Published in final edited form as:

Immunol Today. 1994 August ; 15(8): 367–373. doi:10.1016/0167-5699(94)90175-9.

The CDR1 sequences of a major proportion of human germline Ig V_H genes are inherently susceptible to amino acid replacement

Bernard Chang and Paolo Casali

Dept of Pathology and Kaplan Comprehensive Cancer Center, New York University School of Medicine, 550 First Avenue, New York, NY 10016, USA

Abstract

The variable (V) genes of antigen-selected antibodies are known to exhibit a higher frequency of amino acid replacement mutations in the sequences encoding the antigen-contacting complementarity-determining regions (CDRs) than in those encoding the ‘structural’ framework regions (FRs). Here, Bernard Chang and Paolo Casali analyse the impact of regional differences in the codon composition of human germline Ig V_H and V_L genes on regional differences in the frequency of replacement mutations in the gene products (i.e. the antigen-binding sites of antibody molecules). This analysis reveals that CDR and FR sequences can differ significantly in their inherent susceptibility to amino acid replacement given any single nucleotide change. Thus, the CDR sequences of all the Ig V_H genes analysed comprise a higher frequency of codons susceptible to replacement mutations than would be expected for a random sequence. Conversely, the FR sequences comprise codons less susceptible to replacement mutations than expected. Random accumulation of nucleotide changes throughout the coding sequence of an Ig V-gene segment containing CDRs inherently more prone to replacement mutations than the respective FRs would inevitably yield a higher rate of amino acid replacements in the CDRs than in the FRs. This would provide a fertile structural substrate of hypervariability for antigen selection while still maintaining the structural integrity of the FRs.

The production of antibodies with progressively higher affinity for antigen is an important feature of B-cell clonotypes recruited by repeated antigenic challenge, and is referred to as affinity maturation¹. The molecular basis of this process was first unveiled by the analysis of the binding affinity and primary structure of monoclonal antibodies (mAbs) generated at successive stages of murine immune responses to different conjugated haptens and antigens^{2–10}. The first exposure to antigen results in recruitment of B-cell clonotypes that bind antigen by virtue of the combinatorial and junctional specificity of their unmutated surface receptors (Ig V segments). Subsequent exposures to antigen lead to accumulation of somatic point mutations in the antibody V segments and antigen selection of the high-affinity mutated B-cell clonotypes. Sequential rounds of mutation and selection eventually result in restriction of the response to those B cells with the best ‘fit’ for antigen^{11,12}.

Somatic point mutations in Ig V-gene segments

In the absence of negative or positive selective pressure on the gene product, a random mutational process would entail an even distribution of nucleotide changes yielding amino

acid replacements (R mutations) and nucleotide changes not yielding amino acid replacements (S, or silent, mutations) throughout the coding sequence. However, antigen-selected antibodies have been shown to include a higher frequency of R mutations in the Ig V-gene complementarity-determining regions (CDRs) than in the framework regions (FRs), where the proportion of S mutations may be greater. The high R:S mutation ratio characteristic of 'affinity-mature' Ig V-gene CDR sequences is consistent with their products being under positive pressure to mutate to provide the best fit for antigen. By contrast, the low R:S mutation ratio characteristic of the FR sequences is consistent with their need to be preserved in structure in order to provide the scaffolding for the antigen-contacting CDRs.

Ethical constraints have precluded the performance of time-course studies in humans analogous to those performed in mice. Thus, evidence of affinity maturation in humans has been restricted to observations that high-affinity antibodies induced by exogenous or self antigen bear mutations similar in nature and distribution to those of murine antibodies from the late stages of affinity maturation, i.e. R mutations concentrated in the CDRs. When analysing a hypermutated antibody as a single observation, the notion that a high concentration of R mutations in the CDRs is a mark of antigen selection is based on the assumption that the different regions of the parental germline V gene possess an equal inherent susceptibility to amino acid replacements as a result of random nucleotide changes. This inherent susceptibility to amino acid replacement can be expressed as replacement frequency (Rf , or the proportion of total possible point mutations that result in amino acid replacement), which has been calculated by Jukes and King¹³ as 0.7452 for a germline gene whose product is not subject to selective pressure. This Rf figure yields an inherent R:S mutation ratio (the quotient of total possible R mutations to total possible S mutations) of 2.925.

Based on some of the most-recent findings from our laboratory^{14–20}, we hypothesized that the CDR and FR sequences of a major proportion of human germline Ig V genes differ significantly in their inherent susceptibility to amino acid replacement given any single nucleotide change. The CDR sequences would comprise a higher frequency of codons susceptible to R mutations (displaying higher inherent Rf and R:S mutation ratio values) than would be expected for a random sequence. Conversely, the FR sequences would comprise codons less susceptible to R mutations (displaying lower inherent Rf and R:S mutation ratio values) than expected. Random accumulation of nucleotide changes throughout the coding sequence of an Ig V-gene segment comprising CDRs inherently more prone to R mutations than the respective FRs would result in higher R:S mutation ratios in the CDRs than in the FRs, even in the absence of antigen selection.

To test our hypothesis, we devised a computer algorithm by which the Rf and R:S mutation ratio values inherent to any given nucleotide sequence can be calculated. For each codon, the algorithm analyses the nine possible single-base changes as independent substitutions, and excludes those that result in a stop translation signal. This principle is exemplified by the analysis of CDR1 and FR codons of the human germline V_HH11 gene (Fig. 1).

Human germline Ig V_H-gene CDRs, particularly CDR1, are highly prone to R mutations

We applied the computer algorithm to the analysis of the inherent Rf and R:S mutation ratio values of the CDRs and FRs of 24 selected human germline V_H genes^{21–34}, including H11 (Table 1). These V_H genes are among those most-frequently expressed in the early and adult B-cell repertoires and encode natural antibodies and antibodies to foreign antigens, as well as autoantibodies in autoimmune patients. In these 24 V_H genes, the inherent Rf_{CDR} values were higher than the expected 0.7452 Rf value for a random sequence ($\alpha = 0.05$; $t = 17.9$, $p = 2.5 \times 10^{-15}$, one-tailed Student's t -test), and higher than the respective Rf_{FR} values ($t = 12.9$, $p = 2.7 \times 10^{-12}$). Conversely, the Rf_{FR} values were lower than the expected value ($t = 2.42$, $p = 1.2 \times 10^{-2}$). The inherent R:S_{CDR} mutation ratios were higher than the 2.925 R:S mutation ratio expected for a random sequence ($t = 14.0$, $p = 4.81 \times 10^{-13}$), and higher than the respective R:S_{FR} mutation ratios ($t = 11.6$, $p = 2.0 \times 10^{-11}$). The R:S_{FR} mutation ratios were significantly lower than the expected value ($t = 2.20$, $p = 1.9 \times 10^{-2}$).

To determine whether the inherently high CDR and low FR susceptibility to R mutations could be attributed to discrete clusters of codons with extreme Rf values, we generated inherent mutational profiles for the 24 different V_H genes by applying the statistical algorithm to the analysis of FR1, CDR1, FR2, CDR2 and FR3 as discrete sequences. The inherent R:S mutation ratios for CDR2 sequences (mean=3.731) were generally higher than both the expected random value (2.925) and the corresponding values for the adjacent FR sequences (R:S_{FR2} mean=3.086, R:S_{FR3} mean=3.060), although the difference was not consistent at the single gene level (not shown). By contrast, the CDR1 sequences displayed strikingly high inherent R:S mutation ratios (mean=7.146 *versus* R:S_{FR1} mean=2.560 and R:S_{FR2} mean=3.086), and in some genes, such as H11 and HHG19, the R:S mutation ratios were more than fivefold those of the adjacent FR sequences (Table 1, Fig. 2). Thus, CDR but not FR sequences of a major proportion of human germline Ig V_H genes display a significantly high structural susceptibility to R mutations. Such an inherent susceptibility to amino acid substitution is contributed mainly by the codon composition of CDR1.

Human germline Ig V_L-gene CDR sequences are not prone to R mutations

Calculation of the inherent Rf and R:S mutation ratio values for the CDRs and FRs of 12 human germline V_L genes^{35–46} (Table 2), selected by criteria similar to those used for the V_H-gene sampling, revealed Rf_{CDR} values that were higher than the expected 0.7452 Rf value ($t = 11.50$, $p = 9.01 \times 10^{-8}$) and the respective Rf_{FR} values ($t = 10.2$, $p = 3.0 \times 10^{-7}$), but lower values overall than those calculated for the V_H-segment CDRs ($t = 5.45$, $p = 3.0 \times 10^{-6}$). The V_L-gene Rf_{FR} values did not significantly differ either from the expected value or from the values calculated for the 24 V_H genes. The inherent R:S_{CDR} mutation ratios were higher than the theoretically expected 2.925 value ($t = 9.95$, $p = 3.9 \times 10^{-7}$) and the respective R:S_{FR} mutation ratios ($t = 10.2$, $p = 3.2 \times 10^{-7}$), but were lower than the R:S mutation ratios of the V_H-segment CDRs ($t = 5.48$, $p = 2.2 \times 10^{-6}$). The V_L-gene R:S_{FR} mutation ratios did not differ significantly from the theoretically expected value, nor did they differ from those of the 24 V_H genes. The inherent mutational profiles for the V_L-gene segments (as exemplified by the analysis of two V_κ and one V_λ genes in Fig. 2) showed that

inherent R:S mutation ratios were, on average, only moderately higher in the CDR1 sequences (mean = 3.757) and CDR3 sequences (mean = 4.516) than those in the adjoining FR sequences (R:S_{FR1} mean = 2.514, R:S_{FR2} mean = 3.340, R:S_{FR3} mean = 3.018) and the theoretically expected 2.925 R:S mutation ratio. Interestingly, the average inherent R:S_{CDR2} mutation ratio (2.945) was actually lower than those of the adjoining FR sequences (see above). Thus, as compared with the sequences of V_H genes, those of V_L genes exhibit a smaller overall differential between inherent R:S_{CDR} and R:S_{FR} mutation ratios and virtually no codon clusters with high inherent R:S mutation ratios.

Inherent Ig V_H-gene susceptibility to R mutations and assessment of antigen selection

The present findings indicate that when assessing the antigen-selected nature of the somatic point mutations in an expressed Ig V gene, the R:S mutation ratio values inherent to the progenitor germline gene sequence, not the theoretical expected value of 2.925, must be used as the term of comparison. We have applied this principle to the analysis of seven human mAbs that we formally proved to have undergone somatic hypermutation, and for which, in each case, the respective autologous V_H germline gene was identified by cloning and sequencing genomic DNA from polymorphonuclear leukocytes (PMN) of the same subject whose B cells were used for the generation of the mAb-producing clone in question (Table 3). These mAbs differ widely in a variety of features, displaying foreign- and self-antigen specificity, mono- and polyreactivity, and comprising α, γ and μ H-chain isotypes. For each expressed V_H gene, we calculated the number of expected R mutations and the expected R:S mutation ratio, using the formula described in Box 1 and the Rf_{CDR} and Rf_{FR} inherent to the respective progenitor germline gene (Table 1). Some of the germline genes, such as WHG16G, V_H4.18, H11 and V_H4.11, displayed a high inherent Rf_{CDR} ; while others, such as hv1263, V_H26c and 51P1, displayed a relatively low inherent Rf_{CDR} . Consistent with selection by antigen, all mAb V_H segments displayed higher and lower R:S mutation ratios in the CDRs and FRs, respectively, than those theoretically expected given the inherent germline Rf_{CDR} and Rf_{FR} values. The possibility that the excess of R mutations in the CDRs (with only the exception of mAb 426.12.3F1.4) and scarcity of R mutations in the FRs resulted solely from chance was negated by the significantly low probability values (Table 3) calculated according to the binomial distribution model⁴⁷ (Box 1). Thus, antigen selection of expressed V_H genes results in CDR R:S mutation ratios significantly higher than those expected from random mutation, even when the progenitor germline Ig V_H-gene CDR sequences display a high inherent susceptibility to amino acid replacement.

Box 1

a. Calculation of number of expected R mutations in Ig V-gene CDRs or FRs:

$$\text{Number of expected R mutations} = n \times (CDR\ Rf \text{ or } FR\ Rf) \times (CDR_{rel} \text{ or } FR_{rel})$$

where:

n = total number of observed mutations

Rf = replacement frequency inherent to CDR or FR sequences

CDR_{rel} or FR_{rel} = relative size of the CDRs or FRs

- b. Calculation (according to the binomial distribution model) of the probability that excess or scarcity of R mutations (in Ig V-gene CDRs or FRs) resulted from chance only:**

$$p = \{n! / [k!(n-k)!]\} \times q^k \times (1-q)^{n-k}$$

where:

n = total number of observed mutations

k = number of observed R mutations in the CDRs or FRs

q = probability that an R mutation will localize to CDRs or FRs ($q = CDR_{rel} \times CDR Rf$ or $FR_{rel} \times FR Rf$)

Structural susceptibility to R mutations as a source of Ig V-gene hypervariability

By analysing intrinsic nucleotide substitution preferences derived from studies of passenger murine Ig V transgenes and flanking regions, Betz *et al.*^{48,49} identified 'hotspots' resulting from the preferential targeting by the mutational machinery. In the few Ig V_H genes analysed, these mutational hotspots are particularly frequent within the CDR1 and flanking areas. The inherently high susceptibility to amino acid replacement of certain V_H-gene CDR1 sequences would constitute a source of amino acid hypervariability that, although independent, can be synergistic with that resulting from the mutational hotspot mechanism. This concept is exemplified by our analysis of the murine V_H186.2 gene, also discussed by Betz *et al.*⁴⁹. V_H186.2 is used in the antibody response to the hapten 4-hydroxy-3-nitrophenacetyl (NP) in C57BL/6 mice, and has a CDR1 sequence that is identical to the corresponding sequence of the human V_HH11 gene (Fig. 1). The CDR1 codons of the V_H186.2 gene are preferentially targeted by the mutational machinery, particularly Ser31, Trp33 and Met34. In addition, the composition of these three and the remaining two CDR1 codons confers an extremely high inherent R:S mutation ratio (12.667) on the V_H186.2 CDR1 gene sequence. Thus, in certain Ig V-gene sequences, such as that of the CDR1 of the mouse/human 186.2/H11 V_H segments, the inherently high susceptibility to R mutations would greatly magnify the hypervariability stemming from the preferential targeting by the mutational machinery. The suggestion by Rogozin and Kolchanov⁵⁰ and Betz *et al.*⁴⁹ that the hotspots preferentially targeted by the mutational machinery may show local sequence features, e.g. the DNA sequence RGYW (where R=purine, Y=pyrimidine and W=A or T), may provide a single integrated structural basis for both mechanisms of hypervariability. Indeed, RGYW is present in the CDR1 of the mouse/human 186.2/H11 V_H genes, as well as in several of the CDR1 sequences of other human germline V_H genes that display a high inherent structural susceptibility to R mutations.

Conclusions

The high inherent susceptibility to amino acid replacement of germline gene sequences encoding the Ig V_H-segment CDRs, but not FRs, may constitute an important evolutionary feature reflecting the different functions of these discrete portions of the Ig V-gene product. In addition, the significantly higher inherent R:S mutation ratios of the CDRs of V_H-gene segments as compared with those of the corresponding regions of V_L-gene segments might reflect a function-related evolutionary selection, consistent with the accumulating evidence that H-chain V segments provide the major structural correlate for antigen binding^{51–54}. The fact that the inherent structural hypervariability is maximal within a gene sequence, CDR1, that may also be preferentially targeted by the mutational machinery suggests that the concentration of R mutations observed within CDR1 and flanking areas in high-affinity ‘mature’ antibodies is not solely due to antigen selection. Interestingly, CDR2, not CDR1, is the region with the highest degree of heterogeneity when considering the hypervariability plots of the whole panel of reported human germline V_H genes^{55,56}. This suggests that the phylogenetic selective forces that operate mainly on CDR1, and that are ultimately responsible for the inherent susceptibility of Ig V genes to R mutations, may differ from those operating mainly on CDR2, and that are possibly responsible for the main structural correlate of germline Ig-gene heterogeneity. The precise contribution of the hypervariability mechanism discussed here to Ig V-segment hypermutation and B-cell clonotype selection by antigen will be better defined by further structural studies of the antibody molecule and by increased knowledge of the nucleotide substitution preferences of the enzymatic replicative machinery.

Acknowledgments

This work was supported by U.S.P.H.S. grants AR-40809 and CA-16804. P.C. is a Kaplan Cancer Scholar. B.C. was supported by the U.S.P.H.S. Short Term Training Grant 5T35 DK-07421. This is publication 25 from *The Jeannette Greenspan Laboratory for Cancer Research*. We are most grateful to Julia Phillips-Quagliata, G. Jeanette Thorbecke and George Teebor for critical appraisal of this manuscript.

References

1. Steiner LA, Eisen HN. J Exp Med. 1967; 126:1161–1183. [PubMed: 4168368]
2. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Proc Natl Acad Sci USA. 1984; 81:3180–3184. [PubMed: 6203114]
3. Berek C, Griffiths GM, Milstein C. Nature. 1985; 316:412–418. [PubMed: 3927173]
4. Sablitzky F, Wildner G, Rajewsky K. EMBO J. 1985; 4:345–350. [PubMed: 3926481]
5. Wysocki L, Manser T, Gefter ML. Proc Natl Acad Sci USA. 1986; 83:1847–1851. [PubMed: 3485290]
6. Malipiero UV, Levy NS, Gearhart PJ. Immunol Rev. 1987; 96:59–74. [PubMed: 3298010]
7. Claflin JL, Berry J, Flaherty D, Dunnick W. J Immunol. 1987; 138:3060–3068. [PubMed: 3106498]
8. Berek C, Milstein C. Immunol Rev. 1987; 96:23–41. [PubMed: 3298007]
9. Berek C, Milstein C. Immunol Rev. 1988; 105:5–26. [PubMed: 3058579]
10. French DL, Laskov R, Scharff MD. Science. 1989; 244:1152–1157. [PubMed: 2658060]
11. Manser T. J Exp Med. 1989; 170:1211–1230. [PubMed: 2507726]
12. Weiss U, Rajewsky K. J Exp Med. 1990; 172:1681–1689. [PubMed: 2124253]
13. Jukes TH, King JL. Nature. 1979; 281:605–606. [PubMed: 492328]

14. Harindranath N, Goldfarb IS, Ikematsu H, et al. *Int Immunol*. 1991; 3:865–875. [PubMed: 1718404]
15. Mantovani L, Wilder RL, Casali P. *J Immunol*. 1993; 151:473–488. [PubMed: 7686945]
16. Ikematsu H, Kasaian MT, Schettino EW, Casali P. *J Immunol*. 1993; 151:3604–3616. [PubMed: 8376796]
17. Ikematsu H, Harindranath N, Ueki Y, Notkins AL, Casali P. *J Immunol*. 1993; 150:1325–1337. [PubMed: 8432980]
18. Kasaian MT, Ikematsu H, Balow JE, Casali P. *J Immunol*. 1994; 152:3137–3151. [PubMed: 8144908]
19. Riboldi P, Gaidano G, Schettino EW, et al. *Blood*. 1994; 83:2952–2961. [PubMed: 8180391]
20. Ikematsu H, Ichiyoshi Y, Schettino EW, Nakamura M, Casali P. *J Immunol*. 1994; 152:1430–1441. [PubMed: 8301143]
21. Schroeder HW Jr, Hillson JL, Perlmutter RM. *Science*. 1987; 238:791–793. [PubMed: 3118465]
22. Chen PP, Liu MF, Glass CA, Sinha S, Kipps TJ, Carson DA. *Arthritis Rheum*. 1989; 32:72–76. [PubMed: 2492198]
23. Shin EK, Matsuda F, Nagaoka H, et al. *EMBO J*. 1991; 10:3641–3645. [PubMed: 1935893]
24. Campbell MJ, Zelenetz AD, Levy S, Levy R. *Mol Immunol*. 1992; 29:193–203. [PubMed: 1542297]
25. Chen PP, Liu MF, Sinha S, Carson DA. *Arthritis Rheum*. 1988; 31:1429–1431. [PubMed: 3263866]
26. Berman JE, Mellis SJ, Pollock R, et al. *EMBO J*. 1988; 7:727–738. [PubMed: 3396540]
27. Kuppers R, Fischer U, Rajewsky K, Gause A. *Immunol Lett*. 1992; 34:57–62. [PubMed: 1282498]
28. Chen PP. *Scand J Immunol*. 1990; 31:257–267. [PubMed: 2320951]
29. Rechavi G, Bienz B, Ram D, et al. *Proc Natl Acad Sci USA*. 1982; 79:4405–4409. [PubMed: 6812048]
30. Lee KH, Matsuda F, Kinashi T, Kodaira M, Honjo T. *J Mol Biol*. 1987; 195:761–768. [PubMed: 3116265]
31. Sanz I, Kelly P, Williams C, Scholl S, Tucker PW, Capra JD. *EMBO J*. 1989; 8:3741–3748. [PubMed: 2511001]
32. Weng NP, Snyder JG, Yu-Lee LY, Marcus DM. *Eur J Immunol*. 1992; 22:1075–1082. [PubMed: 1348029]
33. Humphries CG, Shen A, Kuziel WA, Capra JD, Blattner FR, Tucker PW. *Nature*. 1988; 331:446–449. [PubMed: 3123998]
34. Logtenberg T, Young FM, van Es JH, Gmelig-Meyling FHJ, Alt FW. *J Exp Med*. 1989; 170:1347–1355. [PubMed: 2507728]
35. Ermel RW, Kenny TP, Chen PP, Robbins DL. *Arthritis Rheum*. 1993; 36:380–388. [PubMed: 7916589]
36. Chen PP, Albrandt K, Kipps TJ, Radoux V, Liu F, Carson DA. *J Immunol*. 1987; 139:1727–1733. [PubMed: 3114376]
37. Silberstein LE, Jefferies LC, Goldman J, et al. *Blood*. 1991; 78:2372–2386. [PubMed: 1657249]
38. Liu MF, Robbins DL, Crowley JJ, et al. *J Immunol*. 1989; 142:688–694. [PubMed: 2492051]
39. Kipps TJ, Tomhave E, Chen PP, Carson DA. *J Exp Med*. 1988; 167:840–852. [PubMed: 3127527]
40. Klobeck HG, Bornkamm GW, Combriato G, Mocikat R, Pohlenz HD, Zachau HG. *Nucleic Acids Res*. 1985; 13:6515–6529. [PubMed: 2997712]
41. Carroll WL, Yu M, Link MP, Korsmeyer JS. *J Immunol*. 1989; 143:692–698. [PubMed: 2500485]
42. Tsunetsugu-Yokota Y, Minekawa T, Shigemoto K, Shirasawa T, Takemori T. *Mol Immunol*. 1992; 29:723–728. [PubMed: 1603093]
43. Olee T, Lu EW, Huang DF, et al. *J Exp Med*. 1992; 175:831–842. [PubMed: 1740665]
44. Siminovitch KA, Misener V, Kwong PC, Song QL, Chen PP. *J Clin Invest*. 1989; 84:1675–1678.
45. Combriato G, Klobeck HG. *Eur J Immunol*. 1991; 21:1513–1522. [PubMed: 1904362]
46. Yamasaki N, Komori S, Watanabe T. *Mol Immunol*. 1987; 24:981–985. [PubMed: 2889139]

47. Shlomchik MJ, Aucoin AH, Pisetsky DS, Weigert MG. Proc Natl Acad Sci USA. 1987; 84:9150–9154. [PubMed: 3480535]
48. Betz AG, Cristina R, Pannell R, Milstein C, Neuberger MS. Proc Natl Acad Sci USA. 1993; 90:2385–2388. [PubMed: 8460148]
49. Betz AG, Neuberger MS, Milstein C. Immunol Today. 1993; 14:405–411. [PubMed: 8397780]
50. Rogozin IB, Kolchanov NA. Biochim Biophys Acta. 1992; 1171:11–18. [PubMed: 1420357]
51. Amit AG, Mariuzza RA, Phillips SEV, Poljak RJ. Science. 1986; 233:747–753. [PubMed: 2426778]
52. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G. Nature. 1989; 341:544–546. [PubMed: 2677748]
53. Radic MZ, Mascelli MA, Erikson J, Shan H, Weigert M. J Immunol. 1991; 146:176–182. [PubMed: 1898596]
54. Kabat EA, Wu TT. J Immunol. 1991; 147:1709–1719. [PubMed: 1908882]
55. Kabat, EA.; Wu, TT.; Perry, HM.; Gottesman, KS.; Foeller, C. Sequences of Proteins of Immunological Interest. 5. US Department of Health and Human Services; 1991.
56. Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. J Mol Biol. 1992; 227:776–798. [PubMed: 1404388]

CDR1					FR2			FR3			
31	32	33	34	35	44	45	46	78	79	80	81

AGC	TAC	TGG	ATG	CAC	GGG	CTG	GTG	ACG	CTG	TAT	CTG
<u>Ser</u>	<u>Tyr</u>	<u>Trp</u>	<u>Met</u>	<u>His</u>	<u>Gly</u>	<u>Leu</u>	<u>Val</u>	<u>Thr</u>	<u>Leu</u>	<u>Tyr</u>	<u>Leu</u>
Cys	His	Arg	Leu	Tyr	Trp	Met	Leu	Ser	Met	His	Met
Arg	Asn	Arg	Leu	Asn	Arg	Val	Leu	Pro	Val	Asn	Val
Gly	Asp	Gly	Val	Asp	Arg	Pro	Met	Ala	Pro	Asp	Pro
Ile	Phe	Leu	Thr	Leu	Val	Gln	Ala	Met	Gln	Phe	Gln
Thr	Ser	Ser	Lys	Pro	Ala	Arg	Glu	Lys	Arg	Ser	Arg
Asn	Cys	Cys	Arg	Arg	Glu		Gly	Arg		Cys	
Arg		Cys	Ile	Gln							
Arg			Ile	Gln							
			Ile								

Fig. 1.

The human germline V_HH11 gene sequence was analysed using the computer algorithm that calculates the inherent susceptibility to amino acid replacement given any single nucleotide change. The complementarity-determining regions (CDRs) and framework regions (FRs) of the gene sequence displayed extremely different inherent R:S mutation ratios. The sequence of the whole CDR1 and clusters of codons from the FR2 and FR3 are depicted. Each circle above each nucleotide represents a single viable substitution that is possible at that base (substitutions that result in a stop codon are excluded). Closed and open circles represent single-base substitutions that result in an amino acid replacement (R mutation) and no replacement (S mutation), respectively. Listed below each codon are the germline-encoded amino acid (underlined), and all possible amino acid replacements resulting from single-base substitutions at that codon. For example, the first codon in the CDR1 of H11 is AGC, which encodes Ser and has a total of nine possible single-base substitutions for its three nucleotides. Eight of these substitutions result in an amino acid replacement, and one does not. The sequence of the CDR1 of the human V_HH11 gene is identical to that of the mouse V_H186.2 gene CDR1.

Fig. 2.

Inherent R mutation susceptibility profiles of human germline V_H - and V_L -gene sequences. The complementarity-determining region (CDR) and framework region (FR) sequences of the most-commonly expressed human Ig V_H and V_L genes, as listed in Tables 1 and 2, were analysed using the computer algorithm that calculates the inherent susceptibility to amino acid replacement given any single nucleotide change. Inherent R mutation susceptibility profiles were created by plotting the inherent R:S mutation ratio (ratio of total possible single-base substitutions resulting in amino acid replacement to total possible single-base substitutions not resulting in amino acid replacement) for each of the Ig V regions versus the respective V region nucleotide residue numbers. (a) The inherent R mutation susceptibility profiles of the six human germline V_H -gene sequences displaying the highest Rf in CDR1 are depicted in the upper six panels (where Rf indicates replacement frequency, or proportion of total possible point mutations that result in amino acid replacement), (b) The inherent R mutation susceptibility profiles of the three human V_L -gene sequences (two V_K and one V_λ displaying the highest Rf in CDR1 are depicted in the lower three panels. Numbers below each region represent the inherent R:S mutation ratio for that region.

Table 1

Inherent R_f and R:S mutation ratios for human germline V_H genes

Gene ^a	Identity (%)	Ref.	Mutations in CDRs			CDR1 alone			Mutations in FRs		
			Total possible changes ^b	R_f	R:S mutation ratio ^c	R:S mutation ratio	Total possible changes	R_f	R:S mutation ratio	Total possible changes	R_f
$V_H I$											
51P1	100.0	21	191	0.7801	3.548	4.375	650	0.7477		650	0.7477
hvl263	98.3	22	191	0.7801	3.548	4.375	650	0.7477		650	0.7477
V_{H-2}	88.1	23	187	0.8128	4.343	5.833	650	0.7492		650	0.7492
V_{H-3B}	87.4	23	186	0.8118	4.314	6.167	649	0.7504		649	0.7504
$V_H II$											
V_{H2-MC2}	100.0	24	197	0.8020	4.051	4.083	635	0.7433		635	0.7433
$V_H III$											
V_{H26c}	100.0	25	191	0.7801	3.548	6.167	652	0.7423		652	0.7423
8-1B	92.5	26	180	0.7944	3.865	9.750	652	0.7408		652	0.7408
WHG16G	91.2	27	194	0.8144	4.389	9.750	644	0.7376		644	0.7376
22-2B	89.8	26	187	0.8128	4.343	9.250	650	0.7400		650	0.7400
1.9III	89.1	26	185	0.7946	3.868	6.167	655	0.7404		655	0.7404
GL-SJ2	89.1	28	185	0.7946	3.868	6.167	655	0.7404		655	0.7404
hV3005	89.1	28	185	0.7946	3.868	6.167	655	0.7404		655	0.7404
H11	88.4	29	191	0.8063	4.162	12.667	652	0.7362		652	0.7362
HHG19	87.1	27	184	0.8315	4.935	12.667	652	0.7393		652	0.7393
$V_H IV$											
V_{H1-2}	100.0	30	192	0.8177	4.486	6.125	654	0.7217		654	0.7217
V_{H1-4}	98.0	30	174	0.8218	4.613	8.750	654	0.7217		654	0.7217
V_{H4-11}	97.6	31	174	0.8218	4.613	8.750	654	0.7232		654	0.7232
4.33	95.3	32	190	0.8053	4.135	4.700	654	0.7278		654	0.7278
V_{H4-18}	94.9	31	192	0.8177	4.486	6.167	657	0.7245		657	0.7245
V79	92.6	30	188	0.8351	5.064	11.500	658	0.7264		658	0.7264
V_{H4-21}	90.9	31	187	0.8182	4.500	5.857	642	0.7321		642	0.7321
$V_H V$											

Gene ^a	Mutations in CDRs				CDR1 alone		Mutations in FRs		
	Identity (%)	Ref.	Total possible changes ^b	R _f	R:S mutation ratio ^c	R:S mutation ratio	Total possible changes	R _f	R:S mutation ratio
V _H 251	100.0	33	188	0.7766	3.476	4.125	645	0.7659	3.272
V _H 32	94.9	33	189	0.8095	4.150	7.200	647	0.7651	3.256
V _H V1									
6-1G1	100.0	34	184	0.7935	3.842	4.778	678	0.7404	2.852
Mean value			187.2	0.8053	4.167	7.148	651.8	0.7394	2.845
Standard deviation			5.4	0.0161	0.426	2.627	7.4	0.0116	0.175

Abbreviations: *R_f*, replacement frequency or the proportion of total possible point mutations that result in amino acid replacement; R:S, ratio of mutations that yield amino acid replacements to mutations that are silent; CDR, complementarity-determining region; FR, framework region.

^a Within each V_H family, genes are listed according to their sequence identity (%) with an arbitrary prototypic sequence for the family.

^b Total possible mutations are calculated as the total number of possible single-base substitutions excluding those resulting in stop codons.

^c Inherent R:S mutation ratio is the quotient of total possible R to total possible S mutations. For a random sequence, *R_f* should be 0.7452 and R:S should be 2.925 (Ref. 13).

Table 2

Inherent R_f and R:S mutation ratios for human germline V_L genes

Gene ^a	Ref.	Mutations in CDRs			CDRI alone		Mutations in FRs		
		Total possible changes ^b	R _f	R:S mutation ratio ^c	R:S mutation ratio	Total possible changes	R _f	R:S mutation ratio	
V _κ I									
kalc6	35	195	0.8000	4.000	3.318	612	0.7549	3.080	
V _κ III									
kv3g	36	216	0.7917	3.800	3.947	606	0.7492	2.987	
20G2	37	216	0.7824	3.596	4.053	607	0.7512	3.020	
HumKv328h5	38	213	0.7606	3.176	4.053	607	0.7512	3.020	
HumKv325	39	223	0.7713	3.373	4.150	605	0.7504	3.007	
V _κ IV									
V _κ IV	40	268	0.7836	3.621	4.259	607	0.7529	3.047	
V _κ I									
IB9/F2	41	250	0.7760	3.464	3.625	597	0.7236	2.618	
YM-1	42	272	0.7831	3.610	3.567	613	0.7341	2.761	
Humlv1L1	43	254	0.7756	3.456	3.423	593	0.7285	2.683	
Humlv117	44	250	0.7720	3.386	3.385	590	0.7271	2.665	
V _κ III									
V _κ III.1	45	211	0.7915	3.795	3.600	593	0.7420	2.876	
V _κ V									
p3C4/5	46	238	0.7787	3.519	3.700	582	0.7406	2.854	
Mean value		233.8	0.7805	3.566	3.757	601.0	0.7421	2.885	
Standard deviation		23.8	0.0102	0.214	0.322	9.3	0.0107	0.159	

^a See Table 1 for abbreviations and footnotes.

Table 3

R and S mutations in human monoclonal antibody (mAb) and autoantibody V_H genes

Antibody ^a	Origin (Ref.)	Ig class	Reactivity	K _d (M)	Germline gene	R:S ^d	Observed ^e	Theoretically expected	p
mAb57	HS (17)	IgG1	Monoreactive	1.1×10 ^{-10c}	hV1263	CDR:	8,000 (8:1)	3,548	2.147 × 10 ⁻³
			anti-rabies virus glycoprotein ^b			FR:	2,500 (5:2)	2,963	1.943 × 10 ⁻²
mAb412.67.F1.3	SLE (18)	IgA1	Monoreactive	2.7×10 ⁻⁹	WHG16G	CDR:	∞ (6:0)	4,389	3.450 × 10 ⁻³
			anti-dsDNA ^b			FR:	1,000 (2:2)	2,811	1.762 × 10 ⁻²
mAb412.66.F1	SLE (18)	IgA1	Polyreactive	6.0×10 ⁻⁸	V _H 26c	CDR:	∞ (5:0)	3,548	8.800 × 10 ⁻³
			anti-ssDNA ^b			FR:	1,000 (2:2)	2,881	2.400 × 10 ⁻²
mAb61	RA (15)	IgM	Monoreactive	6.0×10 ⁻⁷	V _H 4.18	CDR:	5,000 (5:1)	4,486	2.515 × 10 ⁻⁴
			anti-IgG Fc fragment ^b (rheumatoid factor)			FR:	0,000 (0:1)	2,630	0.000
mAb112	RA (17)	IgM	Monoreactive	4.6 × 10 ⁻⁷	51P1	CDR:	6,000 (6:1)	3,548	3.964 × 10 ⁻³
			anti-IgG Fc fragment ^b (rheumatoid factor)			FR:	1,250 (5:4)	2,963	4.248 × 10 ⁻³
mAb13	IDDM (20)	IgG1	Monoreactive	5.0 × 10 ⁻⁷	H11	CDR:	∞ (4:0)	4,162	4.000 × 10 ⁻³
			anti-human insulin ^b			FR:	0,000 (0:1)	2,791	0.000
mAb426.12.3F1.4	IDDM (16)	IgG1	Polyreactive	1.7 × 10 ⁻⁸					
			anti-ssDNA,						
			anti-human insulin,	4.0 × 10 ⁻⁵	V _H 4.11	CDR:	∞ (3:0)	4,613	1.182 × 10 ⁻¹
			anti-tetanus toxoid	1.0 × 10 ⁻⁶		FR:	0,286 (2:7)	2,613	5.976 × 10 ⁻³

Abbreviations: HS, healthy subject; SLE, systemic lupus erythematosus; RA, rheumatoid arthritis; IDDM, insulin-dependent diabetes mellitus; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; PMN, polymorphonuclear leukocyte; ∞, infinite. For other abbreviations see Table 1.

^a All mAbs listed were formally proved to have undergone somatic hypermutation. For each mAb, the V_H germline gene that putatively gave rise to the expressed mutated segment was identified by analysing PMN DNA from the patient whose B lymphocytes were used to generate the mAb.

^b Possible antigen operating during *in vivo* selection. For polyreactive mAb426.12.3F1.4, the selecting antigen is undetermined.

^c K_d in g μl⁻¹.

^d Observed R:S mutation ratio value is the quotient of observed R to observed S mutations in the mAb CDRs or FRs. The theoretical inherent R:S mutation ratio value is the quotient of total possible R to total possible S mutations in the germline gene. For a random sequence, the R:S mutation ratio should be 2.925 (Ref. 13). Probability values (p) were calculated using the formula in Box 1.

^e Numbers in parentheses are actual numbers of R to S mutations.