTGC AAGGCCAGTGAG..................AAT
```

E	CDR1	FR2	CDR2	FR3
	110 120	130 140 150 160	170 180	190 200
055	AGTATAAGTTACATGCAC	TGGTACCAGCAGNAGTCAAGCACCTCCCCNAAACTCTGGATTTAT	GACACATCCAAANTGGCTTCT	GGNGTCCCTGCNCGN
056	AGTGTAAGTTACATACAG	TGGTTCCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTCT	GTCACATCCAACCTGGCTTCT	GGAGTCCCTGCTCGC
057	AGTGTAAGTTACATACAC	TGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTAT	GCCACATCCAACCTGGCTTCT	GGAGTCCCTGTTCGC
058	ATTCACAATTATTTAGCA	TGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTAT	AATGCAAAAACCTTAGCAGAT	GGTGTGCCATCAAGG
059	ATTTACAGTAATTTGGCA	TGGTTATTCAGCAGAAACAGGGAAAACCCCCCCAGCTTGGTCTAT	GCTGCAACAAACTTAGCAGAT	GGTGTGCCATCAAGG
060	ATTTACAGTTATTTAGCA	TGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTAT	AATGCAAAAACCTTAGCAGAA	GGTGTGCCATCAAGG
061	ATTTACAGTTATTTAGCA	TGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTAT	AATGCAAAAACCTTACCAGAA	GGTGTGCCATCAAGG
062	ATTAGCAAATATTTAGCC	TGGTATCAAGAGAAACCTGGGAAAACTAATAAGCTTCTTATCTAC	TCTGGATCCACTTTGCAATCT	GGAATTCCATCAAGG
063	ATTAACAATTTTTTAAAA	TGGTTTCAGCAAACACTGGGGAAAACTGCTAGGCTCTTGATCTAT	GGTGCAAACAAATTGGAAGAT	GGGGTCCCTTCAAGG
064	ATTGGTAGTAGCTTAAAC	TGGCTTCAGCAGGAACCAGATGGAACTATTAAACGCCTGATCTAC	GCCACATCCAGTTTAGATTCT	GGTGTGCCCAAAAGG
065	ATTCATGGTTATTTAAAC	TTGTTTCAGCAGAAACCAGGTGAAACTATTAAACACCTGATCTAT	GAAACATCCAATTTAGATTCT	GGTGTCCCAAAAAGG
066	ATTAGTGGTTACTTAAGC	TGGCTTCAGCAGAAACCAGATGGAACTATTAAACGCCTGATCTAC	GCCGCATCCACTTTAGATTCT	GGTGTCCCAAAAAGG
067	ATTAGAGGTAATTTAGAC	TGGTATCAGCAGAAACCAGGTGGAACTATTAAACTCCTGATCTAC	TCCACATCCAATTTAAATTCT	GGTGTCCCATCAAGG
068	ATTAATAGCTATTTAAGC	TGGTTCCAGCAGAAACCAGGGAAATCTCCTAAGACCCTGATCTAT	CGTGCAAACAGATTGGTAGAT	GGGGTCCCATCAAGG
069	ATTAATAGCTATTTAACC	TGGTTCCAGCAGAAACCAGGGAAGTCTCCTAAGACCCTGCTCTAT	CGTACAAAGAGATTGGTAGAT	GGGGTCCCATCAAGG
070	ATTAAAAGCTATTTAAGC	TGGTACCAGCAGAAACCATGGNAATCTCCTAAGACCCTGATCTAT	TATGCAACAAGCTTGGCAGAT	GGGGTCCCATCAAGA
071	ATTAGCAATTATTTAAAC	TGGTATCAGCAGAAACCAGATGGAACTGTTAAACTCCTGATCTAC	TACACATCAAGATTACACTCA	GGAGTCCCATCAAGG
072	ATTAGCAATTTTTTATAC	TGGTTTCAGCAGAAATCAGATGGAACTGTTAAACTCCTGATCTAC	TACACCTCAAGATAACACTCA	GGAGTCCCATCAAGG
073	ATTAACAAGTATATAGCT	TGGGACCAACACAAGCCTGGAAAAGGTCCTAGGCTGCTCATACAT	TACACATCTACAATAGAGCCA	GGCATCCCATCAAGG
074	ATTAACAAGTATTTAGCT	TGGTACCAACACAAGCCTGGAAAAGGTCCTAGGCTGCTCATACAT	TACACATCTACATTACAGCCA	GGCATCCCATCAAGG
075	GGGAAGACATACTTGAAT	TGGTTTCTGCAGAGACCAGGACAATCTCCTCAGCTCCTGATCTAT	TTGATGTCCACCCGTGCATCA	GGAGTCTCAGACCGG
076	GGCAACACTTACTTGTAT	TGGTTCCTGCAGAAGCCAGGCCAGTCTCCTCAGCTCCTGATATAT	TATATCTCCAACCTTGCCTCA	GGAGTCCCAGACAGG
077	GGCATCACTTATTTGTAT	TGGTATCTGCAGAAGCCAGGCCAGTCTCCTCAGCTCCTGATTTAT	CAGATGTCCAACCTTGCCTCA	GGAGTCCCAGACAGG
078	GGCAACACTTACTTGTAT	TGGTACCTACAGAGGCCAGGCCAGTCTCCTCAGCTCCTGATATAT	CGGATGTCCAACCTTGCCTCA	GGAGTCCCAGACAGG
079	GGCAACACTTACTTGTAT	TGGTACCTACAGAGGCCAGGCCAGTCTCCTCAACTCCTGATATAT	CGGATGTCCTACCTTGCCTCA	GGAGTCCCAGACAGG
080	GGCAACACTTACTTGTAT	TGGTTCCTGCAGAGGCCAGGCCAGTCTCCTCAGCTCCTGATATAT	CGGATGTCCAACCTTGCCTCA	GGAGTCCCAGACAGG
081	GGCATCACTTTTTTATAT	TGGTATCTCCAGAGGCCAGGCCAGTCTCCTCAGCTCCTGATATAT	CGGGTGTCCAATCTGGCCTCA	GGAGTCCCAAACAGG
082	GGAAACACCTATTTACAT	TGGTACCTGCAGAAGCCAGGCCAGTCTCCAAAGCTCCTGATCTAC	AAAGTTTCCAACCGATTTTCT	GGGGTCCCAGACAGG
083	GGAAACACCTATTTAGAA	TGGTACCTGCAGAAACCAGGCCAGTCTCCAAAGCTCCTGATCTAC	AAAGTTTCCAACCGATTTTCT	GGGGTCCCAGACAGG
084	GGAAACACCTATTTGAAG	TGGTACCTCCAGAAACCAGGCCAGTCTCCACAGCTCCTGATCTAC	AGGGTTTCCAACCGATTTTCT	GGGGTCCTAGACAGG
085	GGGATCACCTATTTGTCT	TGGTACCTGCACAAGCCTGGCCAGTCTCCACAGCTCCTCATCTAT	GGGATTTCCAACAGATTTTCT	GGGGTGCCAGACAGG
086	GGAAACACCTATTTATAT	TGGTACCTGCAGAAACCAGGCCAGTCTCCAAAGCTCCTGATCTAC	AGGGTTTCCAACCGATTTTCT	GGGGTCCCAGACAGG
087	GGAAACACCTATTTGAAC	TGGTACCTCCAGAAACCAGGCCAGTCTCCCAGGCTCCTGATCTAC	AGGGTTTCCAACCGATTTTCT	GGGGTCCTAGACAGG
088	GGGTTCACCTATTTAGAA	TGGTACCTGCAGAAACCAGGNNNNNNNNNAAAGCTCCTGATATAT	GGGATTTCCAACCGATTTTCT	GGGGTCCCAGACAGG
089	GGAAAGACATATTTGAAT	TGGTTGTTACAGAGGCCAGGCCAGTCTCCAAAGCGCCTAATCTAT	CTGGTGTCTAAACTGGACTCT	GGAGTCCCTGACAGG
090	CAAAAGAACTATTTGGCC	TGGTACCAGCAGAAACCAGGACAGTCTCCTAAACTTCTGGTATAC	TTTGCATCCACTAGGGAATCT	GGGGTCCCTGATCGC
091	CAAAAGAACTCTTTGGCC	TGGTACCAGCAGAGACCAGGGCAGTCTCCTAAACTGCTGATTTAC	TGGGCATCCACTAGGGAATCT	GGGGTCCCTGATCGC
092	CCGAAGAACTACTTGACC	TGGTACCAGCAGAAACCAGGGCAGCCTCCTAAACTGTTGATCTAC	TGGGCATCCACTAGGGAATCT	GGGGTCCCTGATCGC
093	CAAAAGAACTACTTGGCC	TGGTACCAGCAGAAACCAGGGCAGCCTCCTAAACTGTTGATCTAC	GGGGCATCCACTAGGGAATCT	GGGGTCCCTGATCGC
094	CAGAAGAACTACTTGGCC	TGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATCTAC	TGGGCATCCACTAGGGAATCT	GGAGTCCCTGATCGC
095	GAAAGAAGCTACTTGGCT	TGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATCTAC	TGGGCATCCACTAGGGAATCT	GGGGTCCCTGATCGC
096	AAAAGAACTAACTTGGCC	TGGTACCANAAGAAACCAGGGCAGCCTCCNAAACTGTTGATCTCC	GTGGATGCGCGACCCNCACAC	GGAGTCCCTGATCGC
097	AAGGTGCACTACTTGGCT	TGGTACCAGAAGAAACCAGAGCAATCTCCTAAACTGCTGATATAC	GGGGCATCCAACCGATACATT	GGGGTCCCTGATCGC
098	GTGAGTAATGATGTAGCT	TGGTACCAACAGAAGCCAGGGCAGTCTCCTAAACTGCTGATATAC	TATGCATCCAATCGCTACACT	GGAGTCCCTGATCGC
099	GTGGGTAATAATGTAGCC	TGGTACCAACAGAAGCCAGGACAGTCTCCTAAACTGCTGATATAC	TATGCATCCAATCGCTACACT	GGAGTCCCTGATCGC
100	GTGGGTGCTGCTATAGCC	TGGTATCAACAGAAACCAGGACAATCTCCTAAACTACTGATTTAC	TGGGCATCCACCCGGCACACT	GGAGTCCCTGATCGC
101	GTGGGTACTGCTGTAGCC	TGGTATCAACAGAAACCAGGACAATCTCCTAAACTACTGATTTAC	TCGGCATCCAATCGGTACACT	GGAGTCCCTGATCGC
102	GTGGTCACTAATGTAGCC	TGGTATCAACAGACACCAGGACAATCTCCTAAAGCACTGATTTAC	TCGGCATCCTACCGGTACAGT	GGAGTCCCTGATCGC
103	GTTCGTACTGCTGTTGCC	TGGTATCAACAGAAACCAGGGCAGTCTCCTAAAGCACTGATTTAC	TTGGCATCCAACCGGTACACT	GGAGTCCCTGATCGC
104	GTGGGTACTAATGTAGCC	TGGTATCAACAGAAACCAGGGCAATCTCCTAAAGCACTGATTTAC	TCGGCATCCTACCGGTACAGT	GGAGTCCCTGATCGC
105	GTGGGTACTAATGTAGCC	TGGTATCAGCAGAAACCAGGGCAATCTCCTAAAGCACTGATTTAC	TCGGCATCCTACCGGTACAGT	GGAGTCCCTGATCGC
106	GTGAGTACTACTGTGGCC	TGGTATCAGCAGAAACCAGGGCAATCTCCTAAACTACTGATTTAT	TCGGCATCCTACCGGTACACT	GGAGTCCCTGATCGC
107	GTGAGTACTGCTGTAGCC	TGGTATCAACAGAAACCAGGACAATCTCCTAAACTACTGATTTAC	TCGGCATCCTACCGGTACACT	GGAGTCCCTGATCGC
108	GTGGTTACTTATGTTTCC	TGGTATCAACAGAAACCAGAGCAGTCTCCTAAACTGCTGATATAC	GGGGCATCCAACCGGTACACT	GGGGTCCCCGATCGC
109	GTGGGTACTTATGTATCC	TGGTATCAACAGAAACCAGAGCAGTCTCCTAAACTGCTGATATAC	GGGGCATCCAACCGGTACACT	GGGGTCCCCGATCGC

```
F   --------------------------------------FR3----------------------------------------┐ ┌------CDR3----------
           210       220       230       240       250       260       270       280        290       300
055 TTCAGTGGNAGTGGGNCTGGNAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCNGAAGATGTTGCCACTTATTACTGT TTTCNGGGGAGTGGGTACCCA
056 TTCAGTGGNAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGGAGTAACCCA
057 TTCAGTGGAAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAACAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGC CAGCAGTGGAGTAGTAACCCA
058 TTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCAGCCTGAAGATTTTGGGAGTTATTACTGT CAACATTTTTGGAGTACTCCT
059 TTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCAGCAGCCGGAGGATTTTGGGAGTTATTACTGT CAACATTTTTGGAGTGCTCCT
060 TTCAGTGGCAGTGGATCAGGCACACAGTTTTCTCTGAAGATCAACAGCCTGCAGCCTGAAGATTTTGGGAGTTATTACTGT CAACATCATTATGTTACTCCG
061 TTCAGTGGCAGTGGATCAGGCACACAGTTTTCTCTGAAGATCAACAGCCTGCAGCCTGAAGATTTTGGGAGTTATTACTGT CAGCATCATTATGGTCCTCCG
062 TTCAGTGGCAGTGGATCTGGTACAGATTTCACTCTCACCATCAGTAGCCTGGAGCCTGAAGATTTTGCAATGTATTACTGT CAACAGCATAATGAATACCCG
063 TTCAGTGGAACTGGATATGGACAGATTTCACTTTCACCATCAGCAGCCAGGAGGAAGAAGATGTGTCAACTTATTTCTGT CTACAGCATAGGTATCTCCCT
064 TTCAGTGGCAGTAGGTCTGGGTCAGATTATTCTCTCACCATCAGCAGCCTTGAGTCTGAAGATTTTGTAGACTATTACTGT CTACAATATGCTAGTTCTCCT
065 TTCAGTGGCAGTAGGTCTGGGTCAGATTATTCTCTCATTATCGGCAGCCTTGAGTCTGAAGATTTTGCAGACTATTACTGT CTACAATATGCTAGTTCTCCT
066 TTCAGTGGCAGTAGGTCTGGGTCAGATTATTCTCTCACCATCAGCAGCCTTGAGTCTGAAGATTTTGCAGACTATTACTGT CTACAATATCTTAGTTATCCG
067 TTCAGTGGCAGTGGGTCTGGGTCAGATTATTCTCTCACCATCAGCAGCCTAGAGTCTGAAGATTTTGCAGACTATTACTGT CTACAGCGTAATGCGTATCCG
068 TTCAGTGGCAGTGGATCTGGGCAAGATTATTCTCTCACCATCAGCAGCCTGGAGTATGAAGATATGGGAATTTATTATTGT CTACAGTATGATGAGTTTCCT
069 TTCAGTGGCAGTGGATCTGGGCAAGAATATTCTCTCACCATCAGCAGCCTGGAGTATGAAGATATGGGAATTTATTTTTGT CTTCAGTATGATGAATTTCTT
070 TTCAGTGGCAGTGGATCTGGGNAAGATTATTCTCTAACCATCAGCAGCCTGGAGTCTGACGATACAGCAACTTATTACTGT CTACAGCATGGTGAGAGCCCT
071 TTCAGTGGCAGTGGGTCTGGAACAGATTATTCTCTCACCATTAGCAACCTGGAGCAAGAAGATATTGCCACTTACTTTTGC CAACAGGGTAATACGCTTCCT
072 TTCAGTGGCAGTGGGTCTGGAACAGATTATTCTTTCACCATTAACAACCTGGAGTAAGAAGATGTCGCCACTTATTCTTGA CAACAGGGTATATT
073 TTCAGTGGAAGTGGGTCTGGGAGAGATTATTCCTTCAGCATCAGCAACCTGGAGCCTGAAGATATTGCAACTTATTATTGT CTACAGTATGATAATCTGTAC
074 TTCAGTGGAAGTGGGTCTGGGAGAGATTATTCCTTCAGCATCAGCAACCTGGACGCGGAAGAGATTGCAACTTATTATTGT CTACAGTATGATAGTCTGTAC
075 TTTAGTGGCAGTGGGTCAGGAACAGATTTCACCCTGGAAATCAGTAGAGTGAAGGCTGAGGATGTGGGTGTGTATTACTGT CAACAACTTGTAGAGTATCCT
076 TTCAGTGGCAGTGGGTCAGGAACTGATTTCACACTGAGAATCAGTAGAGTGGAGGCTGAGGATGTGGGTGTTTATTACTGT ATGCAAGGTCTAGAATATCCT
077 TTCAGTAGCAGTGGGTCAGGAACCGACTTCACACTGAGAATCAGCAGAGTGGAGGCTGAGGATGTGGGTGTTTATTACTGT GCTCAAAATCTAGAACTTCCT
078 TTCAGTGGCAGTGGGTCAGGAACTGCTTTCACACTGAGAATCAGTAGAGTGGAGGCTGAGGATGTGGGTATTTATTATTGT ATGCAACATCTAGAATATCCT
079 TTCAGTGGCAGTGGGTCAGGAACTGCTTTCACACTGAGAATCAGTAGAGTGGAGGCTGAGGATGTGGGTATTTATTATTGT ATGCAACATCTAGAATATCCT
080 TTCAGTGGCAGTGGGTCAGGAACTGCTTTCACACTGAGAATCAGTAGAGTGGAGGCTGAGGATGTGGGTGTTTATTACTGT ATGCAACATCTAGAATATCCG
081 TTCAGTGGCAGTGAGTCAGGAACTGATTTCACACTGAGAATCAGCAGAGTGGAGGCTGAGGATGTGGGTGTTTATTACTGT GCTCAACTGCTAGAACTCCC
082 TTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAAGATCAGCAGAGTGGAGGCTGAGGATCTGGGAGTTTATTTCTGC TCTCAAAGTACACATGTTCCT
083 TTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAAGATCAGCAGAGTGGAGGCTGAGGATCTGGGAGTTTATTACTGC TTTCAAGGTTCACATGTTCCT
084 TTCAGTGGTAGTGGATCAGGGACAGATTTCACACTGAAAATCAGCAGAGTGGAGGCTGAGGATTTGGGAGTTTATTTCTGC CTCCAAGTTACACATGCTCCT
085 TTCAGTGGCAGTGGTTCAGGGACAGATTTCACACTCAAGATCAACACAATAAAGCCTGAGGACTTGGGAATGTATTACTGC TTACAAGGTTCACATCAGCCG
086 TTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAATATCAGCAGAGTGGAGGCTGAGGATATGGGAGTTTATTACTGC TTTCAAGGTACACATGTTCCT
087 TTCAGTGGTAGTGGATCAGGGACAGATTTCACACTGAAAATCAGCAGAGTGGAGGCTGAGGATTTGGGAGTTTATTTCTGC CTCCAAGTTACACATGTCCCG
088 TTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAAGATCAGCAGAGTGGAGGCTGAGGATGTAGGAATTTATTACTGT TTTCAAGGTATACATGTTCCG
089 TTCACTGGCAGTGGATCAGGGACAGATTTCACACTGAAAATCAGCAGAGTGGAGGCTGAGGATTTGGGAGTTTATTATTGC TGGCAAGGTACACATTTTCCT
090 TTCATAGGCAGTGGATCTGGGACAGATTTCACTCTTACCATCAGCAGTGTGCAGGCTGAAGACCTGGCAGATTACTTCTGT CAGCAACATTATAGCACTCCG
091 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGTGTGAAGGCTGAAGACCTGGCAGTTTATTACTGT CAGCAATATTATAGCTATCCG
092 TTCACAGGCAGTGGATCTGGAACAGATTTCACTCTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCAGTTTATTACTGT CAGAATGATTATAGTTATCCG
093 TTCACAGGCAGTGGATCTGGAACCGATTTCACTCTTACCATCAGCAGTGTGCAGGCTGAAGACCTGGCACTTTATTACTGT CAGAATGATCATACTTATCCG
094 TTCACAGGCAGTGGATCTGGGACAGATTTTACTCTTACCATCAGCAGTGTACAAGCTGAAGACCTGGCAGTTTATTACTGT CATCAATACCTCTCCTCG
095 TTCACGGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGTGTGCGGGCTGAAGACCTGGCAGTTTATTACTGC AAGAATCTTATGGATCTTCCC
096 TTCACAGGAAGTGGTTCTGGGAGNGATTATACTCTCACAGTCAGCAGTGTGAAGGCTGAAGACCTGGCACTTTACTACTGT CAACAACATTATAGNTATCCG
097 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTGACCATCAGCAGTGTACAGGTTGAAGACCTCACACATTATTACTGT GCACAGTTTTACAGCTATCCT
098 TTCACTGGCAGTGGATATGGACGGATTTCACTTTCACCATCAGCACTGTGCAGGCTGAAGACCTGGCAGTTTATTTCTGT CAGCAGGATTATAGCTCTCCT
099 TTCACTGGCAGTGGATCTGGGACAGATTTCACTTTCACCATCAGCAGTGTGCAGGTTGAAGACCTGGCAGTTTATTTCTGT CAGCAGCATTATAGCTCTCCG
100 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATTAGCAATGTGCAGTCTGATGACTTGGCAGATTATTTCTGT CAACAATATAGCGGGTATCCT
101 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAATATGCAGTCTGAAGACCTGGCAGATTATTTCTGC CAGCAATATAGCAGCTATCCT
102 TTCTCAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAATGTGCAGTCTGGAGACTTGGCAGAGTATTTCTGT CAGCAATATAACAGCTATCCT
103 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATTACCAATGTGCAATCTGAAGACCTGGCAGATTATTTCTGT CTGCAACATTGGAATTATCCG
104 TTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAATGTGCAGTCTGAAGACTTGGCAGAGTATTTCTGT CAGCAATATAACAGCTGTCCA
105 TTCACAGGCAGTGGATCTGGGACGGATTTCACTCTCACCATCAGCAATGTGCAGTCTGAAGACTTGGCAGTGTATTTCTGT CAGCAATATAACAGCTATCCG
106 TTCACTGGCAGTGGATCTGGGACGGATTTCACTTTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCAGTTTATTACTGT CAGCAACATTATAGTACTCCT
107 TTCACTGGCAGCGGATCTGGGACGGATTTCACTTTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCAGTTTATTACTGT CAGCAACATTATAGTACTCCG
108 TTCACAGGCAGTGGATCTGCAACAGATTTCACTCTGACCATCAGCAGTGTGCAGGCTGAAGACCTTGCAGATTATCACTGT GGACAGGGTTACAGCTATCCG
109 TTCACAGGCAGTGGATCTGCAACAGATTTCACTCTGACCATCAGCAGTGTGCAGTCTGAAGACCTTGCAGATTATTTCTGT GGACAGAGTTACAGCTATCCC
```

similar members from a single V_k gene family (*V_k4/5*) were present in different groups (IV and VI).

Organization of V_k proteins into subgroups using < 13 mismatches up to Trp35 as a criterium (Potter et al. 1982) better reflected primary structure similarities, although such organization frequently led to multiple assignments, in which cases only a single assignment for the sequence representing the best match was included (Table 1). Moreover, as will be shown below, this classification repeatedly failed to adequately reflect overall similarity at the nucleic acid sequence level. Finally, some sequences (discussed below) could not be assigned unambiguously to any existing V_kTrp subgroup.

We then determined whether V_k nucleic acid sequences could be organized into gene families (analogous to V_H genes), and how such families related to V_k protein groups and subgroups. For this purpose, all V_k genes in the data bank were arranged in groups of > 80% sequence similarity, which were termed V_k gene families. The characteristics of these families and their relationship to V_k protein groups and subgroups are detailed below. A quick summary outlining how the different classifications correspond to each other is presented in Table 2.

V_k21 gene family. All *V_k21* genes fulfilled the criteria for a typical *V* gene family, i.e., all members were > 80% similar (mostly > 90%) and differed from all other V_k sequences by at least 25%. This gene family corresponded completely to protein subgroup V_k21 which, in turn, coincided with V_k protein group III. Five germline genes have been cloned (Heinrich et al. 1984) and approximately ten expressed sequences have been published. *V_k21* genes were used in response to influenza hemagglutinin (Clarke et al. 1985, Meek et al. 1989) and major histocompatibility complex class II antigens (Devaux et al. 1985), and encoded some lupus-associated autoantibodies (Shlomchik et al. 1987c). The P3-X63-Ag8.653 myeloma line, a derivative of the MOPC21 myeloma that has lost the ability to express Ig heavy and light chain proteins and is frequently used in hybridoma technology (Kearney et al. 1979), also expressed a non-functional V_k21 mRNA (Strohal et al. 1987). With the exception of an MRL-*lpr/lpr* rheumatoid factor (RF, anti-Ig) V_k sequence (AM12; Shlomchik et al. 1987c), which differed from all known *V_k21* germline genes by > 30 bp and may have derived from an unknown germline gene, all other expressed sequences were very similar to, and hence probably derived from, known *V_k21* germline genes. RFLP (Kofler et al. 1989) and gene cloning analyses (Heinrich et al. 1984) suggested an estimated 6 to 13 *V_k21* germline genes in the genome of most inbred strains of mice.

Finally, an incomplete V_k21 sequence (VM201, Meek et al. 1989), which was therefore not included in our data bank, should be mentioned as it lacked two codons in CDR-1 in comparison to other V_k21 sequences. Unless caused by somatic events, this would make the corresponding germline gene the only V_k gene with 37 codons up to Trp35.

V_k23 gene family. Similar to V_k21, V_k23 sequences were well separated from all other V_k sequences, and formed a gene family that corresponded entirely to its protein counterpart, the V_k23 subgroup (protein group V). One germline gene has been reported (Pech et al. 1981) that was subsequently observed in RFs from BALB/c mice (Shlomchik et al. 1987a), and that probably encoded an (NZB × NZW)F_1 RNA-specific autoantibody (Eilat et al. 1988).

Additional *V_k23* genes, more distant from the above germline gene but closely related to each other, possibly derived from a second *V_k23* germline gene and encoded nitrophenyl-specific anti-idiotypes (Sablitzky and Rajewsky 1984) and a creatine-kinase-specific antibody (Buckel et al. 1987). A nonfunctional V_k23 member was cloned from an MRL/n RF-producing hybridoma and might correspond to another *V_k23* (pseudo) gene (Kofler et al. 1989). Our previous RFLP analyses suggested the presence of four to eight *V_k23* germline genes in the genome of most inbred strains of mice. However, this may represent an over-estimate due to cross-hybridization of the more conserved 3′ portion of the V_k23 probe with V_k1 sequences (Kofler et al. 1989, and below).

◀

Fig. 1. Nucleic acid sequences of 109 V_k genes contained in the V_k database. *Dots* have been introduced to maximize homology; N, undetermined nucleotides; CDR, complementarity determining region; FR, frame work region (according to Kabat et al. 1987). Codes of V_k genes: 1 = *V_k21B*, 2 = *V_k21C*, 3 = *V_k21E*, 4 = *V_k1.6kb*, 5 = *V_k18kb*, 6 = *H37-85*, 7 = *AM10*, 8 = *AM12*, 9 = *Ag8.653k-*, 10 = *L7*, 11 = *T2*, 12 = *D444*, 13 = *MRL/n-RF33B*, 14 = *A8/4*, 15 = *A20/44*, 16 = *MAK33*, 17 = *H1*, 18 = *R11*, 19 = *R1*, 20 = *L8*, 21 = *MRL-Histone7*, 22 = *MRL-DNA22*, 23 = *NQ10-4.6.1*, 24 = *NQ11-1.18*, 25 = *NQ22-87.1*, 26 = *A9*, 27 = *37A4*, 28 = *R2*, 29 = *R9*, 30 = *R13*, 31 = *H2*, 32 = *H3*, 33 = *H4*, 34 = *H6*, 35 = *H8*, 36 = *H9*, 37 = *H13*, 38 = *L6*, 39 = *70Z/3*, 40 = *AM1*, 41 = *2H7*, 42 = *NQ2-6.1*, 43 = *NQ2-48.2.2*, 44 = *NQ10-12.4.6*, 45 = *NQ10-12.5*, 46 = *NQ10-15.3*, 47 = *NQ11-7.12*, 48 = *NQ11-8.1*, 49 = *NQ22-15.18*, 50 = *NQ22-18.7*, 51 = *NQ22-61.1*, 52 = *NQ22-17.18*, 53 = *NQ19-2.4*, 54 = *NQ18-36.44*, 55 = *NQ16-38.18*, 56 = *NQ10-11.1*, 57 = *NQ10-2.12.8*, 58 = *K2*, 59 = *K3*, 60 = *A25.9.7*, 61 = *A31.90*, 62 = *MRL-RF24*, 63 = *PC6684K-*, 64 = *MOPC41*, 65 = *M173B*, 66 = *GLOOP1*, 67 = *BXW-DNA16*, 68 = *L6*, 69 = *40-140*, 70 = *CP5 B5-3*, 71 = *V_kARS*, 72 = *PC3386*, 73 = *38C13*, 74 = *VM113*, 75 = *V_k167*, 76 = *V_k24A*, 77 = *V_k24B*, 78 = *AM28*, 79 = *AM29*, 80 = *A15*, 81 = *25-39*, 82 = *K5.1*, 83 = *K1A5*, 84 = *K18.1*, 85 = *W3129*, 86 = *L XIX 27*, 87 = *JV3*, 88 = *HP9*, 89 = *BXW-DNA14*, 90 = *V_k139*, 91 = *GLOOP5*, 92 = *AM13*, 93 = *VS3*, 94 = *A17*, 95 = *JV10*, 96 = *PY102*, 97 = *S107A*, 98 = *V_kSer*, 99 = *MRL-RF28*, 100 = *CEA66-E3*, 101 = *V-TNP*, 102 = *B6.2*, 103 = *CEM231.6.7*, 104 = *A23*, 105 = *A34*, 106 = *RF49*, 107 = *RF49B*, 108 = *RF34*, 109 = *AM16*.

A

	FR1	CDR1	FR2	CDR2	FR3	CDR3
	10 20	30	40	50	60 70 80	90
001	NIVLTQSPASLAVSLGQRATISC	RASESVDS..YGNSFMH	WYQQKPGQPPKLLIY	LASNLES	GVPARFSGSGSRTDFTLTIDPVEADDAATYYC	QQNNEDP
002	DIVLTQSPASLAVSLGQRATISC	RASESVDS..YGNSFMH	WYQQKPGQPPKLLIY	RASNLES	GIPARFSGSGSRTDFTLTINPVEADDVATYYC	QQSNEDP
003	DIVLTQSPASLAVSLGQRATISC	RASKSVST..SGYSYMH	WYQQKPGQPPKLLIY	LASNLES	GVPARFSGSGSGTDFTLNIHPVEEEDAATYYC	QHSRELP
004	DIVLTQSPASLAVSLGQRATISC	RASQSVST..SSYSYMH	WYQQKPGQPPKLLIK	YASNLES	GVPARFSGSGSGTDFTLNIHPVEEEDTATYYC	QHSWEIP
005	DIVLTQSPASLAVSLGQRATIFC	RASQSVDY..NGISYMH	WFQQKPGQPPKLLIY	AASNLES	GIPARFSGSGSGTDFTLNIHPVEEEDAATYYC	QQSIEDP
006	DIVLTQSPGSLAVSLGQRATISC	RASESVES..SGNNFIH	WHQQKPGQPPXLLIY	RASNLAS	GIPARFSGSGSMTDFTLTINPVEADDVATYYC	QQSNEDP
007	KIVLTQFPASLAVSLRQRATISC	RASESVDS..YGNSFMY	WYQQKPGQPPKLLIY	RASNLES	GVPARFSGSGSRTDFTLTIDPVEADDGATYYC	QQNNEDP
008	DIVLTQSPASLAVSLGQSVTISC	RASESVEY..YGSSLMQ	WYQQKPGQPPKLLIY	GASNVES	GVPARFSGSGSGTDFSLNIHPVEEDDIAVYFC	QQSRKVP
009	DIVLTQFPASLAVSLGQRATISY	RASKSVST..SGYSYMH	WNQQKPGQPPRLLIY	LVSNLES	GVPARFSGSGSGTDFTLNIHPVEEEDAATYYC	QHIREXX
010	DILLTQSPAILSVSPGERVSFSC	RASQ......SIGTSIH	WYQQRTNGSPRLLIK	YASESIS	GIPSRFSGSGSGTDFTLSINSVESEDIADYYC	QQSNSWP
011	DILLTQSPAILSVSPGERVSFSC	RASQ......SIGTSIH	WYQQRTNGSPRLLIK	NASESIS	GIPSRFSGSGSGTDFTPSINSVESEDIAEYYC	QQSYRWP
012	DILLTQSPAILSVSPGERVSFSC	RASQ......SIGTSLH	WYQQRTNGSPRLLIK	YASESIS	GIPSRFSGSGSGTDFTLSINSVESEDVADYYC	QQTNSWP
013	DIVLTQSPATLSVTPGDSVSLSC	RASQ......SIINNLH	LYRUKSHESPRLLIK	YASQSIS	GIPSRFSGSGSGTDFTLSINSVETEDFGMYFC	QQSNSWP
014	DIVLTQSPATLSVTPGDSVSLSC	RASQ......SISNNLH	WYQQKSHESPRLLIN	YASQSMS	GIPSRFSGSGSGTDFTLSINSVETEDFGMYFC	QQSNNWP
015	DIVLTQSPATLSVTPGDSVSLSC	RASQ......SISNNLH	WYQQKSHESPRLLIK	YASQSIS	GIPSRFSGSGSGTDFXLIINNVETEDFGMYFC	QQSNSWP
016	DIVLTQSPATLSVTPRDSVSLSC	RASQ......SISNNLH	WYQQKSHESPRLLIK	YASQSIS	GIPSRFSGSGSGTDFTLSINSVETEDFGMYFC	QQSNSWP
017	QIVLTQSPAIMSASPGEKVTMTC	SARSS.....VSSSYLY	WYQQKPGSSPKLWIY	STSNLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATFYC	QQYSGYP
018	ENVLTQSPAIMAASPGEKVTMTC	SASSS.....VSSSNLH	WYQQKSGTSTKFWIY	RTSNLAS	EVPAPFSGSGSGTSYSLTISSVEAEDAATYYC	QQWSGYP
019	ENVLTQSPAIMAASLGQKVTMTC	SASSS.....VSSSYLH	WYQQKSGASPKPLIH	RTSNLAS	GVPARFSGSGSGTSYSLTISSVEAEDDATYYC	QQWSGYP
020	ENVLTQSPAIMAASLGEKVTMTC	SASSS.....VSSSYLH	WYQQKSGTSPKLWIY	GTSNLAS	GVPARFSGSGAGISYSLTISSMEAENDATYYC	QQWSGYP
021	QIVLTQSPAIMSASPGEKVTMTC	SASSS.....VSSKYLN	WYQQRSGASPKLWIY	GTSNLAS	GVPARFSGSGSGTSYSLTISSVEAEDAATYYC	QQYHSDP
022	QIVLTQSPAIMSASPGERVTMTC	SASSS.....VSSSYLY	WYQQKPGSSPKLWIY	STSNLAS	GVPARFSGSGSGTSYSLTISSMEAED:ATYYC	QQYSGYP
023	EIVLTQSPTTMAXSPGEKITITC	SANSS.....ISSNYLH	WYQQKPGFSPKLLIY	RTSNLAS	GVQARFSGSGXVTSYSLTIGTMEAXDXATYYC	QQGSSIP
024	ENVLTQSPAIMSASPGXKVTMTC	RASSS.....VSSSYLH	WYQQKSGASPKLWIY	STSNLAS	XVPARFSGSGSGTSYSLTISSVEAEDAATYYC	QQYSGYP
025	EIVLTQSPTTMAASPGEKITITC	SASSS.....ISSNYLN	WFQQKPGFSPKLLIY	RTSNLAS	GVPDRFSXSGSXTSYSLTIGTMEAEDVATYYC	QQGSSIP
026	EIVLTQSPALMAASPGEKVTITC	SVSSS.....ISSSNLH	WYQQKSETSPKSWIY	GTSNLAS	GVPVRFSGSGSGTSYSLTISSMEAEDAATYYC	QQWSSYP
027	QIVLTQSPAFMSASLGERVTMTC	TARSS.....VSSSYFH	WYQQKPGSSPKLWIY	STSNLAS	GVPTRFSGSGSGTSYSLTISSMEAEDAATYYC	HQYHRSP
028	EILLTQSPAIIAASPGEKVTITC	SASS.......SVSYMN	WYQQKPGSSPKIWIY	GISNLAS	GVPARFSGSGSGTSFSFTINSMEAEDVATYYC	QQRSSYP
029	QIVLTQSPAIMSASPGEKVTMTC	SASS.......SISYMH	WYQQKPGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	HQRSSYP
030	ENVLTQSPAIMSASLGEKVTMSC	RASS.......SVNYMY	WYQQKSDASPKLWIY	YTSNLAP	GVPARFSGSGSGNSYSLTISSMEGEDAATYYC	QQFTSSP
031	GIVLTQSPTTMTAFPGENVTITC	SASS.......SINYIH	WYQQKSGNTPKQUIY	KTSDLPS	GVPTLFSGSGSGTSYSLTISSVEAEDAATYYC	QQWSGYP
032	QIVLTQSPAIMSASPGEKVTMTC	SASS.......SVSYMH	WYQQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	QQWSSNP
033	QIVLTQSPAIMSASPGEKVTISC	SASS.......SVSYMY	WYQQKPGSSPKPWIY	RTSNLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	QQYHSYP
034	EIVLTQSPAITAASLGQKVTITC	SASS.......SVSYMH	WYQQKSGTSPKPWIY	EISKLAS	GVPARFSGSGSGTSYSLTISSMEAEDAAIYYC	QQWNYPP
035	QIVLTQSPAILSASPGEKVTMTC	SASS.......SVSYMU	WFQQKPGSSPKLWIY	SISNLAS	GVPARFSGSGSGTSYSLTISSVKAEDAATYYC	QQWSSSP
036	QILLTQSPAIMSASPGEKVTMTC	SASS.......SVSYMH	WYQQKPGSSPKPWIY	DTSNLAS	GFPARFSGSGSGTSYSLIISSMEAEDAATYYC	HQRSSYP
037	QIVLTQSPALMSASPGEKVTMTC	SASS.......SVSYMY	WYQQKPRSSPKPWIY	LTSNLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	QQWSSNP
038	QIVLSQSPAILSASPGEKVTLTC	RASS.......SVSFMN	WYQQKPGSSPKPWIY	ATSNLAS	EFPGRFSGEWSGTSYSLAISRVEAEDAATYYC	QQWNSNP
039	QIVLSQSPAILSASPGEKVTMTC	RASS.......SVSYMH	WYQQKLGSSPKPWIY	ATSNLAS	GVPARFSGSGSGTSYSLTISRVEAEDAATYYC	QQWSSNP
040	QIVLTQSPAIMSASPGQKVTITC	SASS.......SVNYMH	WYQQKLGSSPKLWIY	DTSKLAP	GVPARFSGSGSGTSYSLTISSMEAEDAASYFC	HQWSSYP
041	QIVLSQSPAILSASPGEKVTMIC	RASS.......SVSYMH	WYQQKPGSSPKPWIY	APSNLAS	GVPARFSGSGSGTSYSLTISRVEAEDAATYYC	QQWSFNP
042	QILLTQSPAIMSASPGQKVTMTC	SASS.......SVSYMY	WYQQKPGSSPRLLIY	DTSNLAS	GVPVRFSGSGSATSYSLTITRMQAEDAATYYC	QQWSSYP
043	QILLTQSPAIMSASPGQKVTMTC	SASS.......SVSYMH	WYQQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSATSYSLTITSMQAEDAATYYC	QQWSSNP
044	QNVLNQSPXIMSXSPGEKVTMTC	SASS.......SVSYMQ	WFQQXSGTSPKRWIY	DTSKLXS	XVPTRFSXSGSGTXYSLTISSMEAEDAATYYC	QQWSSNP
045	QIVLTQSPXIMSXSPGEKVTMTC	SASS.......SVRYMN	WFQQKSGTSPKRWIY	DTSKLSS	GVPARXSGSGSGTSXSLTISSMEXEDXATYYC	QQWSSNP
046	QIVLTQSPAIMSASPGEKVTMTC	SASS.......SVSYMN	WFQQKSGTSPKRWIY	DTSKLSS	GVPPRFSGSXXGTSYSLTISSMEAEDAATYYC	QQWNSNP
047	QIVLTQSPAIMSASPGEKVTMTC	SASS.......IVSYVQ	WFQQKSGTSPKRWIS	DTSKLPS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	QQWTSNP
048	QIVLIQSPXIMSASPGXKXTMTC	SASS.......SVSYMN	WYQQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTSYSLTISSMEAEDAATYYC	QQWNSNP
049	QIVLTQSPPIMSASPGEKVTMTC	SASS.......SVSYLQ	WFQQKSGTSPKRWIY	DTSKLDS	XVPARFSXSGSGTSYSLTISSMEAEDAATYYC	QQWTSNP
050	QIVLTQSPAIMSASPGEKV' MTC	SASS.......SVSYMN	WFQQKSGTSPKRWVF	ATSKLXS	GVPARFSGSGXGTSYSLTISSMEAEDAATYYC	QQWSSNP
051	QVVLTQSPXIMSASPGXKVTMTC	SASS.......SVSYMQ	WFQQKSGTSPKRLIF	YTSKLTS	GVPARFSXSGXGTSYSLTISSMEAEDAATYYC	QQWSSNP
052	QIVLSQSPAILSASPGERVTLTC	RASS.......SVSYIQ	WFQQKPGSSPKPWIH	ATSKXAS	GVPARFSGSGSGTSYSLTISRVEAEDAATFYC	QQWSSNP
053	QIVLTQSPXIMSASPGEKVTMTC	SASS.......SVSFMQ	WYQQKSGTSPKRWIY	HTSKLAS	GVPARFSXSGXGTSYSLTITSMEAEDAATYYC	QQXSXNP
054	QIVLSQSPAILSASPGEQVTMTC	RASS.......SVSYMH	WYQQKPGSSPKPWIY	ATSNXAS	GVXARFSGSGSGTSYSLTISRVEAEDAATYYC	QQWSSNP

B

	FR1	CDR1	FR2	CDR2	FR3	CDR3
	10 20	30	40	50	60 70 80	90
055	ENVLTQSPXIMSXSXGXKVTMTC	SARS.......SISYMH	WYQQXSSTSXKLWIY	DTSKXAS	XVPXXFSXSGXXNSYSLTISSMEXEDVATYYC	FXGSGYP
056	QIVLSQSPAILSASPGEKVTMTC	RASS.......SVSYIQ	WFQQKPGSSPKPWIS	VTSNLAS	GVPARFSXSGSGTSYSLTISRVEAEDAATYYC	QQWRSNP
057	QIVLSQSPAILSASPGEKVTMTC	RASS.......SVSYIH	WYQQKPGSSPKPWIY	ATSNLAS	GVPVRFSGSGSGTSYSLTINRVEAEDAATYYC	QQWSSNP
058	DIQMTQSPASLSASVGETVTITC	RASG......NIHNYLA	WYQQKQGKSPQLLVY	NAKTLAD	GVPSRFSGSGSGTQYSLKINSLQPEDFGSYYC	QHFWSTP
059	DIQMTQSPASLSVSVGETVTITC	RASE......NIYSNLA	WLFSRNRENPPSLVY	AATNLAD	GVPSRFSGSGSGTQYSLKINSQQPEDFGSYYC	QHFWSAP
060	DIQMTQSPASLSASVGXTVTITC	RASE......NIYSYLA	WYQQKQGKSPQLLVY	NAKTLAE	GVPSRFSGSGSGTQFSLKINSLQPEDFGSYYC	QHHYVTP
061	DIQMTQSPASLSASVGXTVTITC	RASE......NIYSYLA	WYQQKQGKSPQLLVY	NAKTLPE	GVPSRFSGSGSGTQFSLKINSLQPEDFGSYYC	QHHYGPP
062	DVQITQSPSYLAASPGETITINC	RASK......SISKYLA	WYQEKPGKTNKLLIY	SGSTLQS	GIPSRFSGSGSGTDFTLTISSLEPEDFAMYYC	QQHNEYP
063	DVQMTQSPSSLSASLGERVSLTC	QASQ......SINNFLK	WFQQTLGKTARLLIY	GANKLED	GVPSRFSGTGYGTDFTFTISSQEEEDVSTYFC	LQHRYLP
064	DIQMTQSPSSLSASLGERVSLTC	RASQ......DIGSSLN	WLQQEPDGTIKRLIY	ATSSLDS	GVPKRFSGSRSGSDYSLTISSLESEDFVDYYC	LQYASSP
065	DIQMTQSPSSLSASLGERVSLTC	RASQ......DIHGYLN	LFQQKPGETIKHLIY	ETSNLDS	GVPKRFSGSRSGSDYSLIIGSLESEDFADYYC	LQYASSP
066	DIQMTQSPSSLSASLGERVSLTC	RASQ......EISGYLS	WLQQKPDGTIKRLIY	AASTLDS	GVPKRFSGSRSGSDYSLTISSLESEDFADYYC	LQYLSYP
067	DIQMIQSPSSMFGSLGDRVSLSC	RASQ......GIRGNLD	WYQQKPGGTIKLLIY	STSNLNS	GVPSRFSGSGSGSDYSLTISSLESEDFADYYC	LQRNAYP
068	DIKMTQSPSSMYASLGERVTITC	KASQ......DINSYLS	WFQQKPGKSPKTLIY	RANRLVD	GVPSRFSGSGSGQDYSLTISSLEYEDMGIYYC	LQYDEFP
069	DIKMTQSPSSMYASLGERVTISC	KASQ......DINSYLT	WFQQKPGKSPKTLLY	RTKRLVD	GVPSRFSGSGSGQEYSLTISSLEYEDMGIYFC	LQYDEFL
070	DIKMTQSPSSMYASLGERVTITC	KASQ......DIKSYLS	WYQQKPWXSPKTLIY	YATSLAD	GVPSRFSGSGSGXDYSLTISSLESDDTATYYC	LQHGESP
071	DIQMTQTTSSLSASLGDRVTISC	RASQ......DISNYLN	WYQQKPDGTVKLLIY	YTSRLHS	GVPSRFSGSGSGTDYSLTISNLEQEDIATYFC	QQGNTLP
072	DIQMTQTTSSLSASPGDRVTISC	RTSQ......DISNFLY	WFQQKSDGTVKLLIY	YTSRUHS	GVPSRFSGSGSGTDYSFTINNLEUEDVATYSU	QQGI
073	DIQMTQSPSSLSASLGGKVTITC	KASQ......DINKYIA	WDQHKPGKGPRLLIH	YTSTIEP	GIPSRFSGSGSGRDYSFSISNLEPEDIATYYC	LQYDNLP
074	DIQMTQSPSSLSASLGGKVTITC	KASQ......DINKYLA	WYQHKPGKGPRLLIH	YTSTLQP	GIPSRFSGSGSGRDYSFSISNLDAEEIATYYC	LQYDSLY
075	DIVITQDELSNPVTSGESVSISC	RSSKSLLYK.DGKTYLN	WFLQRPGQSPQLLIY	LMSTRAS	GVSDRFSGSGSGTDFTLEISRVKAEDVGVYYC	QQLVEYP
076	DIVMTQAAFSNPVTLGTSASISC	RSSKSLLHS.SGNTYLY	WFLQKPGQSPQLLIY	YISNLAS	GVPDRFSGSGSGTDFTLRISRVEAEDVGVYYC	MQGLEYP
077	DIVMTQAAFSNPVTLGTSASISC	RSSKSLLHS.NGITYLY	WYLQKPGQSPQLLIY	QMSNLAS	GVPDRFSSSGSGTDFTLRISRVEAEDVGVYYC	AQNLELP
078	DIVMTQAAPSVSVTPGESVFISC	RSSKSLLHS.NGNTYLY	WYLQRPGQSPQLLIY	RMSNLAS	GVPDRFSGSGSGTAFTLRISRVEAEDVGIYYC	MQHLEYP
079	DIVMTQATPSVSVTPGESVFISC	RSSKSLLYI.NGNTYLY	WYLQRPGQSPQLLIY	RMSYLAS	GVPDRFSGSGSGTAFTLRISRVEAEDVGIYYC	MQHLEYP
080	DIVMTQAAPSVPVTPGESVSVSC	RSSKSLLHS.NGNTYLY	WFLQRPGQSPQLLIY	RMSNLAS	GVPDRFSGSGSGTAFTLRISRVEAEDVGVYYC	MQHLEYP
081	DIVMTQAAFSNPVTLGTSASISC	RSSKNLLHS.NGITFLY	WYLQRPGQSPQLLIY	RVSNLAS	GVPNRFSGSESGTDFTLRISRVEAEDVGVYYC	AQLLEL
082	DVVMTQTPLSLPVSLGDQASISC	RSSQSLVHS.NGNTYLH	WYLQKPGQSPKLLIY	KVSNRFS	GVPDRFSGSGSGTDFTLKISRVEAEDLGVYFC	SQSTHVP
083	DVLMTQTPLSLPVSLGDQASISC	RSSQSIVHS.NGNTYLE	WYLQKPGQSPKLLIY	KVSNRFS	GVPDRFSGSGSGTDFTLKISRVEAEDLGVYYC	FQGSHVP
084	DAVMTQTPLSLPVSLGDQASISC	RSSQSLENS.NGNTYLK	WYLQKPGQSPQLLIY	RVSNRFS	GVLDRFSGSGSGTDFTLKISRVEAEDLGVYFC	LQVTHAP
085	DVVVTQTPLSLPVSFGDQVSISC	RSSQSLATS.HGITYLS	WYLHKPGQSPQLLIY	GISNRFS	GVPDRFSGSGSGTDFTLKINTIKPEDLGMYYC	LQGSHQP
086	DVVMTQTPLSLPVSLGDQASISC	RSSQSIVHS.NGNTYLY	WYLQKPGQSPKLLIY	RVSNRFS	GVPDRFSGSGSGTDFTLNISRVEAEDMGVYYC	FQGTHVP
087	DAVMTQTPLSLPVSLGDQASISC	RSSQSIENS.NGNTYLN	WYLQKPGQSPRLLIY	RVSNRFS	GVLDRFSGSGSGTDFTLKISRVEAEDLGVYFC	LQVTHVP
088	DIVMTQTPLSLPVSLGDQASISC	RSSQSIVIS.NGFTYLE	WYLQKPXXXXKLLIY	GISNRFS	GVPDRFSGSGSGTDFTLKISRVEAEDVGIYYC	FQGIHVP
089	DVVMTQTPLTLSVTIGQPASISC	KSSQSLLDS.DGKTYLN	WLLQRPGQSPKRLIY	LVSKLDS	GVPDRFTGSGSGTDFTLKISRVEAEDLGVYYC	WQGTHFP
090	DIVMTQSPSSLAMSVGQKVTMSC	KSSQSLLNSSNQKNYLA	WYQQKPGQSPKLLVY	FASTRES	GVPDRFIGSGSGTDFTLTISSVQAEDLADYFC	QQHYSTP
091	GIVMSQSPSSLAVSVGEKVTMSC	KSSQSLFYSSNQKNSLA	WYQQRPGQSPKLLIY	WASTRES	GVPDRFTGSGSGTDFTLTISSVKAEDLAVYYC	QQYYSYP
092	DIVMTQSPSSLTVTAGEKVTMSC	KSSQSLLNSGNPKNYLT	WYQQKPGQPPKLLIY	WASTRES	GVPDRFTGSGSGTDFTLTISSVQAEDLAVYYC	QNDYSYP
093	DIVMTQSPSSLSVSAGEKVTMSC	KSSQSLLNSGNQKNYLA	WYQQKPGQPPKLLIY	GASTRES	GVPDRFTGSGSGTDFTLTISSVQAEDLALYYC	QNDHTYP
094	NIMMTQSPSSLAVSAGEKVTMSC	KSSQSVLYSSNQKNYLA	WYQQKPGQSPKLLIY	WASTRES	GVPDRFTGSGSGTDFTLTISSVQAEDLAVYYC	HQYLSS
095	DVVMSQSPSSLAVSAGEKVAVSC	KSSQSLSTVEPERSYLA	WYQQKPGQSPKLLIY	WASTRES	GVPDRFTGSGSGTDFTLTISSVRAEDLAVYYC	KNLMDLP
096	XIVMTQSXXSLSVSAGXKVTMSC	KSSQXLXNSXVKRTNLA	WYXKKPGQPXKLLIS	VDARPXH	GVPDRFTGSGSGXDYTLTVSSVKAEDLALYYC	QQHYXYP
097	DIVMTQSPTFLAVTASKKVTISC	TXSESLYSSKHKVHYLA	WYQKKPEQSPKLLIY	GASNRYI	GVPDRFTGSGSGTDFTLTISSVQVEDLTHYYC	AQFYSYP
098	SIVMTQTPKFLLVSAGERVTITC	KASQ......SVSNDVA	WYQQKPGQSPKLLIY	YASNRYT	GVPDRFTGSGYGTDFTFTISTVQAEDLAVYFC	QQDYSSP
099	SIVMTQTPKFLPVSAGDRVTMTC	KASQ......SVGNNVA	WYQQKPGQSPKLLIY	YASNRYT	GVPDRFTGSGSGTDFTFTISSVQVEDLAVYFC	QQHYSSP
100	DIVMTQSHKFMSTSVGDRVSITC	KASQ......DVGAAIA	WYQQKPGQSPKLLIY	WASTRHT	GVPDRFTGSGSGTDFTLTISNVQSDDLADYFC	QQYSGYP
101	DIVMTQSQKFMSTSVGDRVSITC	KASQ......NVGTAVA	WYQQKPGQSPKLLIY	SASNRYT	GVPDRFTGSGSGTDFTLTISNMQSEDLADYFC	QQYSSYP
102	DIVMTQSQKFMSTSVGDRVSVTC	KASQ......NVVTNVA	WYQQTPGQSPKALIY	SASYRYS	GVPDRFSGSGSGTDFTLTISNVQSGDLAEYFC	QQYNSYP
103	DIVMTQSQKFMSTSVGDRVSITC	KASQ.......NVRTAV	WYQQKPGQSPKALIY	LASNRYT	GVPDRFTGSGSGTDFTLTITNVQSEDLADYFC	LQHWNYP
104	DIVMTQSHKFMSTSVGDRVSVTC	KASQ......NVGTNVA	WYQQKPGQSPKALIY	SASYRYS	GVPDRFTGSGSGTDFTLTISNVQSEDLAEYFC	QQYNSCP
105	DIVMTQSQKFMSTSVGDRVSVTC	KASQ......NVGTNVA	WYQQKPGQSPKALIY	SASYRYS	GVPDRFTGSGSGTDFTLTISNVQSEDLAVYFC	QQYNSYP
106	DIVMTQSHKFMSTSVGDRVSITC	KASQ......DVSTTVA	WYQQKPGQSPKLLIY	SASYRYT	GVPDRFTGSGSGTDFTFTISSVQAEDLAVYYC	QQHYSTP
107	DIVMTQSHKFMSTSVGDRVSITC	KASQ......DVSTAVA	WYQQKPGQSPKLLIY	SASYRYT	GVPDRFTGSGSGTDFTFTISSVQAEDLAVYYC	QQHYSTP
108	NIVMTQSPKSMSVSVGERVTLTC	KASE......NVVTYVS	WYQQKPEQSPKLLIY	GASNRYT	GVPDRFTGSGSATDFTLTISSVQAEDLADYHC	GQGYSYP
109	NIVMTQSPKSMSMSVGERVTLSC	KASE......NVGTYVS	WYQQKPEQSPKLLIY	GASNRYT	GVPDRFTGSGSATDFTLTISSVQSEDLADYFC	GQSYSYP

Fig. 2. Amino acid sequences deduced from 109 V_k nucleic acid sequences contained in the V_k database. *Dots* have been introduced to maximize homology; X, undetermined amino acids. Remainder of legend as for Figure 1.

Table 1. V_k nucleic acid sequence database*

Code[†]	V_k[‡]	Group[‡]	Subgroup[‡]	Spec[‖]	Strain[§]	Class	Ref[#]
001–005	21	III	21	G	BALB/c	N/A	(1)
006				HA	BALB/c	IgG	(2)
007, 008				RF	MRL/*lpr*	IgG	(3)
009				nf	BALB/c	N/A	(4)
010	23	V	23	G	BALB/c	N/A	(5)
011				nf	N/R	N/A	(6)
012				RNA	$(NZB \times W)F_1$	IgG	(7)
013				nf	MRL/n	N/A	(8)
014, 015				Anti-ID	C57BL/6	IgG	(9)
016				CK	BALB/c	IgG	(10)
017–019	4/5	IV	4, 5	G	BALB/c	N/A	(11)
020				G	BALB/c	N/A	(12)
021				histone	MRL/*lpr*	IgG	(13)
022				DNA	MRL/*lpr*	IgM	(14)
023, 024				OX	BALB/c	IgG	(15)
025				OX	BALB/c	IgG	(16)
026				RF	BALB/c	IgM	(17)
027				ALP	BALB/c	IgG	(18)
028–032		VI	4	G	BALB/c	N/A	(11)
033–038				CaAg	BALB/c	IgG	(19)
039				unknown	BALB/c	IgM	(20)
040				RF	MRL/*lpr*	IgM	(3)
041				CD20	BALB/c	IgG	(21)
042, 043	4/5	VI	4	OX	BALB/c	IgG	(22)
044–048				OX	BALB/c	IgG	(23)
049–054				OX	BALB/c	IgG	(16)
055				OX	BALB/c	IgM	(16)
056				OX	BALB/c	IgG	(15)
057				OX	BALB/c	IgM	(15)
058, 059	12/13	V	12–13	G	BALB/c	N/A	(24)
060, 061				Anti-ID	C57BL/6	IgG	(9)
062	RF	V	ambiguous	RF	MRL/*lpr*	IgM	(13)
063	11	V	11	nf	NZB	N/A	(25)
064	9A	V	9	G	BALB/c	N/A	(26)
065				G	BALB/c	N/A	(27)
066				lysozyme	BALB/c	IgG	(28)
067				DNA	$(NZB \times W)F_1$	IgM	(14)
068	9B	V	9	G	BALB/c	N/A	(5)
069				digoxin	A/J	IgG	(29)
070				BrRBC	CBA/J	IgM	(30)
071	10	V	10	G	A/J	N/A	(31)
072				nf	NZB	N/A	(25)
073	38C	V	ambiguous	unknown	C3H/HeN	IgM	(32)
074				HA	BALB/c	IgG	(33)
075	24/25	II	24	G	BALB/c	N/A	(34)
076, 077				G	BALB/c	N/A	(35)
078, 079				RF	C3H/*lpr*	IgA	(3)
080				RF	BALB/c	IgM	(17)
081				GAC	A/J	IgG	(36)
082–084	1	II	1	G	BALB/c	N/A	(37)
085				dextran	BALB/c	IgA	(38)
086				GAT	BALB/c	IgG	(39)
087				RF	BALB/c	IgM	(17)
088				Anti-ID	BALB/c	IgG	(40)
089	2	II	2	DNA	$(NZB \times W)F_1$	IgM	(14)
090	8	I	8	DNP	BALB/c	IgA	(41)
091				HEL	BALB/c	IgG	(28)
092				RF	MRL/*lpr*	IgA	(3)
093, 094				RF	BALB/c	IgM	(17)
095				RF	129/Sv	IgM	(17)
096				HA	BALB/c	IgG	(33)
097	22	I	22	PC	BALB/c	IgA	(42)
098	19/28	V	28	G	BALB/c	N/A	(43)
099				RF	MRL/*lpr*	IgM	(8)
100			14–15–19	CEA	BALB/c	IgG	(44)
101				TNP	BALB/c	IgM	(45)
102				CASA	N/R	IgG	(46)
103				CEA	N/R	IgG	(47)
104–108				RF	BALB/c	IgM	(17)
109				RF	MRL/*lpr*	IgG	(3)

$V_k4/5$ gene family. V_kTrp subgroups V_k4 (groups IV and VI) and V_k5 (group IV) were encoded by highly similar (around 90%) genes forming a gene family, termed *$V_k4/5$*, that was separated from all other V_k sequences by >25% of their nucleotides. This was the largest V_k gene family, composed of approximately 25-50 members, as deduced from RFLP (Kofler et al. 1989) and gene cloning (Even et al. 1985) studies. Fourteen germline genes (ten *V_k4* and four *V_k5* genes) have been isolated thus far (Even et al. 1985, Höchtl et al. 1982). *$V_k4/5$* genes were found in antibodies specific for galactan (Heller et al. 1987), oxazolone (Kaartinen and Maekelae 1987, Berek and Milstein 1987), dextran (Sikder et al. 1985, Akolkar et al. 1987), the lymphocyte surface marker CD20 (Liu 1987b), alprenolol (Nahmias et al. 1988), red blood cells (Pennell et al. 1988), and DNA, histone, and Ig self antigens (Shlomchik et al. 1987c, Kofler et al. 1987b, Kofler et al. 1988a, Shlomchik et al. 1987b).

$V_k12/13$ gene family. The sequences encoding V_k12–13 proteins (group V) formed another well-defined family that corresponded to all V_k12–13 subgroup proteins (Kabat et al. 1987, Potter et al. 1982). Two germline genes have been published (Nishioka and Leder 1980, Seidman et al. 1978), one of which (*K2*) may be involved in the nitrophenyl-specific anti-idiotypic response (Sablitzky and Rajewsky 1984). A more distant *$V_k12/13$* gene encoded anti-idiotypic light chains in the GAT

---

* Only sequences encoding the entire mature V_k region and differing by >4 bp are contained in this database (see Methods).

[†] Codes of V_k genes are given in legend to Figure 1.

[‡] V_k, V_k gene family (this report); Group, V_k protein groups (Kabat et al. 1987); Subgroup, V_kTrp35 subgroups (Potter et al. 1982).

[‖] Abbreviations: Spec, specificity; Ref, references; G, germline sequence; N/A, not applicable; HA, influenza hemagglutinin; RF, rheumatoid factor; nf, nonfunctional; N/R, not reported, Anti-ID, idiotype-specific antibody; CK, creatine kinase; OX, 2-phenyloxazolone; ALP, alprenolol; CaAg, carbohydrate antigen on human carcinoma cells; CD20, lymphocyte surface marker; BrRBC, bromelain-treated red blood cells; GAC, group A carbohydrate; GAT, Glu^{60} Ala^{30} Tyr^{10} polypeptide; DNP, dinitrophenyl; HEL, hen egg lysozyme; CEA, carcino-embryonal antigen; TNP, trinitrophenyl; CASA, cancer-associated surface antigen.

[§] Strains and their *Igk* haplotypes (Kofler et al. 1989): *Igk*^{a}: MRL/*lpr*, MRL/n; *Igk*^{b}: NZB; *Igk*^{c}: BALB/c, C57BL/6, A/J, C3H, CBA/J, 129/Sv, NZW.

[#] References: 1, (Heinrich et al. 1984); 2, (Clarke et al. 1985); 3, (Shlomchik et al. 1987c); 4, (Strohal et al. 1987); 5, (Pech et al. 1981); 6, (Altenburger et al. 1980); 7, (Eilat et al. 1988); 8, (Kofler et al. 1989); 9, (Sablitzky and Rajewsky 1984); 10, (Buckel et al. 1987); 11, (Even et al. 1985); 12, (Höchtl et al. 1982); 13, (Kofler et al. 1987b); 14, (Kofler et al. 1988); 15, (Berek et al. 1985); 16, (Berek et al. 1987); 17, (Shlomchik et al. 1987a); 18, (Nahmias et al. 1988); 19, (Liu et al. 1987); 20, (Parslow et al. 1984); 21, (Liu 1987); 22, (Kaartinen et al. 1983); 23, (Griffiths et al. 1984); 24, (Seidman et al. 1978); 25, (Kelley et al. 1985); 26, (Seidman et al. 1979); 27, (Max et al. 1980); 28, (Darsley and Rees 1985); 29, (Near and Haber 1989); 30, (Reininger et al. 1987); 31, (Sanz and Capra 1987); 32, (Campbell 1987); 33, (Meek et al. 1989); 34, (Selsing and Storb 1981); 35, (Joho et al. 1984); 36, (Lutz and Davie 1988); 37, (Corbet et al. 1987); 38, (Borden and Kabat 1987); 39, (Schiff et al. 1983); 40, (Ollier et al. 1985); 41, (Riley et al. 1986); 42, (Kwan et al. 1981); 43, (Boyd et al. 1986); 44, (Cabilly et al. 1984); 45, (Hawley et al. 1982); 46, (Sahagan 1986); 47, (Beidler et al. 1988).

Table 2. Correlation between V_k gene families and V$_\text{k}$ protein groups and subgroups*.

V_k gene family	V$_\text{k}$Cys subgroup	V$_\text{k}$Trp subgroup	V$_\text{k}$ protein group
21	21	21	III
23	23	23	V
4/5	4	4	IV, VI
	5	5	IV
12/13	12, 13	12–13	V
RF	ambiguous assignment		V
11	11	11	V
9A	9	9	V
9B	9	9	V
10	10	10	V
38C	ambiguous assignment		V
24/25	24	24	II
	25	25	II, I
1	1, 3, 26	1	II
2	2	2	II
8	8	8	I
22	22	22	I
19/28	–	28	V
	14, 15, 19	19	V

* Relatedness between V_k gene families and V$_\text{k}$Trp subgroups 20 and 27, and V$_\text{k}$Cys subgroups 6, 7, 16, 17, and 18 (for which only partial protein sequences are known), could not be determined.

(Glu^{60} Ala^{30} Tyr^{10}) system (Ollier et al. 1985). In RFLP analyses, two strongly and several weakly hybridizing restriction fragments were observed (Kofler et al. 1989). Whether the latter corresponded to additional, more distant, V_k*12/13* germline genes or are due to high similarity (>80%) in portions of the probe with other V_k genes (particularly those of V_k gene families *9A, 9B, 10,* and *11*) remains to be determined.

V_k*RF gene family.* The MRL-RF24 V$_\text{k}$ protein (Kofler et al. 1987b), a member of the large protein group V, had 12 mismatches up to Trp35 from two V$_\text{k}$12–13 proteins (K2 and MOPC129), but differed from the remaining V$_\text{k}$12–13 proteins (and all other V$_\text{k}$ proteins) by >12 residues. Thus, this protein could not be unambiguously assigned to known V$_\text{k}$ subgroups. Its nucleic acid sequence differed from all V$_\text{k}$ sequences by >25%, thus forming a distinct V_k gene family, termed V_k*RF*. Used as a probe, this gene identified a single restriction fragment that was absent in haplotype *Igk^f* (Kofler et al. 1989). The corresponding (as yet uncloned) germline gene probably also encoded a BALB/c (Bruck et al. 1986) and a C57BL/6 (Sablitzky and Rajewsky 1984) idiotype-specific antibody, as well as an (NZB × NZW)F$_1$ DNA-specific autoantibody (Eilat et al. 1988).

V_k*11, 9A, 9B, 10, and 38C gene families.* The V_k gene families discussed thus far were clearly separated from all other V_k genes by >25% overall sequence dissimilarity and in this respect resembled V_H gene families. The following five gene families, distantly related to V_k*12/13* and V_k*RF*, were less well separated from one another.

V_k*11 gene family.* For this gene family with four to six germline genes by RFLP analysis (Kofler et al. 1989), a single nucleic acid sequence corresponding to a nonfunctional rearrangement from an NZB myeloma (Kelley et al. 1985) was present in the data bank. This sequence fulfilled protein assignment criteria for V$_\text{k}$ protein subgroups 9, 10, and 11; however, it best matched V$_\text{k}$11 proteins. Comparisons with the entire data bank (including some V$_\text{k}$9 and V$_\text{k}$10 sequences) revealed matches of only 76% or less at the nucleic acid level, making this sequence the prototype for the V_k*11* gene family. V$_\text{k}$11 proteins were observed in the beta 2, 1 fructosan response (Kabat et al. 1987).

V_k*9A gene family.* The V$_\text{k}$9 protein subgroup, another member of the large protein group V (Potter et al. 1982), comprised sequences that, at the nucleic acid level, fell into two distinct gene families, termed V_k*9A* and V_k*9B*. The V_k*9A* gene family included two germline genes (Seidman et al. 1979, Max et al. 1980), one of which may be expressed in hen egg lysozyme antibodies (Darsley and Rees 1985). Another expressed V_k*9A* gene from an NZB × NZW F$_1$ anti-DNA IgM (Kofler et al. 1988) was only 88% similar to the other germline gene and probably derived from an unknown V_k*9A* germline gene. In addition, V_k*9A* genes have been observed in GAT- idiotype-specific antibodies (Ollier et al. 1985).

V_k*9B gene family.* The T1 sequence and its germline counterpart, V-L6 (Pech et al. 1981), both assigned to the V$_\text{k}$9 protein subgroup (Potter et al. 1982), differed from V$_\text{k}$9A (and all other V$_\text{k}$) nucleic acid sequences by >20% and, hence, formed a separate family, termed V_k*9B*. Genes from this family encoded antibodies specific for digoxin (Panka and Margolies 1987, Near and Haber 1989) and *Escherichia coli* (Pennell et al. 1988), and bromelain-treated red blood cell autoantibodies from lupus and normal mice (Reininger et al. 1987).

V_k*10 gene family.* This family corresponded to the V$_\text{k}$10 subgroup (protein group V). RFLP data suggested two to three V_k*10* germline genes (Kofler et al. 1989), one of which has been cloned (Sanz and Capra 1987, Wysocki et al. 1987) and probably encoded arsonate (Manser et al. 1987a, Meek et al. 1987), oxazolone (Berek et al. 1985), oligosaccharide (Matsuda and Kabat 1989), bromelain-treated red blood cell (Pennell et al. 1988) and RF-like (Shlomchik et al. 1987c) antibody responses. A more distant V$_\text{k}$10 sequence, with multiple in-frame stop codons, has been observed as a nonfunctional allele of an

NZB myeloma (Kelley et al. 1985), and might correspond to one of the uncloned V_k10 germline genes.

V_k38C gene family (tentative). The very similar (97%) sequences encoding the 38C13 lymphoma (Campbell 1987) and the VM113 anti-hemagglutinin hybridoma (Meek et al. 1989) light chains, respectively, were >20% different from all other $\mathrm{V_k}$ nucleic acid sequences in the database and, hence, could not be assigned to any V_k gene family; the closest matches (77–78%) were observed with a V_k10 germline gene (Sanz and Capra 1987, Wysocki et al. 1987). At the amino acid level, members of four $\mathrm{V_k}$Trp subgroups ($\mathrm{V_k}9$, $\mathrm{V_k}10$, $\mathrm{V_k}11$, and $\mathrm{V_k}12/13$) exhibited equally distant relatedness (nine and more residues difference in the N-terminal 35 amino acids), making unambiguous assignment at the protein level impossible. Whether these sequences were the representatives of a new V_k gene family or corresponded to highly mutated (V_k10) genes remains to be determined.

$V_k24/25$, V_k1, and V_k2 gene families. The next three families were grouped together based on sequence similarity of up to 78% between $V_k24/25$ members and V_k1 and V_k2 genes, respectively, and because the overall similarity between V_k2 and some V_k1 genes exceeded 80%. The latter observation, i. e., similarity of >80% between some, but not all, members of two gene families, obviously constitutes a problem in this type of V_k gene classification (see below).

$V_k24/25$ gene family. Originally, only a single V_k24 germline gene (involved in the phosphocholine response; Malipiero et al. 1987, Gearhart and Bogenhagen 1983) had been reported (Selsing and Storb 1981). Other investigators have cloned this, a related pseudogene, and two additional V_k24 germline genes (Joho et al. 1984). The latter were only about 82–83% similar to the V_k24 prototype and may have encoded *Streptococcus* group A carbohydrate antibody light chains previously assigned to the $\mathrm{V_k}25$ subgroup (Lutz and Davie 1988). Hence, these two $\mathrm{V_k}$Trp subgroups (protein group II) were probably encoded by distant members of a single V_k gene family. In addition to the four cloned $V_k24/25$ germline genes, evidence was obtained for the presence of at least two more germline genes in this family: firstly, RFs from autoimmune and normal mice (Shlomchik et al. 1987a, 1987c) expressed V_k24 genes very similar to each other, but >30 bp different from the closest V_k24 germline gene, suggesting an additional germline gene; secondly, since all cloned $V_k24/25$ genes had 40 codons up to Trp35, the germline gene encoding Hy2.5.13 with 41 N-terminal amino acids (Kabat et al. 1987) has yet to be isolated.

V_k1 and V_k2 gene families. Protein subgroups $\mathrm{V_k}1$ (already previously condensed with Cys23 subgroups $\mathrm{V_k}3$ and $\mathrm{V_k}26$; Potter et al. 1982) and $\mathrm{V_k}2$ were encoded by sequences that, using a stringent family definition, precluded classification into either a single, or two distinct, gene families; all $\mathrm{V_k}1$ nucleic acid sequences were >80% similar, yet the three almost identical $\mathrm{V_k}2$ nucleic acid sequences reported (Akolkar et al. 1987, Kofler et al. 1988, Panka et al. 1988) shared up to 81.7% similarity with some, but only about 75% with other, V_k1 members. Moreover, sequence similarity in the 3′ portion of several V_k1 and V_k2 genes was around 90%. These two "gene families" were, therefore, partially overlapping. However, for reasons of clarity, we have retained them as separate V_k gene families.

Three V_k1 germline genes (Corbet et al. 1987) and approximately 40 expressed $\mathrm{V_k}1$ sequences have been reported. With the exception of an anti-dextran V_k gene (*W3129;* Borden and Kabat 1987) with >15% differences from any known V_k1 gene, all expressed sequences were highly homologous to one of the above germline genes, suggesting that the total V_k1 germline gene number may not exceed four. A more direct complexity estimate in our previous RFLP analysis was hampered by cross-hybridization of the $\mathrm{V_k}1$ probe to non-V_k1 genes due to >80% sequence similarity in the 3′ region of V_k1 and other V_k genes (see below and Kofler et al. 1989). V_k1 genes were used in a variety of responses to foreign and self antigens (reviewed by Schiff et al. 1988, Kofler et al. 1987a). V_k2 germline genes have not yet been reported; the three expressed sequences encoded antibodies to dextran (Akolkar et al. 1987), digoxin (Panka et al. 1988), and DNA (Kofler et al. 1988).

V_k8, V_k22, and $V_k19/28$ gene families. The following three gene families were separated from each other by >20%, and from all other V_k genes by >25%, overall sequence similarity; however, large portions (codons 35 to 94) of their genes had between 80% and 89% common nucleotides, leading to extensive cross-hybridizations (Kofler et al. 1989).

V_k8 gene family. All sequences encoding $\mathrm{V_k}$Trp subgroup $\mathrm{V_k}8$ (protein group I) were around 90% similar and shared up to 78% of their nucleotides with $V_k19/28$ and V_k22 genes. Similarity in codons 35–94 was even higher, reaching 87% with V_k28 genes. The complexity of this gene family was difficult to assess by RFLP analyses due to possible cross-hybridization, however, at least half of the 13–20 fragments hybridizing to a V_k8 probe probably belonged to this large family (Kofler et al. 1989). V_k8 genes encoded antibodies to phosphocholine (Malipiero et al. 1987), dinitrophenyl (Riley et al. 1986), and hen egg lysozyme (Darsley and Rees 1985), as well as RF-like (Shlomchik et al. 1987a, 1987c) and DNA-specific (Eilat et al. 1988) autoantibodies.

V_k22 gene family. The only two, almost identical, $\mathrm{V_k}22$ (protein group I) sequences available for comparison, S107A (Kwan et al. 1981) and HPCA97 (Berek 1984), revealed between 80% and 89% similarity with a large portion (codons 35 to 94) of all $V_k19/28$ genes. The remaining nucleotides were, however, only <70% similar, resulting in an overall similarity of 72%–75%, thus refuting assignment of V_k22 and $V_k19/28$ genes to a common gene family. Similarity with V_k8 genes was in the range of 75%–77% and mismatches were distributed evenly over the entire gene. RFLP analyses suggested one to two V_k22 germline genes; additional weak restriction fragments hybridizing to a V_k22 probe on Southern blots probably corresponded to genes from the $V_k19/28$ and V_k8 families (Kofler et al. 1989). V_k22 genes encoded phosphocholine antibodies (Malipiero et al. 1987).

$V_k19/28$ gene family. Sequences encoding $\mathrm{V_k}$Trp subgroups 19 (comprising $\mathrm{V_k}$Cys14 and 15 sequences) and 28 were >80% similar among each other and differed from all other V_k genes (except V_k8 and V_k22, see above) by >25%. Thus, they were combined to a single V_k gene family, which was termed $V_k19/28$. However, this V_k gene family (like some other V_k gene families, see below) behaved atypically in nucleic acid hybridization studies as compared to V_H gene families: different DNA probes from this family, i. e., a V_k19 and a V_k28 probe, did not hybridize to an identical, but to an overlapping, set of restriction fragments (Kofler et al. 1989). This could be explained by cross-hybridization of the V_k28, but not the V_k19, probe with V_k8 genes.

RFLP data suggested four to six $V_k19/28$ germline genes (Kofler et al. 1989), one of which, a V_k28 germline gene, also known as V_k*Ser,* from haplotypes *Igk-VSer*^{a}, *Igk-VSer*^{b}, *Igk-VSer*^{c}, and *Igk-VSer*^{d}, has been cloned (Boyd et al. 1986, Ponath et al. 1989). $V_k19/28$ genes encoded antibodies to trinitrophenyl (Hawley et al. 1982), carcinoembryonic antigen (Cabilly et al. 1984, Beidler et al. 1988), human breast/lung/colon cancer cells (Sahagan 1986), influenza hemagglutinin (Meek et al. 1989), and an RNA-specific (Eilat et al. 1988) and some RF-like autoantibodies (Kofler et al. 1989, Shlomchik et al. 1987a, 1987c).

Relatedness between V_k gene families and implications for nucleic acid hybridization assays with V_k probes

Figure 3 shows the relatedness between different V_k gene families as reflected by overall nucleic acid sequence similarity. A significant difference from V_H gene families was apparent, since the latter are generally more distantly

	Vk21	Vk23	Vk4/5	Vk12/13	VkRF	Vk11	Vk9A	Vk9B	Vk10	Vk38c	Vk24/25	Vk1	Vk2	Vk8	Vk22
Vk23	65-67														
Vk4/5	62-71	61-64													
Vk12/13	62-64	60-62	60-64												
VkRF	63-66	65-67	59-63	67-70											
Vk11	61	63	58-62	69-71	71										
Vk9A	61-67	61-70	63-67	70-71	69-73	69-72									
Vk9B	63-67	61-65	62-67	70-73	71-72	74-76	73-76								
Vk10	57-67	63-66	60-68	65-71	71-73	71-72	74-77	71-75							
Vk38c	60-65	64-65	59-64	66-70	71	69	69-72	73	72-77						
Vk24/25	63-67	59-60	57-65	57-61	60-62	57	58-63	59-61	58-60	58-59					
Vk1	62-68	62-67	58-62	59-61	59-62	57-62	57-64	58-65	57-63	58-61	70-78				
Vk2	63-66	61	56-60	56-58	61	60	59-61	60-63	58-59	57	74-76	73-81			
Vk8	64-70	61-65	62-66	60-63	65-66	61-63	63-65	65	60-61	61-62	62-68	63-69	64-67		
Vk22	64-65	61-62	60-66	58-60	61	58	60-61	60-62	57-59	58	60-63	62-64	63	75-77	
Vk19/28	65-67	62-66	63-70	58-62	62-66	63-65	62-67	65-70	60-66	59-64	58-65	60-66	60-65	72-78	72-75

Fig. 3. Sequence similarity between different V_k gene families; comparison of known germline genes and derivatives of putative germline genes (i. e., sequences differing from known germline genes by >10% and primarily of the IgM isotype). Indicated are the highest and lowest percentages of nucleic acid sequence similarity between members from two families; single percentage resulted from comparisons yielding identical percentages. *Shading intensities* highlight increased overall similarity between the respective families.

related by sequence similarity. Obviously, if members from different families are only a few percent less similar than those from within a family, cross-hybridizations might occur, particularly if these differences are not evenly distributed over the entire sequence. As described above, large sequence portions with high degrees of similarity were indeed observed in genes from families 8, 22, and 19/28, and thus explain the previously observed cross-hybridizations between those families. Closer scrutiny of the similarities between portions of $\mathrm{V_k}$ sequences from different families revealed that the 3′ region (particularly codons 57–88, corresponding to frame work region 3) were generally more closely related than the remaining sequence, and this portion might precipitate unexpected cross-hybridizations, even between otherwise distant V_k gene families. For example, V_k*10* and V_k*9A* genes had a 135 bp 3′ sequence with 83% similarity, and V_k*RF* and V_k*9B* genes shared 84% of 103 nucleotides at the 3′ end. As a further complication, different genes from a given family may exhibit more or less cross-hybridizations with genes from other families.

Because of the differences in the organization of V_H and V_k genes, nucleic acid sequence hybridization assays with $\mathrm{V_k}$ DNA probes require particular care in the selection of probes and in data interpretation. While in general any member of a V_H gene family used as a probe will recognize its entire family, but will not cross-hybridize with other families, our previous RFLP analyses and the current study strongly suggest that V_k probes may often behave differently. As a rule, probes devoid of the more "promiscuous" 3′ sequences will be more specific; however, such probes may not always hybridize to all members of their gene families, and therefore require the use of two or more genes to probe the entire family.

V_k germline gene complexity

Another question addressed in this study regards the total number of V_k genes in the genome of inbred mice. We estimated the complexity of known V_k gene families by using RFLP criteria (Kofler et al. 1989) and by taking into account expressed and germline genes identified for each family. Regarding expressed sequences, we assumed that IgM sequences with >6, and IgG sequences with >30 mismatches from known germline genes may have derived from as yet unknown germline genes. Allelic differences were also considered, however this was a minor concern as the majority of sequences in the database (91/109) derived from the same haplotype (*Igk*^{c}).

This approach led to a total of about 70–140 genes (Table 3). Obviously, such estimates need to be taken with caution due to the peculiarities of V_k gene probes discussed above, and to inherent limitations of the RFLP technique (discussed by Kofler et al. 1989). Furthermore, possible additional, as yet uncloned, V_k genes and gene families in the mouse genome have not been included. However, although evidence for some additional V_k genes exists, their number might be limited. For two $\mathrm{V_k}$Trp subgroups, $\mathrm{V_k}$27 (group I) and $\mathrm{V_k}$20 (group VII), nucleic acid sequences have not been identified, but the corresponding V_k gene families may be small since only a single sequence for each subgroup has been reported to date. D'Hoostelaere published another novel V_k gene family (*pC9-26*) with approximately six members as suggested by RFLP analyses (D'Hoostelaere et al. 1988), but whether or not this family related to either of the two subgroups above, or to $\mathrm{V_k}$38C, is unknown. Nevertheless, the large number of responses to foreign and self antigens investigated at the nucleic acid sequence level, and repeated isolation of identical sequences, suggest that the majority of the mouse V_k germline repertoire might now be known.

Table 3. V_k germline gene complexity*

V_k gene family	Germline genes	
	Cloned	Estimated
V_k*21*	5	6-13
V_k*23*	1	2–4
V_k*4/5*	14	25–50
V_k*12/13*	2	2-8
V_k*RF*	–	0-1†
V_k*11*	–	4-6
V_k*9A*	2	4-9
V_k*9B*	1	2
V_k*10*	1	2-3
V_k*38C*	–	?
V_k*24/25*	4	6
V_k*1*	3	4-6
V_k*2*	–	1-6
V_k*8*	–	5–16
V_k*22*	–	1-2
V_k*19/28*	–	4-6
Total	35	66–136

* References to cloned V_k germline genes are given throughout the text.

† The one-member V_k*RF* family is deleted in haplotype *Igk*^{f} mice (Kofler et al. 1989).

Acknowledgments. We thank Dr. G. Wick for critically reviewing, and Ms. M. Kat Occhipinti for editing, this manuscript. The work reported herein was supported by a grant from the Austrian Research Council (Project S-41/06). R. Strohal is supported by a fellowship from the Emil Boral Stiftung and G. Kroemer receives an Erwin Schrödinger Stipend (J0307M) from the Austrian Research Council.

References

Akolkar, P. N., Sikder, S. K., Bhattacharya, S. B., Liao, J., Gruezo, F., Morrison, S. L., and Kabat, E. A.: Different $\mathrm{V_L}$ and $\mathrm{V_H}$ germline genes are used to produce similar combining sites with specificity for alpha (1 – – > 6) dextrans. *J Immunol 138:* 4472–4479, 1987

Alt, F.W., Blackwell, K., DePinho, R.A., Reth, M.G., and Yancopoulos, G.D.: Regulation of genome rearrangement events during lymphocyte differentiation. *Immunol Rev 89:* 5–30, 1986

Altenburger, W., Steinmetz, M., and Zachau, H.G.: Functional and non-functional joining in immunoglobulin light chain genes of a mouse myeloma. *Nature 287:* 603–607, 1980

Beidler, C.B., Ludwig, J.R., Cardenas, J., Phelps, J., Papworth, C.G., Melcher, E., Sierzega, M., Myers, L.J., Unger, B.W., Fisher, M., David, G.S., and Johnson, M.J.: Cloning and high level expression of a chimeric antibody with specificity for human carcinoembryonic antigen. *J Immunol 141:* 4053–4060, 1988

Berek, C.: The D segment defines the T15 idiotope: the immunoresponse of A/J mice to *Pneumococcus pneumoniae. Eur J Immunol 14:* 1043–1048, 1984

Berek, C. and Milstein, C.: Mutation drift and repertoire shift in the maturation of the immune response. *Immunol Rev 96:* 23–41, 1987

Berek, C., Griffiths, G.M., and Milstein, C.: Molecular events during maturation of the immune response to oxazolone. *Nature 316:* 412–418, 1985

Berek, C., Jarvis, J.M., and Milstein, C.: Activation of memory and virgin B cell clones in hyperimmune animals. *Eur J Immunol 17:* 1121, 1987

Borden, P. and Kabat, E.A.: Nucleotide sequence of the cDNAs encoding the variable region heavy and light chains of a myeloma protein specific for the terminal nonreducing end of alpha (1–6) dextran. *Proc Natl Acad Sci USA 84:* 2440–2443, 1987

Boyd, T.M., Goldrick, M.M., and Gottlieb, P.D.: Structural differences in a single gene encoding the V_kSer group of light chains explain the existence of two mouse light-chain genetic markers. *Proc Natl Acad Sci USA 83:* 9134–9138, 1986

Brodeur, P.H. and Riblet, R.: The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur J Immunol 14:* 922–930, 1984

Bruck, C., Co, M.S., Slaoui, M., Gaulton, G.N., Smith, T., Fields, B.N., Mullins, J.I., and Greene, M.I.: Nucleic acid sequence of an internal image-bearing monoclonal anti-idiotype and its comparison to the sequence of the external antigen. *Proc Natl Acad Sci USA 83:* 6578–6582, 1986

Buckel, P., Huebner-Parajsz, C., Mattes, R., Lenz, H., Haug, H., and Beaucamp, K.: Cloning and nucleotide sequence of heavy- and light-chain cDNAs from creatine-kinase-specific monoclonal antibody. *Gene 51:* 13–19, 1987

Cabilly, S., Riggs, A.D., Pande, H., Shively, J.E., Holmes, W.E., Rey, M., Perry, L.J., Wetzel, R., and Heyneker, H.L.: Generation of antibody activity from immunoglobulin polypeptide chains produced in *Escherichia coli. Proc Natl Acad Sci USA 81:* 3273–3277, 1984

Campbell, M.J.: Idiotype vaccination against murine B cell lymphoma. Humoral and cellular responses elicited by tumor-derived immunoglobulin M and its molecular subunits. *J Immunol 139:* 2825–2833, 1987

Clarke, S.H., Huppi, K., Ruezinsky, D., Staudt, L., Gerhard, W., and Weigert, M.: Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *J Exp Med 161:* 687–704, 1985

Corbet, S., Milili, M., Fougereau, M., and Schiff, C.: Two V_k germline genes related to the GAT idiotypic network (Ab1 and Ab3/Ab1′) account for the major subfamilies of the mouse V_k-1 variability subgroup. *J Immunol 138:* 932–939, 1987

Cory, S., Tyler, B.M., and Adams, J.M.: Sets of immunoglobulin V_k genes homologous to ten cloned V_k sequences: implications for the number of germline V_k genes. *J Mol Appl Genet 1:* 103–116, 1981

Darsley, M.J. and Rees, A.R.: Nucleotide sequence of five anti-lysozyme monoclonal antibodies. *EMBO J 4:* 393–398, 1985

Devaux, C., Moinier, D., Mazza, G., Guo, X., Marchetto, S., Fougereau, M., and Rierres, M.: Preferential expression of V_k21E on IdX Ia.7 positive monoclonal anti-I-E antibodies. *J Immunol 134:* 4024–4030, 1985

D'Hoostelaere, L.A., Huppi, K., Mock, B., Mallett, C., and Potter, M.: The Igk L chain allelic groups among the Igk haplotypes and Igk crossover populations suggest a gene order. *J Immunol 141:* 652–661, 1988

Dildrop, R.: A new classification of mouse V_H sequences. *Immunol Today 5:* 85–86, 1984

Eilat, D., Webster, D.M., and Rees, A.R.: V region sequences of anti-DNA and anti-RNA autoantibodies from NZB/NZW F1 mice. *J Immunol 141:* 1745–1753, 1988

Even, J., Griffiths, G.M., Berek, C., and Milstein, C.: Light chain germ-line genes and the immune response to 2-phenyloxazolone. *EMBO J 4:* 3439–3445, 1985

Gearhart, P.J. and Bogenhagen, D.F.: Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc Natl Acad Sci USA 80:* 3439–3443, 1983

Griffiths, G.M., Berek, C., Kaartinen, M., and Milstein, C.: Somatic mutation and the maturation of immune response to 2-phenyloxazolone. *Nature 312:* 271–275, 1984

Hawley, R.G., Shulman, M.J., Murialdo, H., Gibson, D.M., and Hozumi, N.: Mutant immunoglobulin genes have repetitive DNA elements inserted into their intervening sequences. *Proc Natl Acad Sci USA 79:* 7425–7429, 1982

Heinrich, G., Traunecker, A., and Tonegawa, S.: Somatic mutation creates diversity in the major group of mouse immunoglobulin kappa light chains. *J Exp Med 159:* 417–435, 1984

Heller, M., Owens, J.D., Mushinski, J.F., and Rudikoff, S.: Amino acids at the site of V_k-J_k recombination not encoded by germline sequences. *J Exp Med 166:* 637–646, 1987

Höchtl, J., Müller, C.R., and Zachau, H.G.: Recombined flanks of the variable and joining segments of immunoglobulin genes. *Proc Natl Acad Sci USA 79:* 1383–1387, 1982

Joho, R., Gershenfeld, H., and Weissman, I.L.: Evolution of a multigene family of V_k germ line genes. *EMBO J 3:* 185–191, 1984

Kaartinen, M. and Maekelae, O.: Functional analogues of the V_kOx1 gene in different strains of mice: evolutionary conservation but diversity based on V-J joining. *J Immunol 138:* 1613–1617, 1987

Kaartinen, M., Griffiths, G.M., Markham, A.F., and Milstein, C.: mRNA sequences define an unusually restricted IgG response to 2-phenyloxazolone and its early diversification. *Nature 304:* 320–324, 1983

Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., and Gottesman, K.S.: *Sequences of Proteins of Immunological Interest,* 4th edn, US Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, 1987

Kearney, J.F., Radbruch, A., Liesegang, B., and Rajewsky, K.: A mouse myeloma cell line that has lost immunoglobulin expression but permits the construction of antibody-searching hybrid cell lines. *J Immunol 123:* 1548–1550, 1979

Kelley, D.E., Wiedemann, L.M., Pittet, A.C., Strauss, S., Nelson, K.J., Davis, J., Van Ness, B., and Perry, R.P.: Nonproductive kappa immunoglobulin genes: recombinational abnormalities and other lesions affecting transcription, RNA processing, turnover, and translation. *Mol Cell Biol 5:* 1660–1675, 1985

Kofler, R.: A new murine Ig V_H gene family. *J Immunol 140:* 4031–4034, 1988

Kofler, R., Dixon, F.J., and Theofilopoulos, A.N.: The genetic origin of autoantibodies. *Immunol Today 8:* 374–380, 1987a

Kofler, R., Noonan, D.J., Strohal, R., Balderas, R.S., Møller, N.P.H., Dixon, F.J., and Theofilopoulos, A.N.: Molecular analysis of the murine lupus-associated anti-self response: involvement of a large number of heavy and light chain variable region genes. *Eur J Immunol 17:* 91–95, 1987b

Kofler, R., Strohal, R., Balderas, R. S., Johnson, M. E., Noonan, D. J., Duchosal, M. A., Dixon, F. J., and Theofilopoulos, A. N.: Immunoglobulin k light chain variable region gene complex organization and immunoglobulin genes encoding anti-DNA autoantibodies in lupus mice. *J Clin Invest 82:* 852-860, 1988

Kofler, R., Duchosal, M. A., and Dixon, F. J.: Complexity, polymorphism and connectivity of mouse V_k gene families. *Immunogenetics 29:* 65-74, 1989

Kwan, S.-P., Rudikoff, S., Seidman, J. G., Leder, P., and Scharff, M. D.: Nucleic acid and protein sequences of phosphocholine-binding light chains. *J Exp Med 153:* 1366-1370, 1981

Liu, A. Y.: Production of a mouse-human chimeric monoclonal antibody to CD20 with potent Fc-dependent biologic activity. *J Immunol 139:* 3521, 1987

Liu, A. Y., Robinson, R. R., Hellström, K. E., Murray, E. D., Jr., Chang, C. P., and Hellström, I.: Chimeric mouse-human IgG1 antibody that can mediate lysis of cancer cells. *Proc Natl Acad Sci USA 84:* 3439-3443, 1987

Livant, D., Blatt, C., and Hood, L.: One heavy chain variable region gene segment subfamily in the BALB/c mouse contains 500-1000 or more members. *Cell 47:* 461-470, 1986

Lutz, C. T. and Davie, J. M.: Genetics and primary structure of V_k gene segments encoding antibody to Streptococcal group A carbohydrate. Comparison of V_k gene structure with idiotope expression. *J Immunol 140:* 641-645, 1988

Malipiero, U. V., Levy, N. S., and Gearhart, P. J.: Somatic mutation in anti-phosphorylcholine antibodies. *Immunol Rev 96:* 59-74, 1987

Manser, T., Parhami-Seren, B., Margolies, M. N., and Gefter, M. L.: Somatically mutated forms of a major anti-p-azophenylarsonate antibody variable region with drastically reduced affinity for p-azophenylarsonate. By-products of an antigen-drive immune response? *J Exp Med 166:* 1456-1463, 1987a

Manser, T., Wysocki, L. J., Margolies, M. N., and Gefter, M. L.: Evolution of antibody variable region structure during the immune response. *Immunol Rev 96:* 141-162, 1987b

Matsuda, T. and Kabat, E. A.: Variable region cDNA sequences and antigen binding specificity of mouse monoclonal antibodies to isomaltosyl oligosaccharides coupled to proteins. T-dependent analogous of $\alpha(1-->6)$dextran. *J Immunol 142:* 863-870, 1989

Max, E. E., Seidman, J. G., Miller, H., and Leder, P.: Variation in the crossover point of kappa immunoglobulin gene V-J recombination: evidence from a cryptic gene. *Cell 21:* 793-799, 1980

Meek, K., Sanz, Z., Rathbun, G., Nisonoff, A., and Capra, J. D.: Identity of the V_k10-Ars-A gene segments of the A/J and BALB/c strains. *Proc Natl Acad Sci USA 84:* 6244-6248, 1987

Meek, K., Johansson, B., Schulman, J., Bona, C., and Capra, J. D.: Nucleotide changes in sequential variants of influenza virus hemagglutinin genes and molecular structures of corresponding monoclonal antibodies specific for each variant. *Proc Natl Acad Sci USA 86:* 4664-4668, 1989

Nahmias, C., Strosberg, A. D., and Emorine, L. J.: The immune response toward beta-adrenergic ligands and their receptors. VIII. Extensive diversity of V_H and V_L genes encoding anti-alprenolol antibodies. *J Immunol 140:* 1304-1311, 1988

Near, R. I. and Haber, E.: Characterization of the heavy and light chain immunoglobulin variable region genes used in a set of anti-digoxin antibodies. *Mol Immunol 26:* 371-382, 1989

Nishioka, Y. and Leder, P.: Organization and complete sequence of identical embryonic and plasmacytoma kappa V-region genes. *J Biol Chem 255:* 3691-3694, 1980

Ollier, P., Rocca-Serra, J., Somme, G., Theze, J., and Fougereau, M.: The idiotypic network and the internal image: possible regulation of a germ-line network by paucigene encoded Ab2 (anti-idiotypic) antibodies in the GAT system. *EMBO J 4:* 3681-3688, 1985

Panka, D. J. and Margolies, M. N.: Complete variable region sequences of five homologous high affinity anti-digoxin antibodies. *J Immunol 139:* 2385-2391, 1987

Panka, D. J., Mudgett-Hunter, M., Parks, D. R., Peterson, L. L., Herzenberg, L. A., Haber, E., and Margolies, M. N.: Variable region framework differences result in decreased or increased affinity of variant anti-digoxin antibodies. *Proc Natl Acad Sci USA 85:* 3080-3084, 1988

Parslow, T. G., Blair, D. L., Murphy, W. J., and Granner, D. K.: Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. *Proc Natl Acad Sci USA 81:* 2650-2654, 1984

Pech, M., Höchtl, J., Schnell, H., and Zachau, H. G.: Differences between germ-line and rearranged immunoglobulin V_k coding sequences suggest a localized mutation mechanism. *Nature 291:* 668-670, 1981

Pennell, C. A., Arnold, L. W., Haughton, G., and Clarke, S. H.: Restricted Ig variable region gene expression among Ly-1^+ B cell lymphomas. *J Immunol 141:* 2788-2796, 1988

Perlmutter, R. M., Kearney, J. F., Chang, S. P., and Hood, L. E.: Developmentally controlled expression of immunoglobulin V_H genes. *Science 227:* 1597-1601, 1985

Ponath, P. D., Hillis, D. M., and Gottlieb, P. D.: Structural and evolutionary comparisons of four alleles of the mouse immunoglobulin kappa chain gene, *Igk-VSer. Immunogenetics 29:* 249-257, 1989

Potter, M.: Antigen-binding myeloma proteins of mice. *Adv Immunol 25:* 141-211, 1977

Potter, M., Newell, J. B., Rudikoff, S., and Haber, E.: Classification of mouse V_k groups based on the partial amino acid sequence to the first invariant tryptophan: impact of 14 new sequences from IgG myeloma proteins. *Mol Immunol 19:* 1619-1630, 1982

Reininger, L., Ollier, P., Poncet, P., Kaushik, A., and Jaton, J. C.: Novel V genes encode virtually identical variable regions of six murine monoclonal anti-bromelain-treated red blood cell autoantibodies. *J Immunol 138:* 316-323, 1987

Reininger, L., Kaushik, A., Izui, S., and Jaton, J.-C.: A member of a new V_H gene family encodes anti-bromelinized mouse red blood cell autoantibodies. *Eur J Immunol 18:* 1521-1526, 1988

Riley, S. C., Connors, S. J., Klinman, N. R., and Ogata, R. T.: Preferential expression of variable region heavy chain gene segments by predominant 2,4-dinitrophenyl-specific BALB/c neonatal antibody clonotypes. *Proc Natl Acad Sci USA 83:* 2589-2593, 1986

Sablitzky, F. and Rajewsky, K.: Molecular basis of an isogeneic anti-idiotypic response. *EMBO J 3:* 3005-3012, 1984

Sahagan, B. G.: A genetically engineered murine/human chimeric antibody retains specificity for human tumor-associated antigen. *J Immunol 137:* 1066, 1986

Sanz, I. and Capra, J. D.: V_k and J_k gene segments of A/J Ars-A antibodies: somatic recombination generates the essential arginine at the junction of the variable and joining regions. *Proc Natl Acad Sci USA 84:* 1085-1089, 1987

Schiff, C., Corbet, S., Milili, M., and Fougereau, M.: Interstrain conservation of the murine GAT-specific antibody V_k repertoire as analyzed at the germline gene level. *EMBO J 2:* 1771-1776, 1983

Schiff, C., Corbet, S., and Fougereau, M.: The Ig germline gene repertoire: economy or wastage? *Immunol Today 9:* 10-14, 1988

Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C., and Leder, P.: Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. *Proc Natl Acad Sci USA 75:* 3881-3885, 1978

Seidman, J. G., Max, E. E., and Leder, P.: A kappa-immunoglobulin gene is formed by site-specific recombination without further somatic mutation. *Nature 280:* 370-375, 1979

Selsing, E. and Storb, U.: Somatic mutation of immunoglobulin light-chain variable-region genes. *Cell 25:* 47-58, 1981

Shlomchik, M., Nemazee, D., Van Snick, J., and Weigert, M.: Variable region sequences of murine IgM anti-IgG monoclonal autoantibodies (rheumatoid factors). II. Comparison of hybridomas derived by lipopolysaccharide stimulation and secondary protein immunization. *J Exp Med 165:* 970-987, 1987a

Shlomchik, M. J., Aucoin, A. H., Pisetsky, D. S., and Weigert, M. G.: Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. *Proc Natl Acad Sci USA 84:* 9150-9154, 1987b

Shlomchik, M. J., Marshak-Rothstein, A., Wolfowicz, C. B., Rothstein, T. L., and Weigert, M. G.: The role of clonal selection and somatic mutation in autoimmunity. *Nature 328:* 805-811, 1987c

Sikder, S. K., Akolkar, P. N., Kaladas, P. M., Morrison, S. L., and Kabat, E. A.: Sequences of variable regions of hybridoma antibodies to alpha (1 - >6) dextran in BALB/c and C57BL/6 mice. *J Immunol 135:* 4215-4221, 1985

Strohal, R., Kroemer, G., Wick, G., and Kofler, R.: Complete variable region sequence of a non-functionally rearranged kappa light chain transcribed in the non-secretor P3-X63-Ag8.653 myeloma cell line. *Nucleic Acids Res 15:* 2771, 1987

Tonegawa, S.: Somatic generation of antibody diversity. *Nature 302:* 575-581, 1983

Winter, E., Radbruch, A., and Krawinkel, U.: Members of novel V_H gene families are found in VDJ regions of polyclonally activated B-lymphocytes. *EMBO J 4:* 2861-2867, 1985

Wysocki, L. J., Gridley, T., Huang, S., Grandea, A. G., and Gefter, M. L.: Single germline V_H and V_k genes encode predominating antibody variable regions elicited in strain A mice by immunization with p-azophenylarsonate. *J Exp Med 166:* 1-11, 1987

Yancopoulos, G. D., Desiderio, S. V., Paskind, M., Kearney, J. F., Baltimore, D., and Alt, F. W.: Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. *Nature 311:* 727-733, 1984

Received July 31, 1989; revised version received September 26, 1989.