preferred orientation of crystallites as in a polycrystalline aggregate. In some photographs, these two types coexist (Fig. 8). This may be indicative of the method of development of the echinoderm skeletal element—a process of oriented polycrystalline growth on, or in, an organic matrix followed by maturation which appears to involve recrystallization by continual fusion and coalescence.

Such an hypothesis allows for the absence of crystal faces within the mature stereom. A recrystallization in the solid state produces a pseudomorph, the shape of which is determined by the polycrystalline aggregate and the original organic framework. Most of the skeletal stereom occurs in the mature state, and, except for the tubercles, only the actively growing surfaces are polycrystalline. Thus, the echinoderm skeletal elements are neither single crystals nor polycrystalline aggregates, but they are a combination of both. These data help to reconcile the contradictions in some of the studies of earlier workers.

KENNETH M. TOWE
Department of Paleobiology,
Smithsonian Institution,
Washington, D.C. 20560

References and Notes

1. D. M. Raup, in *Physiology of Echinodermata*, R. A. Boolootian, Ed. (Interscience, New York, 1966), p. 379.
2. C. D. West, *J. Paleontol.* **11**, 458 (1937).
3. G. Donnay, *Carnegie Inst. Yearbook* **55**, 205 (1956).
4. J. Garrido and J. Blanco, *Compt. Rend.* **224**, 485 (1947).
5. H. Nissen, *Neues Jahrb. Mineral. Abhand.* **117**, 230 (1963).
6. J. D. Currey and D. Nichols, *Nature* **214**, 81 (1967).
7. G. Donnay, J. Wyatt Durham, P. M. Kier and D. M. Raup provided helpful suggestions and criticism. I thank G. Hamilton, J. Merida and T. Phelan for assistance.

7 June 1967

Immunoglobulin Structure: Variation in Amino Acid Sequence and Length of Human Lambda Light Chains

Abstract. *Variation and conservation in the primary structure of human lambda light chains is revealed by complete amino acid sequence of three Bence Jones proteins. These proteins differ in amino acid sequence in from 38 to 48 positions; they are of unequal length in the amino-terminal half of the chain but have identical sequence in the last 105 amino acids.*

Development of the concept that the light chains of immunoglobulins have a variable NH_2-terminal half and an essentially invariant COOH-terminal half began in 1963 with the suggestion of Putnam *et al.* (*1*) based on peptide maps that "All Bence Jones proteins of the same antigenic type share a fixed portion of their sequence and also have a mutable part." The concept was explicitly proposed on the basis of partial analysis of the amino acid sequence of κ-type Bence Jones proteins (*2*) and has received strong support as a result of complete amino acid sequence analysis of one κ-type (*3*) and one λ-type (*4*, *5*) Bence Jones protein of man, and of extensive sequence analysis of two κ-type Bence Jones proteins from the mouse (*6*). We have now verified this concept for human λ-chains by amino acid sequence analysis of two more λ-type Bence Jones proteins reported herein. Further support is offered by the confirmation (*7*) of the sequence we have reported for the COOH-terminal half of human λ-chains (*5*).

We now report the complete amino acid sequence of two human λ-type Bence Jones proteins (designated Ha and Bo, respectively) in comparison to the complete sequence we have published for the human λ-type protein designated Sh (*5*). In Fig. 1, a different numbering system is employed for the Ha and Bo λ-chains since their length is greater than that of the Sh λ-chain. This is given underneath the sequence for protein Ha. However, all citations in the following text refer to the Sh numbering system given over protein Sh in Fig. 1. [This practice is analogous to that introduced by Braunitzer *et al.* (*8*) for sequence comparison of the α-chains and β-chains of hemoglobin, which also are of unequal length.] With this alignment the amino acid sequence of all human λ-chains is presumably identical, from Gln-109 through the COOH-terminal residue Ser-213 (*9*) with the exception that in the Oz(+) serological subtype the arginine at position 190 is replaced by lysine (*10*). Hence, the amino acid sequence is given in Fig. 1 only for the NH_2-terminal portions of proteins Ha and Bo through the residue denoted Gly-108 in the Sh numbering system.

The conclusion that the sequence of the last 105 residues in human light λ-chains is essentially invariant is based on the following findings. By methods already described (*4*, *5*), we have isolated from proteins Ha and Bo tryptic peptides corresponding to all those in protein Sh, beginning with Ala-112 through the COOH-terminal residue Ser-213. Without exception, all the corresponding tryptic peptides for this region of the three proteins [namely peptides T_{16}, T_2, T_{18}, T_{12}, T_1, T_{20}, T_7, T_{11}, T_{19}, and T_{14} of Wikler *et al.*, (*5*)] are identical in amino acid composition, in NH_2-terminal and COOH-terminal groups (in those cases determined), and in such portions of their sequence as we have already completed. These peptides from proteins Ha and Bo also have the same elution position in ion-exchange chromatography and the same electrophoretic mobility at *p*H 3.7 as the corresponding peptides isolated from the Sh protein. Insofar as the data have been obtained, the same statements apply to the chymotryptic peptides we have isolated from this portion of the three λ-chains. Furthermore, the published sequence of our λ-chain Sh is identical in every detail (*5*), except for the position of one amide group, with a sequence from Gln-109 through the COOH-terminus Ser-213 for another λ-type Bence Jones protein (X) (*7*). Thus, residues 109 to 213 are identical in all four λ-type proteins (Sh, Bo, Ha, and X). On the other hand, the position corresponding to Gly-108 in protein Sh is occupied by arginine in proteins Ha and Bo (Fig. 1) and by serine in protein X.

Like κ-chains (*3*), human λ-chains differ in many positions in the NH_2-terminal portion of the molecule. Exactly half of the 108 positions in the variable part of the lambda-type protein Sh are identical with corresponding positions in the lambda-type proteins Bo and Ha (Table 1). However, the remaining 54 positions in this portion of the λ-chain Sh differ from one or both of the other λ-chains (Fig. 1). The positions of variation are distributed in a seemingly random fashion throughout the first half of the λ-chains; yet, there are two short peptide stretches where the amino acids differ in all three λ-chains (residues 48 to 51 and 91 to 95). Conversely, up to residue 108 there are four tetra-

peptide sequences that are identical in all three λ-type proteins, and likewise one pentapeptide and even one heptapeptide sequence (residues 57 to 63). Several of the latter represent areas of strong homology with the κ-chains of man and mouse; the most striking is the sequence Arg-Phe-Ser-Gly-Ser (residues 59 to 63) which has been found in all κ- and λ-chains for which sequence has been reported. This evidence for conservation of certain areas of primary structure in light chains accords with our observation from detailed sequence comparison that *positions that are variable in λ-chains are generally variable in the counterpart position in κ-chains; vice versa, positions that are conserved in κ-chains also tend to be conserved in λ-chains.* Thus, among the 54 positions that are identical in the NH_2-terminal portion of the three λ-chains in Fig. 1, 25 or almost half are also identical with the human κ-chain Ag, and thus far have not been shown to be variable in other human κ-chains.

The three human λ-chains compared in Fig. 1 differ more from each other in sequence than do the only two human κ-type proteins (Ag and Roy) for which extensive sequence data are available (*2, 3*); however, the number of differences in the variable portion of these λ-chains [38 to 48 (see Table 1)] is comparable to the number of differences (42) reported for two mouse κ-chains (*6*). The apparently greater homology in primary structure of the human κ-chains probably reflects the insufficiency of the data, for we have evidence that there is a comparable number of variable loci in two human κ-type proteins now under study in our laboratory (*11*).

Because of the difference in length of the λ-chains, the number of positions that can be compared in direct alignment in Fig. 1 varies from 107 to 111. For the three possible pairings (Sh and Ha, Sh and Bo, Ha and Bo) the number of positions of *identity* in the variable region of these three λ-chains ranges from 60 to 73. This is substantially greater than the number of positions of identity that λ-chains share with the corresponding regions of either human or mouse κ-chains. It is not surprising that the three human λ-chains resemble each other in primary structure more than they do the κ-chains of man or mouse. However, by the same criterion, that is, on the basis of the number of identities in primary sequence in the variable region, these three λ-chains are as closely related to the two mouse κ-chains (45 to 50 positions of identity) as they are to the only human κ-type protein (Ag) for which the sequence has been published (42 to 46 positions of identity). This further substantiates our previous conclusion that there was an early evolutionary divergence of the genes for κ- and λ-chains (*3–5*).

From the limited data available it is difficult to generalize about the nature of the mutations, for example, whether they are predominantly one-step in origin. The only two κ-type proteins extensively studied (Ag and Roy) happen to represent one of the several known subgroups of κ-chains (*3*); whereas the three λ-type proteins of Fig. 1 represent several subgroups of λ-chains. For example, proteins Ha and

Fig. 1. Comparison of the amino acid sequence of the variable portion of three human λ-type Bence Jones proteins (Sh, Bo, Ha). Sequences joined by hyphens were determined by analysis. Areas of undetermined sequence are enclosed in parentheses. Where undetermined, acid or amide groups are indicated by Asx or Glx. A different numbering system is employed for protein Sh (residues 1 to 108 above the sequence) and for protein Ha (1 to 112 below the sequence). The sequences of proteins Bo and Ha are aligned to achieve maximum homology with protein Sh. Thus, four amino acids are inserted in protein Bo and Ha: the NH_2-terminal residue PCA and three residues after Arg-27 of protein Sh. Residues which are nonidentical at the same position in the three chains are designated by underscored letters.

Bo though differing greatly are more like each other than like Sh in such properties as chain length, the presence of pyrrolidone carboxylic acid (PCA), the insertions after Arg-27, and even in the data on homology given in Table 1. The existence of such subgroups may explain why it appears that there are more amino acid interchanges in human λ-chains than in κ-chains that are incompatible with one-step mutations. In the three pairings of these λ-type proteins the minimum number of two-step mutations required ranges from 10 to 16. When all three λ-chains are compared, the number rises to 24 (Table 1). Of course, in some cases where a two-step mutation would be required to explain an interchange at a particular position in one pair of λ-type proteins, only a one-step mutation would be needed to explain the changes in both of these proteins with respect to the third. [See for example, position Ala-32 in the Sh protein which corresponds to a minimum two-step mutation for the change to tyrosine in protein Ha; nevertheless the changes of serine to alanine (protein Bo) and from serine to tyrosine (protein Ha) are both compatible with one-step mutations.] If a two-step mutation results from two successive one-step mutations, the presence of many two-step changes indicates a high frequency of mutation.

Human λ-chains differ not only in amino acid sequence in the variable region, but also in length. Both the Bo and the Ha proteins begin with a blocked NH_2-terminal group (apparently PCA) and also have three additional residues which are placed after Arg-27 in Fig. 1. The evidence for the lack of PCA in protein Sh has been summarized (*5*). The presence of PCA in proteins Bo and Ha is in accord with the report of Hood *et al.* that NH_2-terminal PCA does occur in some human λ-chains (*12*). Figure 2 gives the proof for the sequence of residues 20 through 34 containing the three additional amino acids in this region of protein Ha.

Table 1. Homology of lambda-type Bence Jones proteins. Distribution of probable identities in position and of minimum nucleotide base changes in codons (variable part—112 positions compared).

Positions			Minimum base changes	
No. compared	Identities	Differences	One	Two
	Proteins Sh, Ha, Bo			
107	54	53	29	24
	Proteins Sh, Ha			
108	60	48	32	16
	Proteins Sh, Bo			
107	62	45	31	14
	Proteins Ha, Bo			
111	73	38	28	10

The insertion of three additional amino acids in the sequence of proteins Ha and Bo in the area of Arg-27 of protein Sh has precedent in the finding by Gray, Dreyer, and Hood (*6*) that the mouse κ-type Bence Jones protein M-70 has four additional amino acids after Gln-27 in its sequence compared to mouse κ-type protein M-41. As pointed out by these authors, this variability in chain length is incompatible with some proposed theories of light-chain variation, for example, the proposal of a translational mechanism involving different readings of a single messenger RNA (*13*). Both in human λ-chains and in mouse κ-chains, the difference in length occurs in a region between two positions that have been invariant in all light chains studied, for example, the first half-cystine (Cys-23 in human and mouse κ-chains and Cys-21 in the λ-chain protein Sh) and the first tryptophan (Trp-35 in human and mouse κ-chains and Trp-33 in the λ-chain protein Sh). In Sh the three gaps are arbitrarily placed consecutively. However, if one gap were to be placed in proteins Ha and Bo and an additional one in Sh, there would appear to be greater homology in this area of sequence; both human λ- and mouse κ-chains would then have up to four gaps, each signifying a codon deletion. There may be other places where deletions in λ- (or κ-) chains occur in addition to this region of variable length and the omission of the initial PCA residue. For example, a definite amino acid cannot yet be assigned to position 100 after Phe-99 in protein Bo. Although this omission may only reflect technical problems of sequence analysis, it is interesting that, to get the best alignment of κ- and λ-chains, two gaps had to be placed at this point in κ-chains (*4*). Similarly, in the mouse κ-type protein M-41, a definite amino acid could not be assigned to position 87 (*6*).

For the first time, there is available for light chains a statistically significant number of interchanges to test whether these occur randomly or are restricted to certain exchanges. In the three pairings of the variable portions of these λ-type proteins there are 64 different pairs of exchanged amino acids such as serine and threonine or serine and alanine; this is about one-fifth of the theoretically possible number of pairings of 19 different amino acids. Of the 20 natural amino acids all are involved in the interchange except half-cystine, which is present in an apparently immutable disulfide bridge, and methionine which is absent in these λ-type proteins. With the exception of half-cystine, proline, and possibly glycine, the number of times a given amino acid is involved in the interchanges is closely proportional to its abundance in the variable part of the chain. For example, in the variable parts of the three proteins

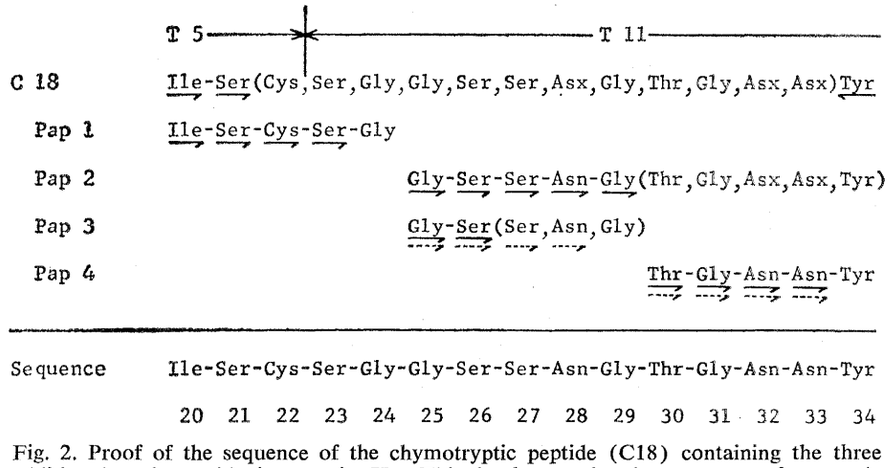

Fig. 2. Proof of the sequence of the chymotryptic peptide (C18) containing the three additional amino acids in protein Ha. This is the overlapping sequence for tryptic peptides T5 and T11, part of which are indicated by the symbols above peptide C18. Papain peptides (Pap) were obtained from C18 as indicated, and their sequence was determined by Edman degradation (———→) and by leucine aminopeptidase (·····→) and carboxypeptidase A (←———) digestion.

there are altogether 22 alanines, 24 threonines, and 24 valines, and these amino acids participate in interchanges in the three pairings 24, 23, and 21 times, respectively (each position in the sequence of one protein is compared twice, once for each of the other proteins). The lower frequency of cystine, proline, and glycine in the interchanges may be related to the importance of these amino acids in the maintenance of the conformation of the polypeptide chain and the significance for the specificity of light chains for antibody activity.

The random character of the interchanges in the variable regions of λ-chains is emphasized by the fact that, for each pair of proteins, the ratio of minimum one-step to two-step mutations is about two to one (Table 1). This is the same base ratio as that required to satisfy the relationship of amino acid changes in the α- and β-chains of human hemoglobin (*14*). Of course, many interchanges that are compatible with one-base changes in the corresponding codons are equally explicable by two-base changes. Hence, except for the principle of parsimony, the results in Table 1 do not favor a predominance of one-step changes.

The sequence variation described herein for λ-chains has the following characteristics: (i) restriction of the interchanges to approximately the first half of the chain with initiation of a constant sequence at Gln-109, (ii) an irregular distribution of variable loci in the NH_2-terminal region, yet a conservation of half the positions including the intrachain disulfide bridge, (iii) the random nature of the amino acid interchanges, and (iv) variable length of the chain. Except for restriction of the variable loci to the NH_2-terminal region, all these characteristics are also found on phylogenetic comparison of the primary structure of proteins such as hemoglobin and cytochrome *c* (*14*) and are likewise observed on comparison of polymorphic proteins within the same species, such as the α-, β-, γ-, and δ-chains of human hemoglobin (*8*) or the β-chains of sheep hemoglobin (*15*). The normal polymorphic proteins, however, are limited in number; whereas no two pathological light chains (Bence Jones proteins) have yet been reported to be identical in sequence, and the number of normal light chains may exceed one thousand. The mechanism for generation of both polymorphic and phylogenetic differences in primary structure of proteins is generally accepted to be mutational changes in the genes of the germ line, followed by evolutionary selection. In all respects, the constant regions of κ- and λ-chains must have diverged by a similar mechanism. If the variable regions of light chains likewise did so, there must be many genes for κ-chains and also for λ-chains. If not, some new mutational mechanism must be found which is somatic in origin and which can account for restriction of variation to one end of the chain and thus of localization of mutation (or codon misreading) to one end of the gene. It is doubtful that the mechanism is as simple as codon misreading because of the large number of observed substitutions. Tyrosine, for example, is exchanged for nine different amino acids, namely Lys-29, Asn-29, Asp-30, Ala-32, Ser-32, Phe-47, His-85, Arg-89, and Trp-89. Of these, four represent two-step mutations or would require a double misreading. The same objections apply to somatic mutation mechanisms based on a defective gene repair enzyme that acts after a stretch of DNA has been removed from one sister strand (*16*). Furthermore, such a mechanism should set up a gradient so that the sequence errors should increase in going from the NH_2-terminal end of the chain toward the starting signal, which is assumed to be in the middle of the gene. This clearly is not the case in λ-chains, for the number of variable loci (underscored positions) in the first quarter of the sequence in Fig. 1 is about the same as in the last quarter and nearly twice as great as in the third quarter.

The complete sequence data presented herein for three λ-chains offer an excellent opportunity for testing the several hypotheses on the origin of antibody diversity which have been devised to explain the rather limited sequence data hitherto available on κ-type light chains. The data on λ-chains likewise offer strong support to the hypothesis that light and heavy chains evolved from a common ancestral gene (*4, 17*). Not only are human λ-chains as closely related to mouse κ-chains as to human κ-chains, but also there is about as much homology between the Fc fragment of rabbit heavy chain and human λ-chains as there is between the Fc fragment and human κ-chains. Indeed, an alignment of the four immunoglobulin chains, rabbit Fc, human κ and λ, and mouse κ, can be made in which there are several stretches of 25 consecutive residues of each chain where 80 percent of the amino acids are identical either in all four chains or at least in any two or three of the chains (*18*). These structural relations which are evidence for a common evolutionary origin of the light and heavy chains of immunoglobulins are in accord with our suggestions that κ- and λ-chain specialization preceded interspecies differentiation and that the sequence variation in light chains results from an accumulation of mutations through many separate genes for κ- and λ-chains rather than through some novel mechanism of somatic hypermutation (*14*).

FRANK W. PUTNAM
TOMOTAKA SHINODA
KOITI TITANI, MAURICE WIKLER
Division of Biological Sciences,
Indiana University, Bloomington

References and Notes

1. F. W. Putnam, C. W. Easley, J. W. Helling, *Biochim. Biophys. Acta* **78**, 231 (1963); F. W. Putnam and C. W. Easley, *J. Biol. Chem.* **240**, 1626 (1965).
2. K. Titani, E. Whitley, Jr., L. Avogardo, F. W. Putnam, *Science* **149**, 1090 (1965); N. Hilschmann and L. C. Craig, *Proc. Nat. Acad. Sci. U.S.* **53**, 1403 (1965).
3. F. W. Putnam, K. Titani, E. Whitley, Jr., *Proc. Roy. Soc. Ser. B.* **166**, 124 (1966); K. Titani, E. Whitley, Jr., F. W. Putnam, *Science* **152**, 1513 (1966).
4. K. Titani, M. Wikler, F. W. Putnam, *Science* **155**, 828 (1967).
5. M. Wikler, K. Titani, T. Shinoda, F. W. Putnam, *J. Biol. Chem.* **242**, 1668 (1967).
6. W. R. Gray, W. J. Dreyer, L. Hood, *Science* **155**, 465 (1967).
7. C. Milstein, J. B. Clegg, J. M. Jarvis, *Nature* **214**, 270 (1967).
8. G. Braunitzer, K. Hilse, V. Rudloff, N. Hilschmann, *Advances Protein Chem.* **19**, 1 (1964).
9. Abbreviations for amino acid residues: Lys, lysine; His, histidine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Asx, aspartic acid or asparagine, identity not established; Thr, threonine; Ser, serine; Glu, glutamic acid; Gln, glutamine; Glx, glutamic acid or glutamine, identity not established; Pro, proline; Gly, glycine; Ala, alanine; Val, valine; Ile, isoleucine; Leu, leucine; Tyr, tyrosine; Phe, phenylalanine; Trp, tryptophan; CyS, half-cystine; and PCA, pyrrolidone carboxylic acid. The first amino acid of protein Sh is designated Ser-1, and so forth.
10. E. Appella and D. Ein, *Proc. Nat. Acad. Sci. U.S.* **57**, 1449 (1967).
11. E. J. Whitley, Jr., personal communication.
12. L. Hood, W. R. Gray, W. J. Dreyer, *J. Mol. Biol.* **22**, 179 (1966).
13. M. Potter, E. Appella, S. Geisser, *ibid.* **14**, 361 (1965).
14. T. H. Jukes, *Molecules and Evolution* (Columbia Univ. Press, New York, 1966), p. 178.
15. S. H. Boyer, P. Hathaway, F. Pascasio, C. Orton, J. Bordley, M. A. Naughton, *Science* **153**, 1539 (1966).
16. S. Brenner and C. Milstein, *Nature* **211**, 242 (1966).
17. R. L. Hill, R. Delaney, R. E. Fellows, H. E. Lebovitz, *Proc. Nat. Acad. Sci. U.S.* **56**, 1762 (1966).
18. F. W. Putnam, K. Titani, M. Wikler, T. Shinoda, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
19. We thank Ofra Leon, Mary Barnhart, and Rebecca Prior for amino acid analysis and technical assistance. Supported by grant CA-08497 from NIH.

1 June 1967; revised 12 July 1967