# ATTEMPTS TO LOCATE COMPLEMENTARITY-DETERMINING RESIDUES IN THE VARIABLE POSITIONS OF LIGHT AND HEAVY CHAINS *

Elvin A. Kabat and T. T. Wu

*Departments of Microbiology, Neurology, & Human Genetics & Development, Columbia University, The Neurological Institute, Presbyterian Hospital New York, N.Y.*
*and Departments of Physics and Engineering Sciences Northwestern University Evanston, Ill.*

Examination of the sequences of Bence-Jones proteins and myeloma immunoglobulin light chains from humans and from mice has led to the recognition of the variable and constant regions, and the accumulation of data on heavy chains indicates that they too contain a variable and a constant region.[1-5] The variable region comprises approximately the amino terminal half of the light chain and the amino terminal quarter of the heavy chain, and it is these regions which are generally believed to be responsible for antibody complementarity. The genetic control of the constant regions of both chains is readily explainable on classical genetic principles, but the genetics of the variable regions is far from clear and no generally agreed upon concept of the genetic determination of antibody complementarity has as yet been formulated. The recognition of subgroups in the variable regions of human (Reference 2, p. 133),[6, 7] and mouse [8] light chains and in human heavy chains,[9, 11] from sequence analyses, mainly of the first 20–25 amino terminal residues, has led to the designation of genes for these subgroups. It is generally accepted that the light chain and the heavy chain are each under the control of two genes, one for the variable and one for the constant region, and that a translocation results in the joining of these two genes.[12] These conclusions, however, do not account for antibody complementarity nor do they localize the combining site to any specific portions of the variable region.

When the subgroups of the variable regions were first recognized, it was noted [7, 4] that certain positions, notably those near 30 and 91–96, showed greater variability than could be accounted for by the subgroups. As further sequences accumulated, it became clearer that there were two regions of hypervariability, one following cysteine 23 and the other following cysteine 88 and comprising residues 24–34 and 89–96 respectively, and it was of special interest that these two regions were brought into close proximity by the disulfide bond $I_{23}$-$II_{88}$, and that insertions or deletions occurred in these regions.[13, 4]

A more detailed analysis indicated the presence of three hypervariable regions. Franěk (Reference 5, p. 311) tabulated the positions showing non-homologous replacements and recognized a region from residues 50–55 in

* Aided by grants from the National Science Foundation, GB-8341 and GB-25686, a General Research Support Grant from the United States Public Health Service to Columbia University, and Biomedical Sciences Support grant FR 7028-03 from the National Institutes of Health to Northwestern University.

addition to the other two regions. A statistical analysis was made by Wu and Kabat [14, 5] of the complete and partial sequence data available on 77 Bence-Jones proteins and immunoglobulin light chains, considering human κ, human λ, and mouse κ chains of various subgroups as a single population which was aligned for maximum homology. Defining variability as

$$\frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}},$$

three hypervariable regions were found involving residues 24–34, 50–56 and 89–97. It was proposed that at least the first and third of these regions, and possibly all three, together with similar regions in the heavy chains might be the complementarity determining regions and that amino acid side chains in these regions might make contact with the antigenic determinant, the remainder of the residues of the variable region being essentially structural and involved in three-dimensional folding.

The significance of the three hypervariable regions was greatly reinforced by the findings of Weigert and coworkers [15] who examined ten mouse λ chains; from composition analyses on peptides, six had apparently identical sequences in the variable regions and the remaining four showed variation only in one or another of three hypervariable regions. FIGURE 1 shows the sequences of Weigert and associates superimposed on the original plot of Wu and Kabat.

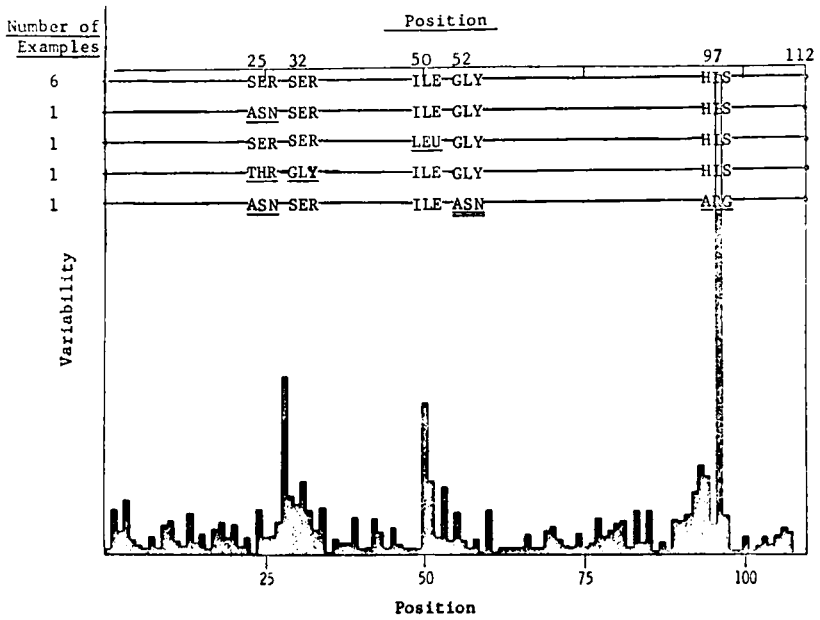


FIGURE 1. Mouse λ myeloma sequences of Weigert and coworkers [15] superimposed on the variability against position plot for human κ, human λ and mouse κ Bence-Jones proteins and light chains of Wu and Kabat.[14] The positions at which no differences among the mouse λ chains were found are indicated by a line. Sequences are given at positions at which differences occurred, and the number of nucleotide changes indicated by the underlining.[15]

It should be noted that their positions 50 and 52, if aligned for homology relative to our data, would be at 48 and 50. The mouse λ chains thus seem to be much more homogeneous in the variable region than mouse κ or human κ or λ chains, so that the changes appear to be restricted to the hypervariable regions.

It is also noteworthy that Capra and Kunkel [16] have found two cases of hypergammaglobulinemic purpura with antibodies of restricted specificity to have identical sequences for the first 40 N-terminal residues of their light chains that include the first hypervariable region; both chains were unusual in having Thr-Val at residues 13 and 14. It would be of great interest to establish whether the combining sites are identical and whether the sequence of the entire variable region of the light chains (and ultimately of the heavy chains) proves to be identical. Wang and colleagues [17, 18] have shown that in a patient with both a myeloma γG2κ and a macroglobulin γMκ protein, the κ light chains of both were identical in amino acid composition, bands in urea-starch gel, peptide maps, and in optical rotatory dispersion and circular dichroism. The light chains have not been sequenced, but the first 27 residues from the N-terminal region of the heavy chains and the idiotypic specificity of both proteins were identical.

We have continued to tabulate sequences of light chains and have now accumulated complete and partial sequences data on 121 human κ and λ and mouse κ and λ sequences. The tabulation only includes data for positions at which the sequence is reported as unequivocal. FIGURE 2 shows the variability vs. position plots for all 121 light chains as a single population (A), for 33 human κI chains (B), for 58 human κ chains (C), and for 58 human and 27 mouse κ chains (D). It is evident that the hypervariable regions may be seen in all plots. When two values are given for any position this is due to uncertainty in regard to Glx and Asx residues. It should be emphasized that apart from the N terminus, the number of complete sequences is still not as large as desirable and that the data for positions 40–85 are based on less than 20 proteins, while those at positions 1–23 are based on between 59 and 117 proteins.

It is of considerable interest that Singer and Thorpe found the invariant Tyr 86 of the light chain to be affinity-labeled in anti-DNP antibodies,[25] and Goetzl and Metzger [26] showed that position 34 in the above alignment (actually position 32) in the mouse λ myeloma protein with anti-DNP activity was labeled. Tyr 86 is very close to the hypervariable region, and position 34 is in the first hypervariable region. The identification of residues by affinity labeling in antibodies and myeloma proteins with different specificities will obviously be of importance in defining the relation of residues in the hypervariable regions to antibody complementarity.

Sequence data on heavy chains have been accumulating, and we have examined these data for hypervariable regions. Complete sequences of seven heavy chain variable regions, including five γG1 (Eu, He, Daw, Cor, Nie), 1 γM (Ou), one mouse myeloma protein (MOPC 173), and partial sequences of 7 γG1, 2 γG2, 2 γG3, 1 γG4, 8 γM, 2 γA, 1 γE, four rabbit, two horse and one shark, were available. As for light chains, subgroups of the variable regions of the heavy chains have been recognized and termed $V_{HI}$, $V_{HII}$, $V_{HIII}$, and $V_{HIV}$.[9, 20, 10, 11] The degree of homology of these variable region sequences is truly extraordinary. FIGURE 3A shows a plot of variability against position considering all the heavy chain sequences available as a single population. In aligning for maximum homology, gaps of two residues are placed between

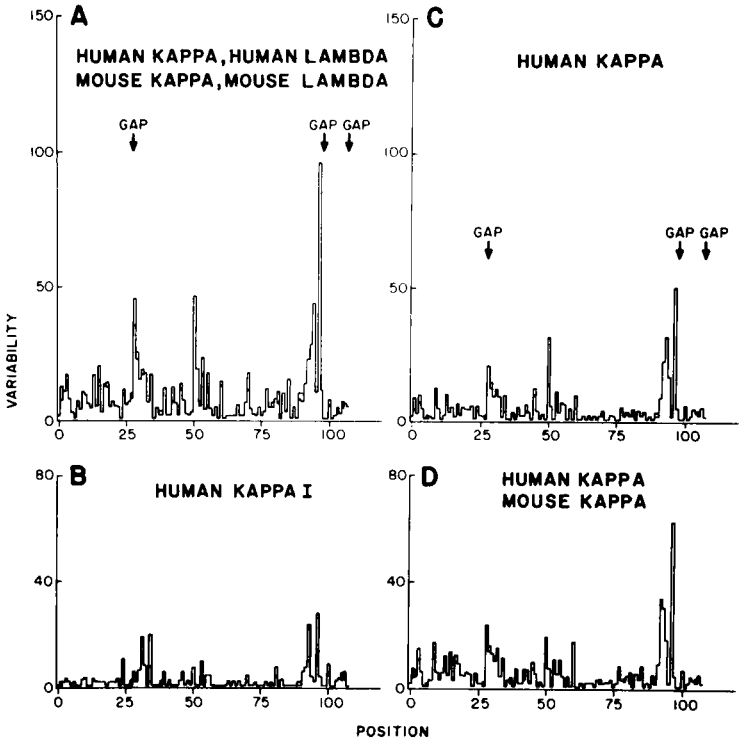


FIGURE 2. Variability at different amino acid positions for the variable region of light chains. GAP indicates positions at which insertions have been found. Data used: Wu and Kabat,[14] plus the following light chains: Pot, Die, Car, Tei, Joh (16); Dav, Fin (19); Ou (20); Hau (21); Til, Wil, Sal, Pom (22); Ful (23); G 173, F 31C, B J 149, B J 321, B J 63, A 603, A 870, A 384, A 467, F 47, H 37, B J 843, B J 674, A 8, A 15, B J 773, B J 265, GLP C 1, A 167 (24); S 104, X P 8, J 698, H 2061, J 558, HOPC 1, RPC 20, S 176, H 2020, S 178 (15).

residues 34 and 35 of $V_{HI}$, $V_{HIII}$, and mouse proteins, of one residue between positions 54 and 55, of three residues between positions 85 and 86 in some $V_{HII}$ proteins, and of three to five residues between positions 100 and 101 in certain $V_{HI}$, $V_{HII}$ and $V_{HIII}$ proteins. Hypervariable regions are seen as for the light chains. The first hypervariable region comprises residues 31–35, and the last hypervariable region involves residues 95–102; in both instances, as with the light chains, gaps occur in both regions. There also are indications of two other regions which are somewhat more variable. One of these runs from residues 50–65 and corresponds approximately to the third light chain hypervariable region but is considerably larger; and the fourth, which was noted by Drs. Capra and Keogh, includes residues 81 and 83–85. A similar distribution is seen in FIGURE 3B, in which only the human heavy chain sequences are plotted; in this instance the hypervariable region extends from 50–54 rather than from 50–65, although somewhat more variability is seen for residues 61, 62, and 64, and residues 81, 83, 84, and 85 also show some-

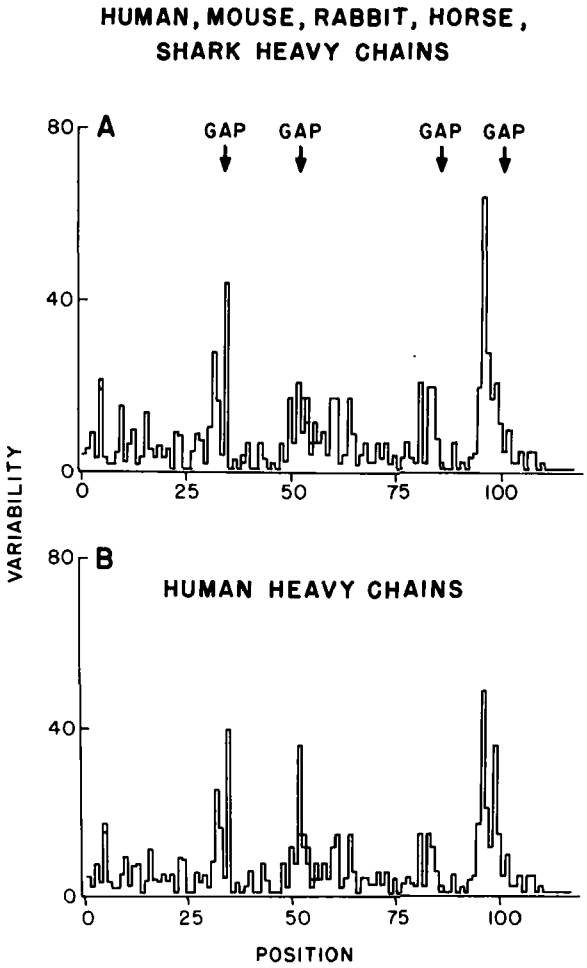


FIGURE 3. Variability at different amino acid positions for the variable regions of heavy chains. GAP indicates the positions at which insertions have been placed. Data used: Eu;[27] Ca;[28] Ste, horse γGab, horse γGT;[29] Dee;[30] He;[9] Daw, Cor;[31] Ou;[32] Car;[33] Sa;[34] Vin;[35] Til;[18] Sha;[36] Nie;[11] Tei, Was, Jon, Ben;[22] Fi, Vu;[37] Zuc;[38] Bus, Dau, Dos, Bal;[39] Di, Wo, Na, Hu, Re;[10] MOPC 173;[40] Rabbit;[41] Rabbit Aal, Aa2, Aa3;[42] Shark.[43]

what increased variability. Residue 5 in both plots shows substantial variability. The two major hypervariable regions are clearly evident. Unlike the light chain hypervariable regions, which begin after each Cys residue, the heavy chain hypervariability begins at residue 31 and 95 respectively, while the Cys residues are at positions 22 and 92. Thus while they too are brought into close proximity by the disulfide bond, there appears to be some displacement from the Cys residues.

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.