Proc. Nat. Acad. Sci. USA
Vol. 72, No. 6, pp. 2180–2184, June 1975

Diversity of Light Chain Variable Region Sequences among Rabbit Antibodies Elicited by the Same Antigens

(homogeneous antibody/antibody combining site/amino-acid sequences/hypervariable regions)

MICHAEL N. MARGOLIES*, L. EDWARD CANNON III*, A. DONNY STROSBERG†, AND EDGAR HABER*

* Departments of Surgery and Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Mass. 02114; and
† Chemie der Proteinen, Vrije Universiteit Brussels, Belgium

Communicated by Herman Kalckar, March 20, 1975

ABSTRACT **We report the complete variable region sequences of three homogeneous rabbit antibody light chains and the partial sequences of five others. When these are compared to other published rabbit light chain sequences, two regions of markedly increased variability are revealed, which are homologous in position to the first and third hypervariable regions of murine and human myeloma light chains. In addition, there is increased variability among the first three residues at the amino-terminal end. A hypervariable region homologous to that identified at positions 50 to 56 in myeloma light chains is not present in these rabbit antibody light chains. The available three-dimensional models of Fab fragments based on x-ray crystallography indicate that neither the amino-terminal portion of the light chain nor the region homologous to positions 50 to 56 forms a part of the combining site.**

Comparison of the hypervariable regions among six light chains from antibodies to Type III pneumococcal polysaccharide and among four from antibodies to Type VIII pneumococcal polysaccharide suggests that a large number of different sequences may be found in antibodies specific for these relatively simple antigens.

Certain residues outside of the hypervariable regions are invariant in the rabbit light chains and correspond to residues that are required for proper chain folding in human and murine myeloma light chains, indicating that the general conformation of myeloma light chains is the same as that of light chains of elicited antibodies.

An examination of amino-acid sequences from many paraprotein light chains (L chains) (1) and heavy chains (H chains) (2, 3, 36) has revealed segments of "hypervariability" within the variable region, which are thought to include residues involved in antigen binding. X-ray crystallographic studies of hapten-binding human (4) and murine (5) myeloma protein Fab fragments reveal that amino-acid residues from five of these seven hypervariable (hv) regions line the binding cavity. A comparison of the amino-acid sequence of several L chains from homogeneous antibodies would permit a direct test of hypotheses concerning antibody complementarity derived from the studies of myeloma light chains. Prolonged and repetitive immunization of rabbits with bacterial vaccines (6, 7) results in the frequent production of monoclonal antibodies that are suitable for determination of primary sequence. Occasionally responses to hapten-protein conjugates are also monoclonal (8). The amino-acid sequences of the light chain variable region (V_L) in rabbit antibody to *p*-azobenzoate (8), group C streptococcus (9), and two antibodies specific for Type III pneumococcal polysaccharide (S3) (10, 11) have been described. We report here complete V_L amino-acid sequences of a Type VIII pneumococcal polysaccharide (S8)-specific antibody, two S3-specific antibodies, as well as partial V_L sequences for three additional S8 and one additional S3 antibody and an antibody specific for *Micrococcus lysodeikticus* (*M. lyso.*) cell wall (12). These are aligned and compared with the published L chain sequences in an effort to reveal those regions of the sequence that are related to antigen specificity and those portions of the L chain required for V_L tertiary structure.

Abbreviations: H chain, immunoglobulin heavy chain; L chain, light chain; V_L, light chain variable region; S3, Type III pneumococcal polysaccharide; S8, Type VIII pneumococcal polysaccharide; hv, hypervariable; *M. lyso.*, *Micrococcus lysodeikticus*.

METHODS

Rabbits 3315 and 3322 were immunized with Type VIII pneumococcal vaccine, and rabbits 3374, 3368, and 3381 with Type III pneumococcal vaccine (13). Rabbit 120 was immunized with *M. lyso.* vaccine (12). Antibodies 3368, 3374, 3315, 3381, and 120 were isolated from serum by chromatography on DEAE-cellulose. Antibodies 3322A, 3322B, and 3322C were unique components isolated from the antiserum of a single rabbit by elution from an S8-bovine-Ig-Sepharose immunoadsorbent with cellobiose (14). The L chains of antibodies 3322A, B, and C reacted with b_5 allotype-specific antisera; all other L chains reacted with b_4 allotype-specific antisera. All antibodies demonstrated a single L chain on polyacrylamide disc gel electrophoresis. The strategy for the V_L sequence determination depended on tryptic cleavage of L chains at arginines after protection of lysine residues. The three to six peptides thus obtained were fractionated by gel filtration. In the case of L chain 3315 where the entire L chain sequence was obtained, additional tryptic peptides of unblocked light chain as well as chymotryptic and peptic peptides were sequenced. All peptides larger than 10 residues in length were sequenced in a Beckman 890B Sequencer using a modified 0.1 M Quadrol peptide program or a 0.1 M Quadrol protein microsequence program (37). Smaller peptides were sequenced by manual Edman degradation. The phenylthiohydantoin derivatives were identified by gas–liquid chromatography and on silica gel plates or polyamide sheets (15), thus providing dual identification of virtually all residues, including unequivocal assignment of amides. Carboxyl-terminal residues were confirmed by digestion with carboxypeptidases A and B. The ordering of the fragments was aided by sequenator runs on the intact L chain extending 38 to 58 steps and

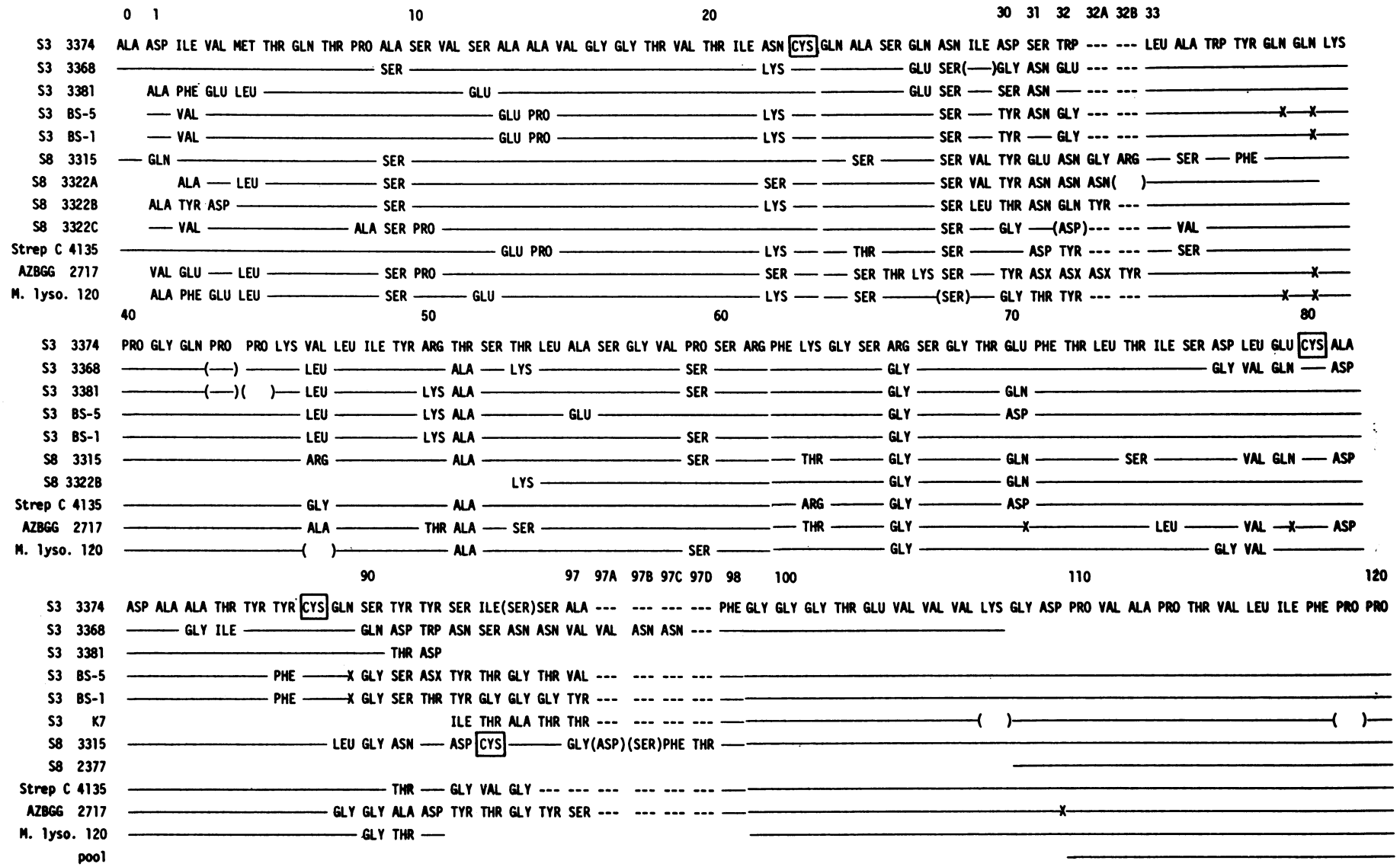

FIG. 1. V_L sequences from rabbit homogeneous antibodies are aligned by homologous invariant residues. The first five residues of K7 are aligned arbitrarily because this partial sequence begins in an hv region. Numbering is according to the human $V_{\kappa I}$ Ag. A horizontal line indicates residue identity with sequence 3374; a broken line indicates a deletion. Residues in parentheses were identified by a single method only. An X indicates that the amide or acid was not distinguished. The amino-acid sequence of chains BS-1, BS-5, 4135, and 2717 (AZBGG) were reported by others (8–11). We have reported the amino-acid sequences of the fragments of K7, nonimmune Ig L chain (pool), and 2377 previously (19). The sequences of 3374, 3368, 3315, 3381, 3322A, B, and C, and 120 are newly reported here.

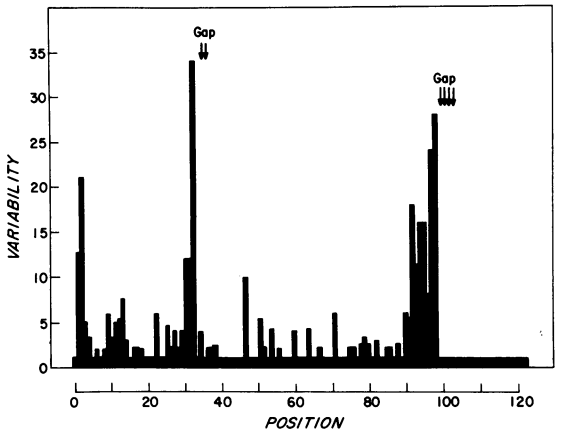

FIG. 2. Variability plot according to the method of Wu and Kabat based on the data from Fig. 1, except for the first 20 positions, for which data from over 50 rabbit antibody L chains are included (21). Variability is defined as the number of different amino acids at a given position divided by the frequency of the most common amino acid at that position.

on the products of the specific acid cleavage between aspartic acid 109 and proline 110 (16). The details of antibody purification, isolation of peptides, and amino-acid sequence analysis will be reported separately‡.

RESULTS

The complete V_L sequences for the S3 antibodies 3374 and 3368 and the S8 antibody 3315 are shown in Fig. 1, as well as partial sequences for the S3 antibody 3381, the S8 antibodies 3322A, 3322B, and 3322C from the same antiserum, and the *M. lyso.* antibody 120. These have been aligned by homology with the V_L sequences of S3 antibodies BS-1 (10) and BS-5 (11), streptococcal C antibody 4135 (9), and *p*-azobenzoate antibody 2717 (8). A variability plot according to the method of Wu and Kabat (1), based on the data in Fig. 1, is depicted in Fig. 2. Inspection of the sequences and the variability plot reveals three segments of significant variability: (*a*) The amino-terminal three residues show several amino-acid sequence alternatives in addition to three different chain lengths. Similar findings have been reported previously (17). (*b*) A region of marked variability is seen beginning at position 30, which is characterized by variation in both length and sequence. The length of this hv region is between three and five residues when the chains are aligned by cysteine 23 and the relatively constant sequence that begins at leucine 33. (*c*) The sequences of positions 91 through 97 are also unique to each chain and this hv region is between 5 and 11 residues in length when it is aligned by the constant cysteine 88 and the invariant sequence beginning at phenylalanine 98. None of the hv region sequences of the six S3 antibodies nor those of the S8 antibodies are identical. The amino-acid sequence outside of these hv regions reveals either invariant residues or limited substitutions except for position 46, where five alternatives occur. Significant variability is not present in the region that includes residues 50 through 56, which are homologous to the second hv region of myeloma L chains described by Franek (18) and by Wu and Kabat. The amino-acid sequence carboxyl-terminal to phenylalanine 98 is invariant in all the rabbit by L chains examined, including non-immune L chains and the S3 chain K7 and S8 chains 2377 and 2388, which we reported previously (19).

‡ By the authors and F. W. Chen, K. J. Fraser, and M. Van Hoegaerden.

DISCUSSION

On the basis of a statistical examination of human and murine myeloma L chain sequences, Wu and Kabat concluded that there were three areas of hypervariability within V_L (1). It was suggested that some of these segments were folded together in the intact antibody so as to form the antigen-binding site. Four such regions have also been described in paraprotein H chains (2, 3, 36). Affinity-labeling studies have supported this model by demonstrating that many of the labeled residues were located within or near the hv regions (20). Recent x-ray crystallographic analyses of human myeloma Fab' fragment New (4), which binds vitamin K_1OH, and of the Fab fragment of the phosphorylcholine-binding murine myeloma McPC 603 (5) have shown that *two* of the V_L hv regions and *three* of the H chain hv regions form the combining site. Correlation of these data with an examination of the available rabbit L chain sequences from homogeneous elicited antibodies permits a critical evaluation of the relationship between hv region sequence and antigen specificity.

The hypervariability of the amino-terminal three residues, which is unique to the rabbit V_L, can only partially be accounted for by the three different rabbit chain lengths (17). The amino-acid sequence of the amino-terminal region appears to bear no relation to antigen specificity; this conclusion is further supported by a statistical study of amino-terminal sequences of more than 50 homogeneous rabbit L chains (21). For example, L chains 3381 and 120 from antibodies directed against two different antigens have identical sequences for the first four amino-terminal residues (Fig. 1). Similarly, identical sequences in this region are present in the S3-specific L chains 3368 and 3374 as well as the streptococcus C antibody L chain 4135. These findings, as well as the observations from x-ray crystallography that the L chain amino-terminus is not involved in the binding site, lead us to conclude that for this region hypervariability does not relate to antigen specificity. Comparisons among the rabbit L chains (Figs. 1 and 2) demonstrate two additional regions of hypervariability that are homologous to the *first* and *third* myeloma L chain hv regions, which have been shown by x-ray crystallographic analysis to include residues lining the binding site (4, 5, 22, 23). In the rabbit chains, these are located after positions 29 and 90, respectively. Each is characterized by variation in length. The length of the former hv region varies between three and five residues and ends at the invariant leucine 33, which is also present in human kappa chains (24). The hv region after position 90 is the longest hv segment of V_L, and varies in length from 5 (4135) to 11 (3315) residues in these rabbit L chains. X-ray crystallography has demonstrated that the size of the combining site depends largely on hv region length (4, 5). The homologous region in human myeloma kappa chains, however, shows no variation in length (24). This hv region in rabbit L chains ends at an invariant phenylalanine 98 which is also seen in the human and murine L chains.

An unexpected finding was the absence of significant variability in the nine rabbit V_L sequences (Fig. 1) at positions 50 to 56, homologous to the second hv region previously described (1, 18). On the basis of these sequences we would

predict that this region is not involved in the antigen-binding site in rabbit L chains. This contention is supported by the x-ray crystallographic analysis of McPC 603 (5) in which the second hv region does not line the binding site and does not interact with the hapten. The crystallographically determined structure of REI, a human V_L kappa dimer (23), may be superimposed on McPC 603 in this region (E. Padlan and D. Segal, personal communication), which indicates that there is no significant structural difference in the second hv region between these two proteins. Moreover, the homologous portion of the lambda chain of Fab' New adjacent to a seven-residue deletion is not involved in the binding site. Thus only two L chain hv regions participate in antigen binding.

Hypervariability *per se* is not a sufficient criterion to identify that portion of the sequence involved in binding to antigen, as exemplified by the rabbit V_L amino-terminal sequence variation and the apparent lack of an hv region at positions 50 to 56 in elicited rabbit antibody L chains. An analogous situation obtains for the H chain, where the third hv region identified by Kehoe and Capra (2) also is not involved in the combining site by x-ray crystallographic analysis (4, 5). Hypervariability of a given amino-acid residue may reflect a role in antigen binding or may represent a position outside of the binding site ("framework") (25), where, because of a lack of steric constraint, a change in amino-acid side chains may occur without altering chain folding. Position 46 in the rabbit L chains, where five amino-acid alternatives occur among eight chains, may represent such a position (Fig. 2).

If the first and third V_L hv regions contribute to the binding site structure, a comparison of these sequences among antibodies to a given determinant should reveal whether or not there is a single sequence for that determinant, or whether there are several possible structures fulfilling the requirement for binding. In six S3 antibody L chains for which hv sequence data are available (Fig. 1) the hv regions differ in sequence and, in the instance of 3368, in length. The S3 polysaccharide is an unbranched repeating polymer of the disaccharide cellobiuronic acid. Although the structure of this antigen is seemingly simple, the number of unique hv sequences in S3-specific L chains already appears to be large. Since it may be argued that even repeating polymers represent a number of antigenic determinants, an example of antibodies of still more limited antigen specificity was studied in rabbit 3322. This animal exhibited five to seven S8-specific antibodies in a single antiserum. Of these, 3322A, B, and C were eluted from an S8 immunoadsorbent with cellobiose, indicating that this disaccharide is immunodominant. Nevertheless, these L chains exhibit different amino-acid sequences and length in the hv region beginning at position 30, suggesting that there is more than one possible set of combining site sequences complementary to a given antigenic determinant. Moreover, the three chains differ in their amino-terminal sequence and chain length. Thus, if differences in sequence of the amino-terminal portion of the light chain represent different germ line genes (26), it appears that the same antibody specificity may be associated with different V region genes. This contrasts with the findings (27) that amino-terminal sequences of mouse myeloma H chains that bind the same antigen are either identical or very similar.

Our results clearly show that elicited antibodies of similar specificities have hv regions of markedly different sequence and length. With the available data, it is not possible by inspection to distinguish a V_L specific for S3 from one specific for S8. The diversity seen among hv regions of these *elicited* antibodies is in marked contrast with the observations that hv regions of H chains of phosphorylcholine-binding mouse myeloma proteins are identical (27). Of the four V_L first hv regions reported, three were identical. Cebra *et al.* (28) have shown restricted heterogeneity of sequence in the first and second hv regions of guinea pig H chains from pooled antibodies specific for a given hapten, and Capra and Kehoe have shown striking similarity in sequences of the entire V_H of two human immunoglobulins binding IgG (29). Not only does great diversity exist among rabbit antibody L chain hv regions, but comparison of a limited number of rabbit H chains indicates that diversity may also be a feature of their hv regions (21). The seemingly simple relationships shown in other systems may reflect selection on the basis of inbreeding (28), idiotypic identities (29), or the method and site of induction of a neoplasm (26, 27), which is not pertinent to the elicited rabbit antibodies reported here.

Changes in the circular dichroism pattern upon hapten-binding of each of three S3-specific antibodies isolated from the same antiserum have been reported (30). The difference spectra obtained indicated that either a tyrosine or tryptophan residue was involved in binding. These observations can be accounted for by the presence of tyrosine or tryptophan in all the S3 L chain hv regions sequenced thus far. Tyrosine, however, is found as the most common residue (16 times) in the hv regions of *all* of the rabbit L chains regardless of specificity. There is also a remarkable preponderance of asparagine, glycine, threonine, and serine. L chain 3315 is unusual because it has an extra free cysteine residue in the third hv region. This substitution has been reported for the first hv region in human myeloma λ chains X (31), Bau (32), and Du (33). It seems premature to associate a specific hv region residue with a given antigen specificity, contrary to the hypothesis suggested by the guinea pig sequence data (28).

The amino-acid sequence of the rabbit L chains outside of the hv regions demonstrates considerable conservation of structure. The rabbit κ_B L chains have seven half-cystine residues. The intradomain disulfide bridges (23 to 88 and 134 to 194) (34) are homologous to those of human and murine kappa chains. The extra disulfide bridge in κ_B chains joins position 80 in the V region with 171 in the C region (19, 34). The x-ray crystallographic structure of Fab' New and of the human L chain dimer Mcg (22) reveals that the homologous residues in these L chains are located at a distance from each other consistent with disulfide bridge formation. Thus the *inter*domain disulfide bridge is not inconsistent with the *intra*domain folding seen in L chains with two disulfide bridges.

In addition to the conserved cysteines, Wu and Kabat noted certain invariant residues: glutamine 6, tryptophan 35, proline 59, arginine 61, aspartic acid 82, tyrosine 86, phenylalanine 98, and glycines at positions 16, 41, 57, 64, 68, 99, and 101. Poljak and coworkers (4) and Epp and coworkers (23) have noted that the glycine residues are often involved in hairpin turns of the V_L polypeptide chain, which allow folding into antiparallel β-pleated sheets. The glycines are often preceded by a proline or serine residue. The same folding probably obtains in rabbit L chains, as these glycines are also invariant (Fig. 1). Several of the other invariant nonpolar residues are

located in the λ New structure with their side chains in the hydrophobic interior (4). All of these residues are present in the rabbit L chains, with the exception of position 59, where proline is replaced in several chains by a serine. The sequence beginning at phenylalanine 98 is entirely conserved in all the rabbit b_4 chains thus far examined (Fig. 1), including the region where the L chain may be cleaved into its variable and constant domains under mild acidic conditions (16, 19). Poljak and coworkers have reported that the residues in Fab' New λ chain homologous to rabbit L chain phenylalanine 98, glycine 99, and glycine 101 are invariant because the phenylalanine 98 is involved in an internal interchain hydrophobic pocket including a constant tryptophan from the H chain. This interaction is believed important for assembly. The glycine residues referred to are tightly packed and, therefore, cannot be replaced by amino acids that have a side chain.

Stanton *et al.* (35) have suggested that the sequence Arg-Phe-Ser-Gly-Ser-Gly homologous to positions 61 to 66 in the rabbit L chains reported here is highly conserved among many species. The x-ray crystallographic analysis of λ New indicates that side chains at positions homologous to 61, 62, and 64 are sterically hindered while those at the other positions are not. It is striking that, among the rabbit L chains reported here, the hindered positions are indeed invariant, while the others show variations from previously reported prototypes, indicating that tertiary structure is conserved among species while positions not required for maintenance of structure (residues 63 and 66) may vary (Fig. 1).

A preliminary account of a portion of this work was presented at the Miami Winter Symposium, January 1975. We are grateful to A. Brauer, I. Caplier, K. Dewoina, U. Lion, D. J. Perry, M. Vanderlinden, W. Verheulpen, and S. Wong for technical assistance. This work was supported by National Institutes of Health Grant AIO4967, a Grant-In-Aid from the American Heart Association, and Belgian Government Grants no. 02022 and no. 02023. M.N.M. is an Established Investigator of the American Heart Association.

1. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–250.
2. Kehoe, J. M. & Capra, J. D. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2019–2021.
3. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 845–848.
4. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 3440–3444.
5. Segal, D. M., Padlan, E. A., Cohen, G. H., Silverton, E. W., Davies, D. R., Rudikoff, S. & Potter, M. (1974) in *Progress in Immunology*, eds. Brent, L. & Holborow, J. (North-Holland Publishing Co., Amsterdam), pp. 93–102.
6. Haber, E. (1970) *Fed. Proc.* **29**, 66.
7. Krause, R. (1970) *Fed. Proc.* **29**, 59–65.
8. Appella, E., Roholt, O. A., Chersi, A., Radzimski, G. & Pressman, D. (1973) *Biochem. Biophys. Res. Commun.* **53**, 1122–1129.
9. Chen, K. C. S., Kindt, T. J. & Krause, R. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1995–1998.
10. Jaton, J.-C. (1974) *Biochem. J.* **141**, 1–13.
11. Jaton, J.-C. (1974) *Biochem. J.* **141**, 15–25.
12. Van Hoegaerden, M., Janssens, R. & Kanarek, L. (1973) *Arch. Int. Physiol. Biochim.* **81**, 210–211.
13. Chen, F. W., Strosberg, A. D. & Haber, E. (1973) *J. Immunol.* **110**, 98–106.
14. Cheng, W. C., Fraser, K. J. & Haber, E. (1973) *J. Immunol.* **111**, 1677–1689.
15. Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624–628.
16. Fraser, K. J., Poulsen, K. & Haber, E. (1972) *Biochemistry* **11**, 4974–4977.
17. Haber, E. & Strosberg, A. D. (1973) in *Specific Receptors of Antibodies, Antigens and Cells*, eds. Pressman, D., Tomasi, T. B., Grossberg, A. C. & Rose, N. R. (S. Karger, Basel), pp. 236–258.
18. Franek, F. (1970) in *Developmental Aspects of Antibody Formation and Structure*, eds. Sterzl, J. & Riha, I. (Academic Press, New York), pp. 311–313.
19. Strosberg, A. D., Fraser, K. J., Margolies, M. N. & Haber, E. (1972) *Biochemistry* **11**, 4978–4985.
20. Givol, D. (1973) in *Contemporary Topics in Molecular Immunology*, eds. Reisfeld, R. A. & Mandy, W. J. (Plenum Press, New York), pp. 27–50.
21. Haber, E., Margolies, M. N., Cannon, L. E. & Rosemblatt, M.S. (1975) in *Molecular Approaches to Immunology* (Academic Press, New York), in press.
22. Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973) *Biochemistry* **12**, 4620–4631.
23. Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974) *Eur. J. Biochem.* **45**, 513–524.
24. Gally, J. A. & Edelman, G. M. (1972) *Annu. Rev. Genet.* **6**, 1–46.
25. Cohn, M., Blomberg, B., Geckeler, W., Raschke, W., Riblet, R. & Weigert, M. (1974) in *The Immune System—Genes, Receptors, Signals*, eds. Sercarz, E. E., Williamson, A. R. & Fox, C. F. (Academic Press, New York), pp. 89–117.
26. Barstad, P., Farnsworth, V., Weigert, M., Cohn, M. & Hood, L. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 4096–4100.
27. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183**, 962–964.
28. Cebra, J. J., Koo, P. H. & Ray, A. (1974) *Science* **186**, 263–265.
29. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 4032–4036.
30. Holowka, D. A., Strosberg, A. D., Kimball, J. W., Haber, E. & Cathou, R. E. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3399–3403.
31. Milstein, C., Clegg, J. B. & Jarvis, J. M. (1968) *Biochem. J.* **110**, 631–654.
32. Baczko, K., Braun, D. G., Hess, M. & Hilschmann, N. (1970) *Hoppe-Seyler's Z. Physiol. Chem.* **351**, 763–767.
33. Buchwald, B. M. (1971) *Canadian J. Biochem.* **49**, 900–902.
34. Strosberg, A. D., Margolies, M. N. & Haber, E. (1974) *Fed. Proc.* **33**, 726.
35. Stanton, T., Sledge, C., Capra, J. D., Woods, R., Clem, W. & Hood, L. (1974) *J. Immunol.* **112**, 633–640.
36. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* **190**, 382–391.
37. Brauer, A., Margolies, M. N. & Haber, E. (1975) *Biochemistry*, in press.