THE BASIC SEQUENCES OF IMMUNOGLOBULIN κ CHAINS: SEQUENCE STUDIES OF BENCE JONES PROTEINS Rad, Fr 4 AND B6

C. MILSTEIN

MRC Laboratory of Molecular Biology, Cambridge, England

Received 17 February 1969

In previous reports [1,2] I proposed that the available sequence information on the variability of the κ light chains of immunoglobulins could be considered as resulting from a departure from three basic sequences. These three basic sequences were defined by the recurrence of certain residues along the individual chains that were invariably, or almost invariably, linked. However, those results were based on data involving residues 1–27 and 59–107, residue 107 being considered as the end of the variable N-terminal half of κ chains. The reason for not including residues 27–59 previously was mainly lack of sequence information on proteins of one of the basic sequences. I have now collected considerable sequence data on three Bence Jones proteins of the type κIII which further support the idea of the existence of three well defined families of κ chains.

The results presented in fig. 1 have been obtained by isolating the tryptic peptides of fully reduced and carboxymethylated proteins by paper electrophoretic and chromatographic techniques as outlined previously [3]. The isolated tryptic peptides were analysed and their sequence established either by the "dansyl-Edman" procedure [4] or by further digestion of the isolated peptides with other proteolytic enzymes, followed by isolation and characterization of the products. The overlaps of the tryptic peptides were not done in individual proteins except in very few instances. However, the similarity of the three proteins permitted me to establish some important overlaps when differences in basic residues occurred. The arrangement of the remaining tryptic peptides was done by comparison with the other basic sequences, which are shown in fig. 2. The sequence of certain peptides (enclosed in brackets in fig. 1) was not completed, although the most probable sequence can be easily derived by comparison with homologous peptides.

Interesting, from the point of view of trypsin specificity, was the unequivocal observation of the splitting of an Arg-Pro bond (residues 39–40 in fig. 1) by trypsin in protein Fr 4. The isatin positive peptide TB5 was detected in good yields after 5 hr digestion with Worthington crystalline trypsin (enzyme:substrate ratio 1:100). An Arg-Pro bond was also observed to be split by trypsin in the Bence Jones protein X [3] at a position clearly homologous to the one reported here.

In order to define the basic sequences as shown in fig. 2, the available proteins and fragments (as explained in the legend to fig. 2) have been classified in three groups by the occurrence of linked groups of residues [1,2]. Assignment of a residue to a given position in the basic sequence was based on the finding of that residue in at least two out of three proteins studied. Each individual protein (not shown) thus presents a number of variants of one basic sequence scattered along the chain in a fairly random fashion, although the occurrence of "hot spots" and highly conservative residues [2] is obvious. Thus the repeated occurrence of a given residue was necessary for its provisional assignment to the basic sequence. Other conventions used for the postulated basic sequences are discussed in the legend to fig. 2.

The most interesting fact emerging from the present studies is that the three proteins shown in fig. 1 require a gap or an addition to be inserted when compared with the proteins of the other two groups. Thus,

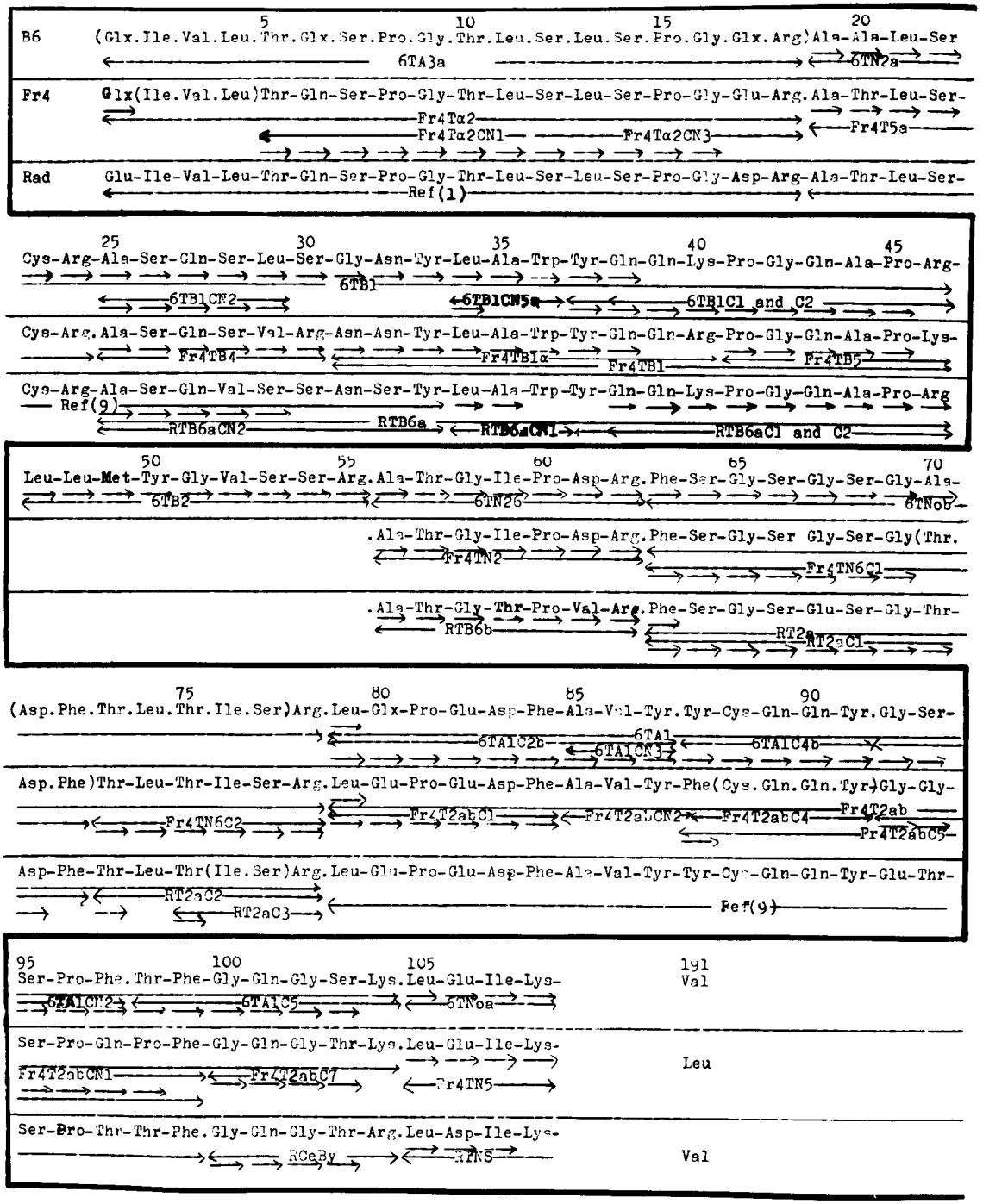

Fig. 1. Partial sequence of the N-terminal half of three kappa Bence Jones proteins. Peptides of the C-terminal half have not been included. Double arrows indicate the isolated peptides and single arrows the results of dansyl-Edman. The points used to separate residues within a bracket indicate that the proposed sequence is based on the similarity with other proteins. Some data are taken from refs. [1,9]. The Bence Jones proteins were kindly provided by Dr. Feinstein (Rad), Dr. Franklin (Fr 4) and Dr. Baglioni (B6).

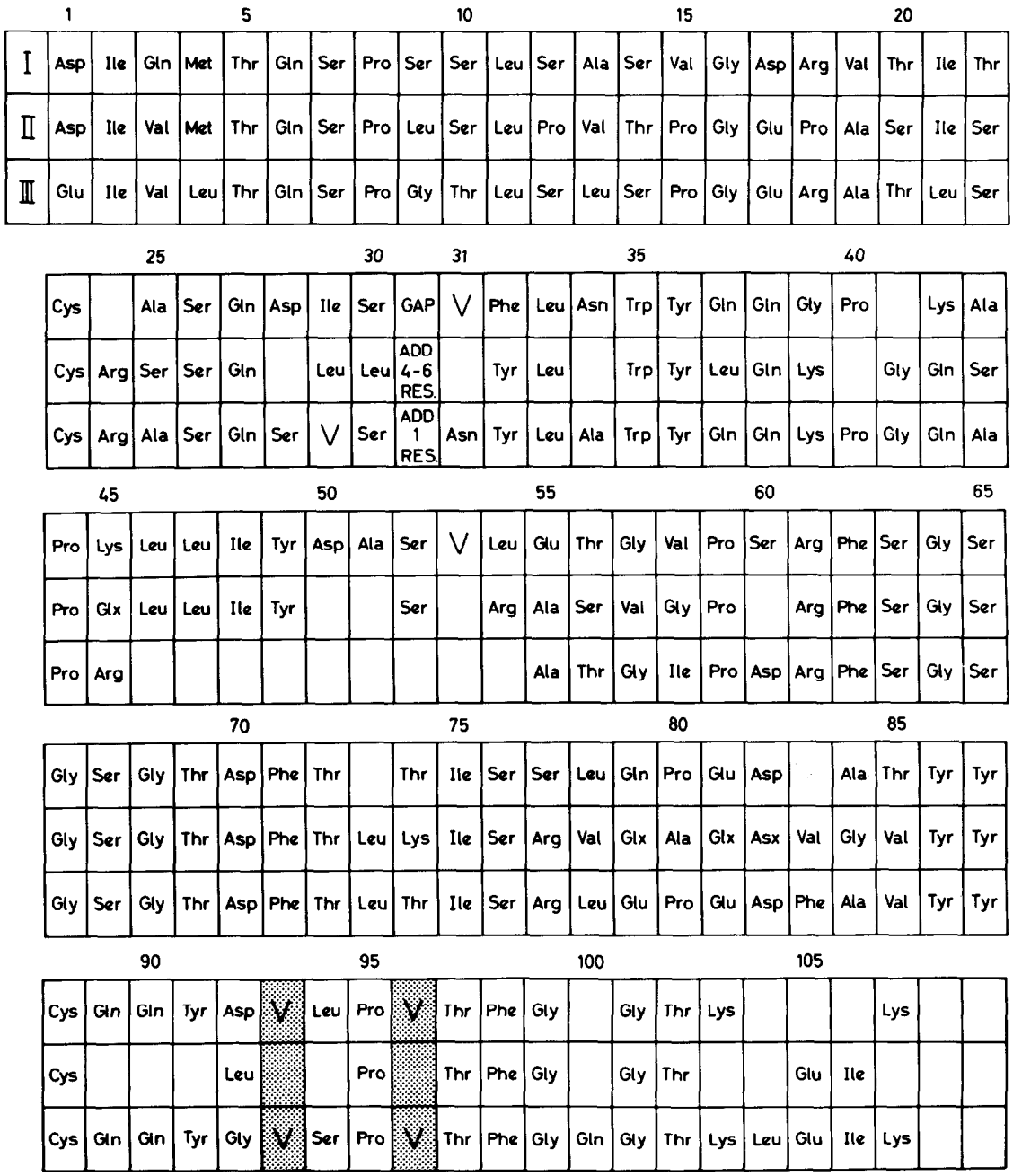

Fig. 2. Three basic sequences in κ chains. Each basic sequence has been constructed by dividing, on similarity grounds, the sequence data of individual κ chains into three groups. Only the residues that occur repeatedly in each group of proteins are indicated in the basic sequences. If more than one residue occurs repeatedly one of them was included only if its frequency was at least 2:1. If more than four variants (or three when the position is established only in three proteins) occur at a given position in a basic sequence this is indicated by a V. Residues 93 and 96 have been shaded to indicate the positions where the largest number of variants has been observed (six variants for basic sequence I). When neither of these criteria was fulfilled the position was left blank. The proteins included in this computation and the corresponding references are: Bel and Man [2], Mil [5], Roy and Cum [6], Car, Ale and Dee [8], Ker and BJ [2,9], Ag [10], Eu [11], B6, Rad and Fr 4 (fig. 1).

the three basic sequences have a size difference with the location of the gap in a homologous place. Basic sequence I is the shortest (107 residues), basic sequence III being one residue longer. The length of two known proteins of basic sequence II varies from four to six residues longer [5,6]. Based on the frequency of the N-terminal sequence [7] of myeloma light chains, it appears that the basic sequence III is one of the two major families, the more common being basic sequence I. This is in agreement with the yield from normal light chains of a tryptic peptide characteristic of basic sequence III [8]. The important question of the number of structural genes which give rise to the three basic sequences is still dubious; there must be at least three (one per basic sequence), but it appears that in fact the minimum number should be increased by at least one and probably more [8]. However, I prefer to keep the number of basic sequences at three. A subdivision of the basic sequences may be the more convenient way to express the need of a larger number of structural genes.

The skilled technical assistance of Mr. J.M.Jarvis is gratefully acknowledged.

References

[1] C.Milstein, Nature 216 (1967) 330.
[2] C.Milstein, Symp. III Proc. 5th FEBS Meeting, Prague (Academic Press, New York, 1968) in press.
[3] C.Milstein, J.B.Clegg and J.M.Jarvis, Biochem. J. 110 (1968) 631.
[4] W.R.Gray, in: Methods in Enzymology, vol. XI, eds. S.P.Colowick and N.O.Kaplan (Academic Press, New York, 1967) p. 469.
[5] W.J.Dreyer, W.R.Gray and L.Hood, Cold Spring Harbor Symp. Quant. Biol. 32 (1967) 353.
[6] N.Hilschmann, Hoppe-Seyler's Z. Physiol. Chem. 348 (1967) 1077.
[7] H.D.Niall and P.Edman, Nature 216 (1967) 262.
[8] C.Milstein, Celia P.Milstein and A.Feinstein, Nature 221 (1969) 151.
[9] C.Milstein, Biochem. J. 101 (1966) 352.
[10] F.W.Putnam, K.Titani and E.Whitley, Proc. Roy. Soc. (London) B (1966) 124, as shown in Atlas of Protein Sequence and Structure (1968) p. 178.
[11] B.A.Cunningham, P.D.Gottlieb, W.H.Konigsberg and G.M.Edelman, Biochemistry 1 (1968) 1983.