Proc. Nat. Acad. Sci. USA
Vol. 72, No. 7, pp. 2785–2788, July 1975
Immunology

Evolutionary and structural influences on light chain constant (C_L) region of human and mouse immunoglobulins

(structure-function correlation and evolution/λ-κ comparisons/comparative structures)

ELVIN A. KABAT*, EDUARDO A. PADLAN†, AND DAVID R. DAVIES†

* Fogarty International Center and † Laboratory of Molecular Biology, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014; and * Departments of Microbiology, Human Genetics and Development and Neurology, Columbia University and the Neurological Institute, Presbyterian Hospital, New York, N.Y. 10032

Contributed by Elvin A. Kabat, April 17, 1975

ABSTRACT **A comparison of five constant region sequences of human and mouse κ and λ immunoglobulin chains has been undertaken in order to reveal sequence homologies and evolutionary relationships. Simultaneously, a comparison with the three-dimensional structure of one mouse κ-chain (McPC 603) has suggested structural reasons why many of the residues are invariant or conserved along κ versus λ lines. There are a number of residues that have remained invariant despite exposed positions for reasons that do not appear to be connected with the folding of this C_L domain.**

The constant region (C_L domain) of immunoglobulin light (L) chains contributes substantially to the functioning of an immunoglobulin molecule. While it is not involved directly in the specificity and complementarity of the antibody combining sites, it is joined to its counterpart in the heavy (H) chain, the C_H1 domain, by a variety of noncovalent interactions as well as in most immunoglobulins by a disulfide bond at its C-terminal or subterminal Cys. This -S-S- bond is not essential to L–H association, which is maintained noncovalently even after reduction and alkylation. In one immunoglobulin subclass of IgA2, the L–H bond does not occur, but an L–L dimer is formed, which remains noncovalently linked to the H chains. Bence Jones proteins in the form of L–L dimers also occur and may be held together by noncovalent forces (1).

There are two subclasses of light chains, κ and λ, which are present in almost all species examined; both are found in all five classes of immunoglobulins, IgG, IgM, IgA, IgD, and IgE, but each immunoglobulin molecule contains two identical κ or two identical λ chains.

The availability of the sequences of the C_L domain of human κ, human λ, mouse κ, and two mouse λ light chains (2), together with x-ray data on the three-dimensional structure of this domain (3–5) made it desirable to evaluate, if possible, structural influences on the evolution of these domains.

The present study attempts, residue by residue among these five chains, to relate preservation or variation of sequence to structure and function as evaluated from a three-dimensional model of the C_L and its interactions with the C_H1 domain. The findings show some interesting stretches of sequence in which invariance predominates, and others in which evolutionary divergence has been essentially along κ and λ lines. Positions at the surface of protein molecules that are accessible to solvent may undergo many mutational changes which do not affect three-dimensional folding, while residues that are buried tend to be invariant or highly conserved (6). In C_L domains mutations involving residues that contact the C_H1 domain also tend to be restricted.

The present study shows that residues preserved along κ and λ lines generally show conservative substitutions if in the interior of the domain or if buried. Fewer exposed residues which have diverged along κ and λ lines are homologous. Certain residues remain invariant despite an essentially exposed position and for no obvious reason. Of special interest is the observation that at only two and four positions, respectively, were human κ identical with human λ and mouse κ identical with mouse λ, while human κ and mouse κ were identical at 29 and human λ and mouse λ at 39 positions.

MATERIALS AND METHODS

The model of the Fab fragment of mouse McPC 603 constructed from x-ray data at 3.1 Å resolution was used (5) as well as published information on the C_L regions of a Bence Jones dimer (4) and human Fab fragment (3). Sequences of human κ, human λ, mouse κ, and two mouse λ chains were available (2). These sequences were aligned for maximum structural and sequence homology from residues 101 to 215 and modified to include the additional data reported (7). Each residue was located in the model of McPC 603 and classified according to whether it was: exposed, 0; mainly exposed, 1; partly exposed, partly buried, 2; mainly buried, 3; completely buried, 4; or in contact with C_H1, C. In addition, each position was classified as: invariant; four of five chains and three of five chains identical; human κ and human λ identical; mouse κ and mouse λ identical; human and mouse κ identical; human and mouse λ identical; and human κ, human λ, mouse κ and mouse λ different.

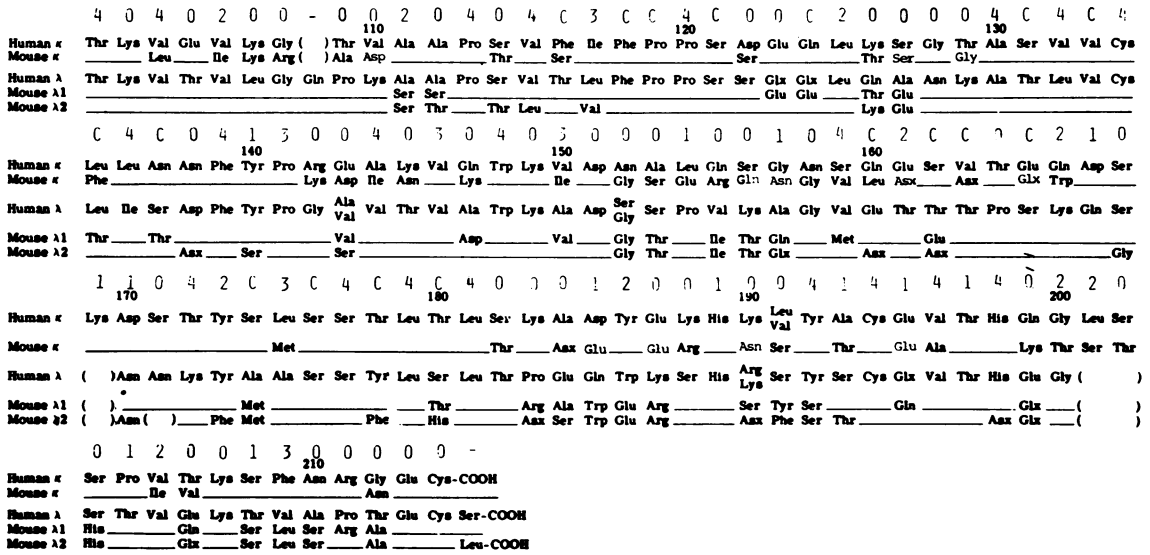
RESULTS

Table 1 lists the sequences of the five chains from positions 101 to 215. Above each residue is its classification from its position in the model.

Fig. 1 summarizes the sequence data in Table 1 with respect to the identities specified above. It is evident that clusters of invariant residues and those with 4/5 chains identical occur, notably at positions 118 to 123, 148 to 152 (excluding 150), 176 to 182 (180 3/5 identical), and 194 to 200 (exclud-

Abbreviations: C and V, constant and variable regions of immunoglobulin chains; L and H, light and heavy chains of immunoglobulins.

Table 1. Sequences of the switch and constant regions of human and mouse immunoglobulin light chains and their location from the model of mouse κ myeloma protein McPC 603 constructed from x-ray data



Sequence data are from Gally (2) with the mouse κ chain data revised according to Svasti and Milstein (7). Residues at 169 to 177 have been realigned to remove the gap in all λ chains at 178 and replace it by a gap at 169. Residues 201 and 202 have been moved to 203 and 204, leaving a gap at 201 and 202.

ing 195). There are also several clusters in which κ versus λ diversification predominates, 128 to 132 (130 invariant) and especially 164 to 172.

Of the 114 residues considered, 28 were invariant, 17 and 8, respectively, showed 4/5 and 3/5 chains identical. Of the kinds of identities of two chains in the remaining positions human κ and human λ were identical at only two positions and mouse κ and mouse λ at only four positions, while human and mouse κ were the same at 29 positions and human λ and one or both mouse λ chains at 39 positions. At two positions a human κ and a mouse λ chain were the same (*) and at three positions human λ and mouse κ chains had the same residue (●). At but eight positions were all four chains different.

Table 2 summarizes the sequence data in Fig. 1 in relation to the location of each residue in the model. The most striking finding is that of the eight positions at which the four chains had different amino acids (Fig. 1), six residues were completely exposed to solvent and the remaining two were mainly exposed. Moreover, of the two positions at which human κ and human λ were the same (one of which was the Oz marker at position 190) and of the four positions at which mouse κ and mouse λ were identical, all but one were completely or mainly exposed. The sixth, residue 135, was a contacting residue adjacent to invariant Cys 134. Three of the five positions in which human and mouse identities occurred but which were not both κ or both λ as indicated by the symbols (* and ●) in Fig. 1 were completely exposed;

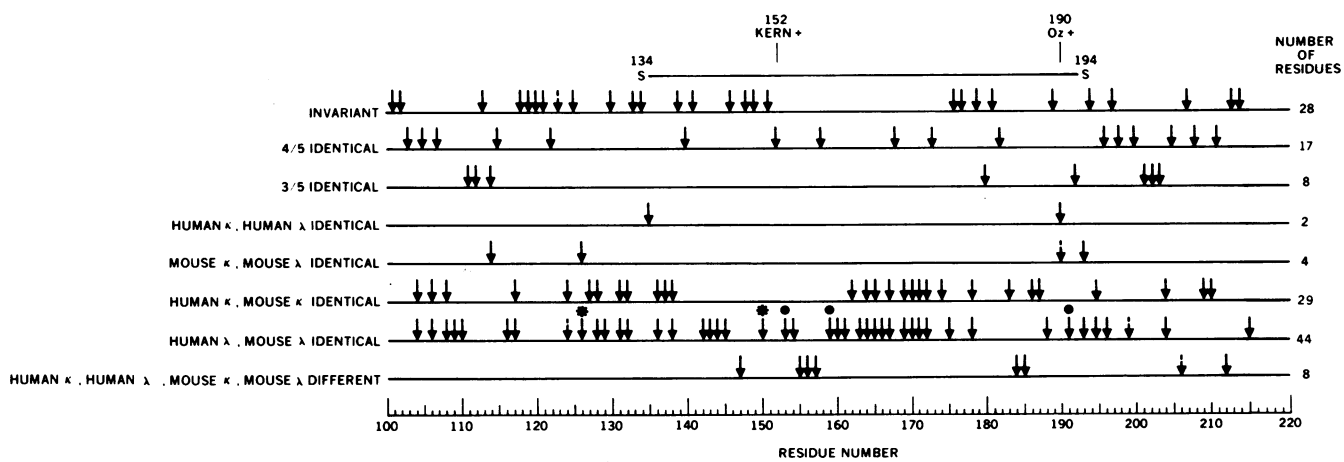


FIG. 1. Distribution of identical residues in the switch and constant regions of human and mouse immunoglobulin light chains. The arrows indicate identical residues in two or more of the five chains as well as at positions at which the four classes of chains differed. When more than one arrow occurs at a given position, there was identity among two sets of chains. Thus, at position 169 human and mouse κ chains had Lys while human λ and both mouse λ chains had a gap. Arrows with an * and a ● indicate the few residues identical in human κ and mouse λ and human λ and mouse κ, respectively. A dashed arrow indicates an Asx or Glx at that position in one or more chains.

Table 2. Location from x-ray structure and degree of evolutionary preservation of residues in the C_L domain and switch region of human and mouse light chains

		Identities—number of residues							
Location	Code	Invariant	4/5	3/5	Human κ Human λ	Mouse κ Mouse λ	Human κ Mouse κ	Human λ Mouse λ	Human κ, Human λ Mouse κ, Mouse λ different
Exposed	0	7	7	3	1	2	10	16	6
Mainly exposed	1	2	2			1	4	4	2
Partly exposed, partly buried	2	1	4	1			2	3	
Mainly buried	3	2		2			2	1	
Completely buried	4	11	4	1			3	5	
Contacts H chain	C	5		1	1	1	8	7	
Total		28	17	8	2	4	29	36‡	8

‡ Excludes residues with * and • in Fig. 1 and residue 215 not specifiable; gaps in κ at position 108 and in λ chains at positions 169 and 201 and 202.

these included the Inv marker 191. Position 152, which in human λ chains carries the Kern marker, was also exposed (3).

Of the 28 invariant residues 18 were contact residues or were mainly or completely buried, while nine were mainly or completely exposed. Considering together the 25 positions in which 4/5 and 3/5 residues were identical, 12 were completely or mainly exposed, only one was a contact residue, while five were completely buried.

Of the 29 positions at which human κ and mouse κ had the same amino acid and the six (footnote Table 2) at which human λ and one of the mouse λ chains had the same amino acid, seven in each were contacting residues and three and six, respectively, were completely buried while 15 and 20, respectively, were completely or mainly exposed.

Of the seven invariant exposed residues six were charged, 2 Glu, 1 Asp, and 3 Lys; the seventh was Cys 214 which forms the –S–S– bond to the H-chain or in some cases to another L-chain. The location of these six residues and the two mainly exposed residues His and Thr in the model provides no insight into the basis for their invariance.

The remaining 19 invariant residues, those partly or completely buried and the contacting residues, included 12 strongly hydrophobic residues, 1 Trp, 2 Phe, 4 Pro, 3 Leu, and 2 Val; the remaining residues were Cys 134 and Cys 194, which form the domain S–S bond, two less strongly hydrophobic residues Ala and Thr, and three Ser.

Examining the residues with 4/5 and 3/5 identities, the partly, completely buried, and contacting residues were also overwhelmingly hydrophobic, consisting of 2 Tyr, 5 Val, 1 Thr, and 1 Ala, the others being 1 Gly and 1 His and the gap at position 201. The nonidentical residues at these positions were also hydrophobic in almost all cases, involving replacements of 2 Ile, 2 Leu, and Ala for the five Val residues and of Phe for one Tyr; the remaining Tyr 192 was replaced by Ser in both mouse λ chains, the Thr by Ser or His, the Ala by Asx, and the Gly by Thr. There is obviously an extraordinary preservation of structure in terms of these groups of residues.

Among the 15 and 16 residues that have diverged along κ versus λ lines and which are partly, mainly, and completely buried or are contacting residues, eight are identical pairs; three of these involve conservative hydrophobic substitutions, Ile-Leu, Val-Leu, and Leu-Ile, at positions 117, 132, and 136, respectively, involving mainly and completely buried residues; the remaining pairs are the substitutions Thr-Lys at 172, Gln-Glu 124, Ser-Thr 131, Glu-Ser 165, and Thr-Tyr 178, the last four being contacting residues. The presence in κ chains of an additional residue at 169 causes Thr 172 to be completely buried in the mouse C_κ domain; in the λ chains, Lys 172 is probably completely exposed to solvent. The unpaired residues are 1 Phe, 1 Tyr, 2 Ser, 1 Asp, and 1 Thr in the κ, and 1 Ala, 1 Val, 1 Gln, 1 Glu, 1 Lys, and 3 Thr in the λ group.

The κ versus λ differences show a predominance of completely and mainly exposed residues with 14 of 29 and 20 of 36 residues in κ and λ, respectively, falling into this group; nine are pairs.

The region 101–108 consists of two invariant residues, 101 and 102, followed by six positions in which residues with 4/5 identities alternate with residues which have evolved along κ and λ lines, including the gap at position 108; four of the positions are completely exposed, one is partly exposed, and two are completely buried. Arg 107 marks the end of the mouse V_κ domain; C_κ starts with Ala 109. The additional residue at position 108 in λ chains could be accommodated by a hairpin bend facilitated by Gly 107 or by Pro 109.

There is a cluster of invariant residues from 118 to 123 consisting of three contacting residues, 118, 119, and 121, one partly buried 120, and two exposed residues, 121 Ser (4/5) with an Asp alternative, and an invariant Glu 123. The region 124–140 contains some invariants but is largely made

up of residues which have evolved along κ and λ lines; residues 130–137 consist exclusively of contacting and completely buried residues of which three are invariant, in another three κ and λ differ, and in the others more variation has occurred. At 135 human κ and λ have Leu while mouse κ has Phe and mouse λ Thr and at position 137 human and mouse κ have Asn, human λ Ser, and mouse λ Thr.

The most striking region which has been preserved along κ versus λ lines, 160 to 175, has five contacting, one completely and four partly buried residues, and two mainly and three completely exposed residues. It is followed by a largely invariant cluster, 176 to 181, of contacting and buried residues. The remainder of the molecule is mainly exposed except for the region around Cys 194, in which buried and mainly exposed residues alternate and there is no clustering of invariant or κ versus λ residues.

DISCUSSION

The general structural relationships described for the C_L region of human and mouse κ and λ chains are an initial attempt to understand the basis for the evolutionary preservation of certain regions as essentially invariant and the preservation of others along κ versus λ lines, an evolutionary divergence which took place about 200 million years ago (8). The principles established from sequence and structural findings on other proteins generally apply equally well to the immunoglobulins. Thus the buried residues and those contacting the heavy chain tend to be largely invariant and hydrophobic, while residues that are exposed or mainly exposed may vary, and are generally polar. These also include the few residues that have evolved along human versus mouse lines. There are, however, substantial numbers of hydrophobic residues that are invariant or identical in 4/5 or 3/5 chains and that occur in regions of the molecule for which no obvious structural basis for their conservation may be assigned.

The marked clustering of residues which have been preserved along κ versus λ lines in certain portions of the chain, most notably at positions 160 to 175, or have been maintained invariant, such as 118 to 123, suggests that these may have unique functions.

The C_L domain has been extraordinarily preserved once κ versus λ diversification occurred, since at only eight positions were all four chains different and at only five other positions were human and mouse chains identical despite a κ versus λ difference (Fig. 1).

The immunoglobulin findings were compared with values for human and mouse hemoglobins, using the α chain as equivalent to κ and the β (γ) chain equivalent to λ (8). Since hemoglobin chains are longer, the data were normalized to 115 residues. The findings for hemoglobin are strikingly different from the immunoglobulin results. Thirty-six residues were invariant, and 4/5 and 3/5 chains were identical at 12 and at 49 positions, respectively. Human and mouse α and human and mouse β were identical at 13 and 11 positions, respectively. At only one position each were mouse α and mouse β, human α and mouse β, and human β and mouse α identical. There were no positions at which all four chains differ. Thus there were 97 residues in which three or more chains were identical in hemoglobins in contrast to 53 such residues in the immunoglobulins. This comparison indicates a higher degree of evolutionary conservation of residues that did not differentiate along α versus β lines than residues that differentiated along κ versus λ lines. The immunoglobulin C_L domain thus shows a higher degree of evolutionary adaptability than do the hemoglobins.

1. Kabat, E. A. (1968) *Structural Concepts in Immunology and Immunochemistry* (Holt, Rinehart and Winston, New York), 305 pp.
2. Gally, J. A. (1973) in *The Antigens*, ed. Sela, M. (Academic Press, New York), pp. 161–298.
3. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 3440–3444.
4. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M. & Westholm, F. A. (1974) in *Progress in Immunology II, Vol. 1: Immunochemical Aspects*, eds. Brent, L. & Holborow, J. (North-Holland Publishing Co., Amsterdam), pp. 103–113.
5. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 4298–4302.
6. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13**, 669–678.
7. Svasti, J. & Milstein, C. (1972) *Biochem. J.* **128**, 427–444.
8. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C.).