Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9138–9142, December 1990
Immunology

Resolution of hypervariable regions in T-cell receptor β chains by a modified Wu–Kabat index of amino acid diversity

(immunoglobulin fold/evolution/protein sequence features)

RITA JORES*, PEDRO M. ALZARI†, AND TOMMASO MEO*

*Unité d'Immunogénétique, Institut National de la Santé et de la Recherche Médicale U 276, Institut Pasteur, 75724 Paris, France; and †Unité d'Immunologie Structurale, URA 359, Centre National de la Recherche Scientifique, Institut Pasteur, 75724 Paris, France

Communicated by Elvin A. Kabat, August 29, 1990

ABSTRACT **The Wu–Kabat variability coefficient is a well-established descriptor of the susceptibility of an amino acid position to evolutionary replacements. It conveniently highlights stretches of accentuated amino acid variation that, for example, in an antibody molecule account for most of the antigen contacts (complementarity-determining regions). Diverse opinions are held as to why the index yields unclear results when applied instead to the polypeptide sequences of the T-cell antigen receptor. We show that a simple modification enhances the resolving power of the index by increasing the weight on the frequency distribution of the amino acids in the formula. Application of the improved index to T-cell receptor β chains highlights four unambiguous hypervariable regions, three of which are positioned similar to immunoglobulin complementarity-determining regions along the chain. In a Fab-like three-dimensional model of the T-cell receptor, the four hypervariable regions coincide with the four loops on the surface of the domain and form a contiguous area available for binding.**

Antibodies remain the leading example of proteins whose pattern of amino acid sequence variation has been a significant clue both to their three-dimensional structure and genetic design. Thus, in a statistical examination of light- and heavy-chain polypeptides from human and mouse immunoglobulins, several invariant residues were recognized and interpreted as sites of strict structural constraints (1), and three linear stretches of hypervariable amino acid positions were delineated and predicted to cluster in space to form the antibody combining site (2–4), hence their designation as complementarity-determining regions (CDRs) (2). High-resolution x-ray crystallography results on Fab structures and antigen–Fab complexes (for review, see ref. 5) have verified most of the structural prediction and validated the approach of relating binding specificity to amino acid side-chain differences in CDRs.

The T-cell antigen receptor (TCR) is a heterodimer of disulfide-linked chains (αβ or γδ) homologous to those of the antibody molecules, with which they share a basic design consisting of an amino-terminal variable (V) region and a carboxyl-terminal constant (C) region assembled from discrete gene segments during differentiation. Immunoglobulin V and C regions form distinct domains characterized by a typical fold consisting of two β-pleated sheets connected at their extremities by loops that may vary in length and character. Although V domains ought to encode epitope specificity, in no case have segments of hypervariability similar to the CDRs of immunoglobulin clearly been recognized in TCR V chains.

Attempts to identify the CDRs of the TCR polypeptides began with the very first reports on the cloning of the receptor in man and mouse. Thus Patten *et al.* (6) called attention to the extraordinarily high level of dissimilarity observed within the β-chain V regions relative to immunoglobulin V regions. Computation of the Wu–Kabat coefficient of variability for amino acid positions brought out an unconventionally broad profile of variability that could be partitioned in up to seven hypervariable regions. Patten *et al.* (6) also ascertained that the results could not be due to an artifact of the small number of sequences available, by showing that equally small sets of V regions of immunoglobulin κ (Vκ) or heavy (VH) chains did display the canonical distribution of hypervariability in spite of having been selected in a way that maximized heterogeneity. By contrast, Barth *et al.* (7), after examining an additional set of mouse Vβ sequences, concluded that both the total diversity and the distribution of variability in the Vβ and VH sequences are not significantly different from one another. They took the view that among human VH sequences obtained by direct protein analysis, those with blocked α-amino groups offer a more random representation of VH segments and that small sample size may inflate variability. According to their analysis, the atypical hypervariable regions of Vβ sequences could in fact be seen also in VH chains, if chosen among α-amino-blocked human sequences. However, subsequent applications of the Wu–Kabat variability analysis to a larger number of mouse β chains (8) or to their human homologs (9–11) lent support to the initial hypothesis that relative to immunoglobulins, Vβ sequences possess a higher and more widespread level of amino acid variation, which is difficult to apportion into discrete hypervariable regions. The Wu–Kabat analysis of the primary structure of mouse α-chain V regions, rather than resolving the issue, left the contending views equipoised (12–14).

Compelled by the functional implications of the possible structural differences implied by the Wu–Kabat plots between immunoglobulin and TCR V regions, different methods of sequence analysis were emphasized, which either predict secondary structure potentials (7, 11, 12) or allow construction of three-dimensional models by analogy to existing crystal structures (15–18). These methods concur to support the prediction that the TCR and immunoglobulin V domains are structurally similar. Nevertheless, Schiffer *et al.* (19) sought to improve the resolution of the Wu–Kabat plots and pointed out that, besides sample size effects, illegitimate pooling of sequences leads to artifactual results. They showed that a reassortment of Vβ sequences into two subgroups helps to sharpen the Wu–Kabat variability distribution of the ensemble. However, their subsets based on a tertiary structure criterion and the finer subgrouping attempted by Bougueleret and Claverie (20) on the basis of optimal sequence alignments fell short of defining hypervari-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CDR, complementarity-determining region; TCR, T-cell antigen receptor; V, variable; Vκ, VH, etc., κ chain V region, heavy chain V region, etc., respectively.

able regions in Vβ chains. Of course, all studies pointed to the dominant peak of variability corresponding to immunoglobulin-like CDR3 generated somatically by diversifying mechanisms linked to the joining of the V and C region gene segments (21, 22).

The data reported in this paper demonstrate that a modification in the computation of the Wu–Kabat coefficient of variability allows for a greatly improved sensitivity of the analysis. Application of our index of amino acid variability to Vβ regions effectively delineates four unambigous stretches of hypervariability.

MATERIALS AND METHODS

We established two data bases of amino acid sequences. In the first data base, we pooled most published sequences (ref. 23 and references published therein as well as refs. 24–43) and an additional set of 20 β-chain sequences derived by one of us (R.J., unpublished data) from cDNA clones of a single human thymus. The total set consists of 159 β chains (93 human, 46 mouse, 19 rat, and 1 rabbit) and 82 α chains (46 human, 35 mouse, and 1 rabbit).

To restrict analysis to V genes only, thus eliminating the bias of inflating variation in the joining regions at the expense of V gene segments, we established a second data base. Data base II consists of unique V genes of α and β chains from the first data base arbitrarily truncated three residues downstream of the cysteine at position 90/92 (residue numbering follows ref. 23 for each chain type). The set of immunoglobulin V sequences of this data base is derived from the National Biomedical Research Foundation (ref. 44, release 57) and GenBank data bases (ref. 45, release 17), excluding sequences sharing >95% identity with a previously included one to ensure that each V gene is represented only once. As a result, this data base consists of a total of 94 immunoglobulin H-chain, 111 κ-chain, 41 λ-chain, 87 TCRα, 127 TCRβ, and 72 cytochrome *c* (GenBank) sequences. Manual alignments were made on the basis of the 40 invariant amino acids crucial for the conformation of immunoglobulin V domains (17), by introducing a minimum of gaps.

Calculation of the modified variability index was performed for each position according to the formula $NP \times NPd/NP_{freq}$, where NP is the total number of amino acid pairs [calculated from the number of sequences (n) in the sample; $NP = n \times (n - 1)/2$], NPd is the number of distinguishable pairs, and NP_{freq} is the number of times the most frequent pair occurs in the sample.

As an objective criterion for the identification of localized regions of hypervariability, we eliminated isolated hypervariable sites by submitting the variability indices to a filtering algorithm based on a truncated Fourier series expansion. The chosen number of Fourier coefficients (20 coefficients) is that which results in a fair correspondence of valleys and peaks of the smoothed curve with β-strands and connecting peptides of the immunoglobulin fold. The threshold to define hypervariable regions in TCR β chains was set at 1 rms deviation of the filtered data.

A three-dimensional model of the TCR α/β dimer was built from the crystallographic coordinates of Fab J539 (46) by using program FRODO (47) running on an Evans & Sutherland PS390 graphic display system. To validate the model, we ascertained that the periodicity of variability in β-strands determined in the plots matches the solvent exposure of the side chains in the model.

All computer programs were run on a MicroVax II microcomputer.

RESULTS AND DISCUSSION

Our starting assumption was that the failure to highlight discrete hypervariability regions in TCR sequences was due primarily to their more distant ancestry relative to the immunoglobulin chains (48). The steady accumulation of nonadaptive amino acid substitutions during evolutionary time, in spite of an inherent structure-preserving bias in the choice of replacements, inevitably tends to raise the variability score of the less rapidly evolving positions. The histograms of Fig. 1 summarize for each V gene family the proportions of pairwise sequence combinations falling in various intervals of percent amino acid identity. Although the distributions of Vα and Vβ manifest slight differences in spread, both appear more homogeneous than immunoglobulin V regions and are clustered near the modal value of 30% amino acid identity (i.e., even lower than the minimum level of amino acid identity in immunoglobulin V chains). As the sequences are derived from different mammals, the distributions show for Vα and Vβ chains that there is as much dissimilarity between species as there is among genes of the same species. Such results are

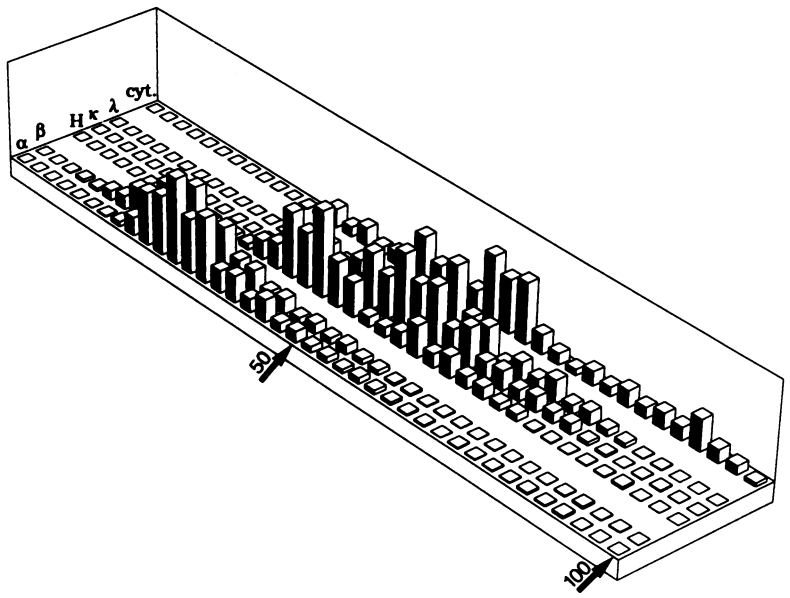

FIG. 1. Distributions of amino acid identity levels within V gene families. Percent of amino acid identity was determined from pairwise comparisons of sequences from data base II. The height of the bars expresses in percent the number of sequence pairs falling in a bin with the given identity level. cyt., cytochrome *c*.

as expected for members of a gene family having arisen from remote amplification events and diverged without undergoing substantial duplications in different species. The approach of selecting Vβ sequences displaying a high degree of amino acid identity—comparable to those of immunoglobulin V regions (bins with the most identical pairs in the tail of the distribution of Fig. 1)—met with the problem of extracting reliable data from subsets that were too small.

Description of diversity embodies a dual concept: the variety (that is, the number of distinguishable items in a collection) and their relative abundance (that is, frequency distribution). The Wu–Kabat coefficient of variability of amino acid positions is biased by the degree of dominance of the most common amino acid. Since it is calculated as the number of different amino acids at a given position divided by the frequency of the most common amino acid at this position (2), it fails to use information on the relative abundance of the various amino acids other than the most common one in the sample. An example may help to clarify this point. Two collections of 16 amino acids such as aaaaaaabbbcccddd and aaaaaaabbbbbbbcd have the same Wu–Kabat coefficient (9.14) but differ according to the modified coefficient ($120 \times 10/21 = 57.1$ and $120 \times 8/49 = 19.5$, respectively). Due to the finiteness of amino acid number and to structure-preserving restrictions in the process of amino acid substitutions, it is plausible that, over evolutionary time, the degree of diversity of an amino acid position will be inevitably influenced by the abundance relationships of the various amino acids rather than by their variety. Only among more recently diverged proteins (e.g., immunoglobulin V regions) is variety a sufficient indicator of diversity.

A trivial modification of the Wu–Kabat coefficient that takes into account the frequency of more than one amino acid consists of replacing single amino acids with pairs in their equation. Thus our diversity index is defined as the number of distinguishable amino acid pairs occurring at a given position divided by the frequency of the most common amino acid pair at that position. As each amino acid is paired—except with itself—with all other amino acids in the sample, regardless of identity, n sequences will yield a total of $n \times (n - 1)/2$ pairs. Incidentally, the computation of pairs introduces a conspicuous scale-up effect as it replaces the number of different amino acids in the numerator of the Wu–Kabat formula (maximum 20 if each amino acid is represented) with the number of distinguishable pairs (maximum 210 if each amino acid is represented at least twice) and because of the pronounced effect of the number of sequences in the sample

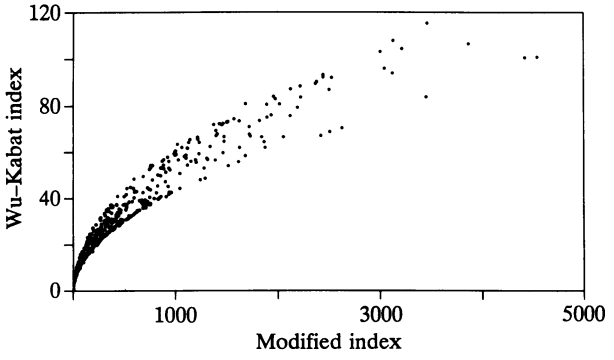

FIG. 2. Relationship between the Wu–Kabat index and the modified index of variability. Variability indices for 813 positions from various sets of related sequences from data base II were determined by both methods and plotted against each other. The shape of the distribution is likely to arise because the degree of diversity of closely related sequences is affected primarily by the number of different amino acids, whereas, for more distantly related ones, diversity depends on number and relative frequencies of amino acids.

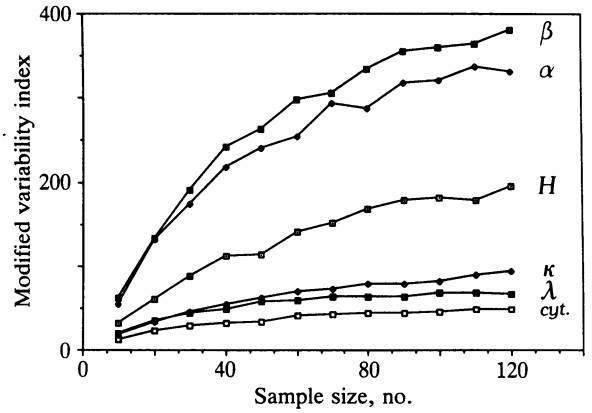

FIG. 3 Sample size dependence of the modified variability index. Variability indices, averaged along the entire sequence for each gene family, are plotted as a function of sample size. A given point $X(n)$ was obtained as the arithmetic mean of 20 runs on different sets of n sequences randomly taken from data base II. cyt., cytochrome c.

on the numerical value of the coefficient. However, the significance of the improvement is one of sensitivity rather than of scaling. In Fig. 2 the relationship between the two diversity indices is sought by plotting values obtained with both equations on 500 amino acid positions from diverse gene families. The scattering of the data points clearly shows that the two variability coefficients do not describe the same

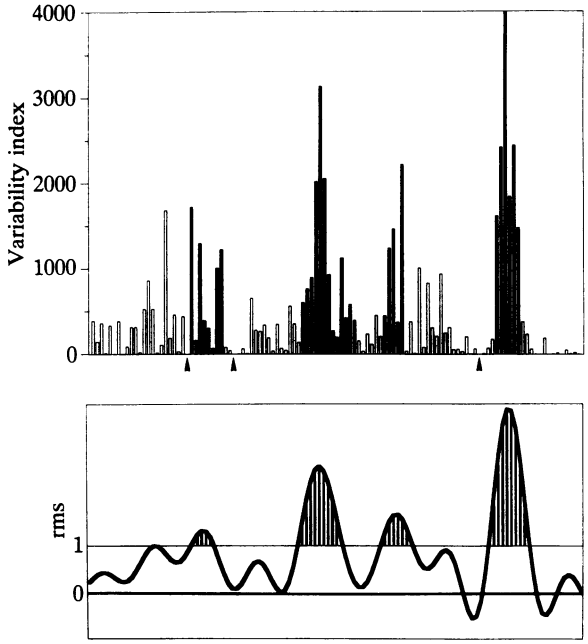

FIG. 4. Variability plot of TCR β chains. Variability indices calculated according to the modified formula are plotted for the 159 TCR β chains from data base I (*Upper*). Positions 30A, 62A, and 63 and positions 100–102 (ref. 23) are not shown because a gap was inserted in >50% of the aligned sequences. Regions of hypervariability have been objectively identified by applying on the data a smoothing algorithm based on truncated Fourier series expansion (*Lower*). The number of coefficients (20 coefficients) was chosen so that valleys and peaks correspond roughly to β-strands and connecting loops of the immunoglobulin fold. The smoothed plot is expressed in units of rms deviations of the filtered data (1 rms = 1190). A reference line at 1 rms isolates four clusters of hypervariable positions marked by hatched bars. The exact boundaries of these clusters were arbitrarily defined on the unfiltered plot where hypervariable regions are marked by solid bars. Arrowheads indicate the two invariant cysteines (positions 23 and 92) and the invariant tryptophan (position 34).

parameter. Furthermore, the shape of the distribution of the data points emphasizes the superiority of the modified coefficient for better resolving the Wu–Kabat values that fall in the high-variability range.

Nevertheless, the pronounced influence already alluded to of the number of sequences on the value of the modified coefficient prompted us to examine directly its dependence on sample size. As a way to simulate a range of amino acid positions of low and high diversity, we computed arbitrary means by averaging values of variability along the entire sequence of diverse gene families. As expected, the graph of Fig. 3 shows that the rate by which the value of the diversity index increases to an asymptote varies among gene families (i.e., the higher the true variability, the stronger the dependence of the diversity index on sample size). Whereas reliable ranks of high and low diversity indices can be determined on a rather small set of sequences, larger samples must be collected to render high variability coefficients only marginally sensitive to sample size. By comparison, the approach to the asymptote of the Wu–Kabat coefficients is obtained with samples of sequences roughly one and one-half times smaller (data not shown).

Application of the modified diversity index to the set of V-region sequences of TCR β chains from data base I resolves four discrete stretches of hypervariable positions that are objectively defined after Fourier filtering (Fig. 4). To visualize the possible spatial relationships among these hypervariable segments and their structural significance, we determined their localization in a three-dimensional model (Fig. 5) of the Vβ polypeptide constructed by analogy with the crystal structure of V region of light chains (cf. refs. 16 and 17). Each hypervariable stretch checks out with each of the four loops or connecting peptides that link the strands of the two β-pleated sheets at the top of the molecule. Thus we refer to these loops according to the label of the β strands (49) bounding them: BC (positions 24–31) links strands from different sheets, while C'C'' (positions 50–62), DE (positions 70–74), and FG (positions 96–105) are hairpin turns that link strands of the same sheet. Compared to the structure of the immunoglobulin V regions, the location of hypervariable regions BC, C'C'', and FG correspond closely to the canonical CDR1, -2, and -3, respectively. Interestingly, regions DE in the recently resolved crystal structure of the idiotype–anti-idiotype (Fab–Fab) complex (50) are part of the combining site of each interacting Fab (E255 and D1.3). An odd region of hypervariability in human immunoglobulin VH sequences was noticed between positions 84 and 91 by Capra and Kehoe (51). Although the functional significance of this segment remains unclear (52), it is remarkable that the four hypervariable segments of the Vβ sequences cluster together in the model at the tip of the molecule and that loop DE contributes to the accessible surface, lying opposite to the sheet–sheet interface of association with the Vα domain.

It is common knowledge that in protein evolution the greatest variations in amino acids occur at positions in the polypeptide where the side chain is on the surface. Such a tendency can be nicely appreciated, for instance, in the alternating degree of variability among the first residues of the Vβ chains (see Fig. 4) where each second residue of the β-strand A is expected to point toward the interdomain sheet. To appraise the contribution of this topological bias to the variability in β chains, we introduced a scheme that essentially reduces the variability score by a factor proportional to solvent exposure (calculated as described in ref. 53). The effect of this surface-dependent penalty was an overall reduction of variability in atypical positions, such as at residue 18 (see Fig. 4), without substantial modification in the hypervariability stretches, reinforcing the distinctive character of these regions.

It is significant that the modified variability plot of the 82 Vα sequences in data base I differs from that obtained with the Vβ sequences. Distinct peaks of variability cannot be unambiguously discerned. Although the variability is high overall, it is striking that at positions corresponding to hypervariable regions in other chain types, the variability in Vα sequences is lower. The relation of these findings to the

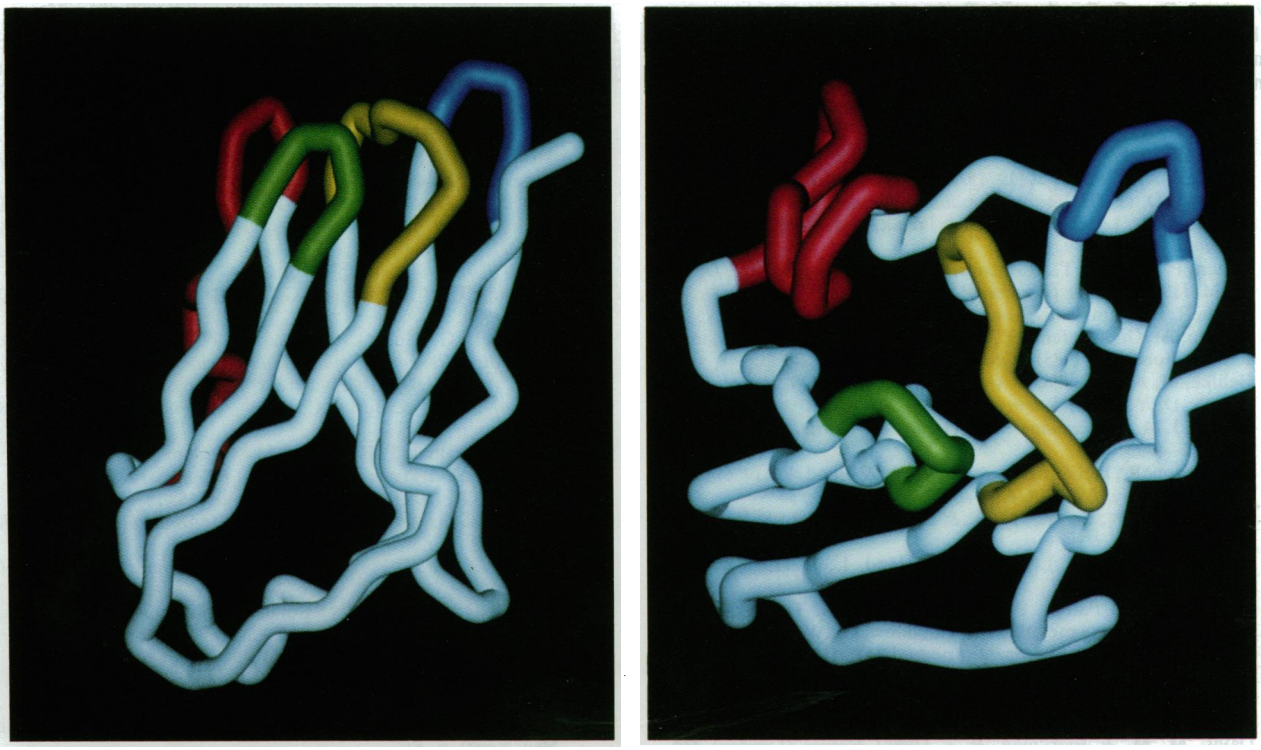

FIG. 5. Topological relationship of TCR β chain hypervariable regions. Graphic display of the three-dimensional Cα backbone structure of the TCR Vβ domain. Hypervariable regions are highlighted in colors. BC, yellow; C'C'', red; DE, green; and FG, blue. (*Left*) Lateral view. (*Right*) Top view.

factors that underlie the appearance of hypervariable regions in germ-line cells deserves further study. Nevertheless, the observations that hypervariability regions are not merely a consequence of surface exposure and that Vα sequences are at odds with all other V gene families suggest that hypervariable regions may not merely represent the accumulation of amino acid replacements that do not interfere with a backbone structure and are crucial to antigen receptor function. Incidentally, the data are difficult to reconcile with the hypothesis that V-region hypervariability is driven by positive selection.

Turning to the properties of the improved coefficient of variability, a few additional advantages are worth mentioning. Definition of diversity on the basis of binary comparisons opens the possibility of merging distinct homology subgroups in a single variability plot. Binary comparisons are made only within groups to yield sets of amino acid pairs that can be pooled and processed in bulk, thus overcoming the problem of calculating reliable variability coefficients from a collection of too exiguous subgroups. Finally, the definition obviates the need for a globally optimized alignment to obtain a unique variability plot, if all pairwise comparisons are referred to an arbitrarily chosen reference sequence. We have tested possible improvements of the modified equation by using ternary and higher order comparisons as a way to take into fuller account the frequency distribution of the different amino acids. The effect is a further accentuation in the scale amplification and in the resolving power of the index in the range of high variability scores. However, an increasing unevenness in the response of the amino acid positions leads to a progressive degradation in the discreteness of the variability plot. Finally, further improvements might be suggested by analogy with diversity measures from other fields. Nei's index of gene diversity (54), used to compare genetic variation among loci or populations, was tested but it proved less discriminative than the Wu–Kabat coefficient. However, the various measures of variability used to describe species abundance in the ecological literature (55) have not been explored.

We thank Drs. G. Bentley, M. Tosi, and O. Acuto for comments on the manuscript. Financial support from the Institut Pasteur and from the Fédération Nationale des Groupements des Entreprises Françaises dans la Lutte Contre le Cancer is acknowledged.

1. Kabat, E. A. (1968) *Proc. Natl. Acad. Sci. USA* **59**, 613–619.
2. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–250.
3. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* **190**, 382–393.
4. Capra, J. D. (1971) *Nature (London) New Biol.* **230**, 61–63.
5. Alzari, P. M., Lascombe, M.-B. & Poljak, R. J. (1988) *Annu. Rev. Immunol.* **6**, 555–580.
6. Patten, P., Yokota, T., Rothbard, J., Chien, Y., Arai, K. & Davis, M. M. (1984) *Nature (London)* **312**, 40–46.
7. Barth, R. K., Kim, B. S., Lan, N. C., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I. L. & Hood, L. (1985) *Nature (London)* **316**, 517–523.
8. Behlke, M. A., Spinella, D. G., Chou, H. S., Sha, W., Hartl, D. L. & Loh, D. Y. (1985) *Science* **229**, 566–570.
9. Leiden, J. M. & Strominger, J. L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4456–4460.
10. Concannon, P., Pickering, L. A., Kung, P. & Hood, L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6598–6602.
11. Tillinghast, J. P., Behlke, M. A. & Loh, D. Y. (1986) *Science* **233**, 879–883.
12. Arden, B., Klotz, J. L., Siu, G. & Hood, L. E. (1985) *Nature (London)* **316**, 783–787.
13. Becker, D. M., Patten, P., Chien, Y., Yokota, T., Eshhar, Z., Giedlin, M., Gascoigne, N. R. J., Goodnow, C., Wolf, R., Arai, K. & Davis, M. M. (1985) *Nature (London)* **317**, 430–434.
14. Hayday, A. C., Diamond, D. J., Tanigawa, G., Heilig, J. S., Folsom, V., Saito, H. & Tonegawa, S. (1985) *Nature (London)* **316**, 828–832.
15. Davis, M. M. & Bjorkman, P. J. (1988) *Nature (London)* **334**, 395–402.
16. Novotny, J., Tonegawa, S., Saito, H., Kranz, D. M. & Eisen, H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 742–746.
17. Chothia, C., Boswell, D. R. & Lesk, A. M. (1988) *EMBO J.* **7**, 3745–3755.
18. Claverie, J.-M., Prochnicka-Chalufour, A. & Bougueleret, L. (1989) *Immunol. Today* **10**, 10–14.
19. Schiffer, M., Wu, T. T. & Kabat, E. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4461–4466.
20. Bougueleret, L. & Claverie, J.-M. (1987) *Immunogenetics* **26**, 304–308.
21. Alt, F. & Baltimore, D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4118–4122.
22. Tonegawa, S. (1983) *Nature (London)* **302**, 575–581.
23. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (U.S. Public Health Service, National Institutes of Health, Bethesda, MD).
24. Leiden, J. M., Fraser, J. D. & Strominger, J. L. (1986) *Immunogenetics* **24**, 17–23.
25. Duby, A. D. & Seidman, J. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4890–4894.
26. Kimura, N., Toyonaga, B., Yoshikai, Y., Triebel, F., Debre, P., Minden, M. D. & Mak, T. W. (1986) *J. Exp. Med.* **164**, 739–750.
27. Chou, H. S., Behlke, M. A., Godambe, S. A., Russell, J. H., Brooks, C. G. & Loh, D. Y. (1986) *EMBO J.* **5**, 2149–2155.
28. Behlke, M. A. & Loh, D. Y. (1986) *Nature (London)* **322**, 379–382.
29. Yoshikai, Y., Kimura, N., Toyonaga, B. & Mak, T. W. (1986) *J. Exp. Med.* **164**, 90–103.
30. Singer, P. A., McEvilly, R. J., Noonan, D. J., Dixon, F. J. & Theofilopoulos, A. N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7018–7022.
31. Kimura, N., Toyonaga, B., Yoshikai, Y., Du, R.-P. & Mak, T. W. (1987) *Eur. J. Immunol.* **17**, 375–383.
32. Klein, M. H., Concannon, P., Everett, M., Kim, L. D. H., Hunkapillar, T. & Hood, L. E. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6884–6888.
33. Mage, R. G., McCartney-Francis, N. L., Komatsu, M. & Lamoyi, E. (1987) *J. Mol. Evol.* **25**, 292–299.
34. Rupp, F., Brecher, J., Giedlin, M. A., Mosmann, T., Zinkernagel, R. M., Hengartner, H. & Joho, R. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 219–222.
35. Spinella, D. G., Hansen, T. H., Walsh, W. D., Behlke, M. A., Tillinghast, J. P., Chou, M. S., Whiteley, P. J., Kapp, J. A., Pierce, C. W., Shevach, E. M. & Loh, D. Y. (1987) *J. Immunol.* **138**, 3991–3995.
36. Hochgeschwender, U., Simon, H.-G., Weltzien, H. U., Bartels, F. A. & Epplen, J. T. (1987) *Nature (London)* **326**, 307–309.
37. Kappler, J. W., Wade, T., White, J., Kushnir, E., Blackman, M., Bill, J., Roehm, N. & Marrack, P. (1987) *Cell* **49**, 263–271.
38. Chou, H. S., Whiteley, P. J., Kapp, J. A., Pierce, C. W., Shevach, E. M. & Loh, D. Y. (1987) *J. Immunol.* **138**, 3991–3995.
39. Malissen, M., Trucy, J., Letourneur, F., Rebai, N., Dunn, D. E., Fitch, F. W., Hood, L. & Malissen, B. (1988) *Cell* **55**, 49–59.
40. Morris, M., Barclay, A. N. & Williams, A. F. (1988) *Immunogenetics* **27**, 174–179.
41. Tan, K.-N., Datlof, B. M., Gilmore, J. A., Kronman, A. C., Lee, J. H., Maxam, A. M. & Rao, A. (1988) *Cell* **54**, 247–261.
42. Triebel, F., Breathnach, R., Graziani, M., Hercend, T. & Debre, P. (1988) *J. Immunol.* **140**, 300–304.
43. Williams, C. B. & Gutman, G. A. (1989) *J. Immunol.* **142**, 1027–1035.
44. George, D. G., Hunt, L. T. & Barker, W. C. (1988) in *The National Biomedical Research Foundation Protein Sequence Database*, ed. Lesk, A. M. (Oxford Univ. Press, Oxford), pp. 17–26.
45. Bilofsky, H. S. & Burks, C. (1988) *Nucleic Acids Res.* **16**, 1861–1864.
46. Suh, S. E., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986) *Proteins* **1**, 74–80.
47. Jones, T. A. (1985) *Methods Enzymol.* **115**, 157–171.
48. Tonegawa, S. & Mak, T. W. (1987) *Annu. Rev. Immunol.* **5**, 585–620.
49. Williams, A. F. & Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6**, 381–405.
50. Bentley, G. A., Bhat, T. N., Boulot, G., Fischmann, T., Navaza, J., Poljak, R. J., Riottot, M.-M. & Tello, D. (1989) *Cold Spring Harbor Symp. Quant. Biol.* **54**, 239–245.
51. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 845–848.
52. Kindt, T. J. & Capra, J. D. (1984) *The Antibody Enigma* (Plenum, New York), p. 81.
53. Richards, F. M. (1985) *Methods Enzymol.* **115**, 440–464.
54. Nei, M. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3321–3323.
55. Peet, R. K. (1974) *Annu. Rev. Ecol. Syst.* **5**, 285–307.