A STUDY OF IMMUNOGLOBULIN STRUCTURE

III. An Estimate of the Variability of Human Light Chains*

By R. Quattrocchi, D.Sc., D. Cioli, M.D., and C. Baglioni, M.D.

(*From the International Laboratory of Genetics and Biophysics, Naples, Italy, and the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*)

(Received for publication 24 March 1969)

The peptide chains of immunoglobulins show extensive variability, since all the proteins analyzed so far have been shown to have a unique amino acid sequence (1). Several genetic theories have been put forward to account for this variability (2). These theories can be subdivided into two major groups: germ line theories which postulate the existence of as many genes in the germ line as there are immunoglobulin chains, and somatic variation theories which postulate that some specific mechanism(s) exists to generate variability in amino acid sequences coded by a few genes.

It seems essential to possess an estimate of the number of different sequences which are found in a given species in order to provide the correct explanation for the variability of immunoglobulins. We have attempted to reach this estimate by analyzing a large number of human light chains by peptide mapping. A convenient source of human light chains is found in Bence Jones proteins, secreted by patients with multiple myeloma. These proteins can be easily purified and compared by peptide mapping (3). This method of analysis allows us to establish whether two proteins are different. However, to establish whether two proteins are identical, it is necessary to determine their amino acid sequence.

It has been questioned whether the light chains obtained from Bence Jones proteins are representative of the population of light chains present in antibodies. The recent demonstration that many human and murine myeloma immunoglobulins possess antibody activity (4–7) suggests that the light chains present in these proteins, and thus the corresponding Bence Jones proteins, can be considered constituents of antibodies.

No two identical proteins have been found in the sample of 102 Bence Jones proteins analyzed. The probability of not finding two identical proteins in sampling populations of different sizes has been calculated and has led to an

* Work supported by the Euratom-Consiglio Nazionale delle Ricerche-Comitato Nazionale Energia Nucleare Contract 012-61-12 BIAI and by Grant AI 08116 of the National Institutes of Health, Bethesda, Maryland.

estimate of the minimal number of sequences of light chains which exist in man. This number is rather large, at least a few thousand.

Materials and Methods

Protein Specimens.—The Bence Jones proteins studied were obtained from the sources previously reported (3). In addition, proteins BJ 65 to BJ 70 were donated by Dr. B. Jirgensons, M.D. Anderson Hospital and Tumor Institute, University of Texas, Houston, Tex.; proteins BJ 74 to BJ 79 were donated by Dr. M. Adinolfi, Wright-Fleming Institute of Microbiology, St. Mary's Hospital Medical School, London, England; proteins BJ 83 to BJ 85 by Dr. M. LaVia, Department of Pathology, University of Colorado Medical Center, Denver, Colo.; proteins BJ 86 to BJ 88 by Dr. P. Heller, Department of Medicine, University of Illinois Medical School, Chicago, Ill.; proteins BJ 90 to BJ 96 by Dr. H. G. van Eijk, Laboratorium for Fysiologische Chemie der Rijkuniversiteit, Utrecht, Holland; protein BJ 97 by Dr. A. Sehon, Department of Biochemistry, McGill University, Montreal, Canada; proteins BJ 103 to BJ 111 and protein BJ 123 by Dr. R. L. Engle, Jr., Department of Medicine, Cornell University Medical School, New York; proteins BJ 112 to BJ 122 by Dr. C. E. Buckley, Duke University, Durham, N. C.; proteins BJ 127 to BJ 131 by Dr. C. Milstein, Medical Research Council Unit of Molecular Biology, Cambridge, England; proteins BJ 125 and BJ 137 by Dr. A. Carbonara, Istituto di Genetica Medica, Universita di Torino, Torino, Italy; protein BJ 142 by Dr. D. Ein, Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Md. All the other proteins were prepared in our laboratory from urines collected from patients of the Department of Hematology, Ospedali Riuniti, Napoli; the Istituto di Patologia Medica, Universita di Napoli; and the Istituto dei Tumori, Napoli. The proteins were purified as previously described (8). The proteins were aminoethylated and isolated by gel filtration on Sephadex G-100 as previously indicated (3). Chromatograms of several proteins have been shown previously (3); identical criteria were used in the present investigation for the isolation of the aminoethylated Bence Jones proteins; the fractions containing Bence Jones protein were combined and lyophilized.

Analysis of Proteins.—The lyophilized proteins were digested with trypsin and analyzed by peptide mapping as previously described (8). Peptides were isolated from tryptic digests by a combination of gel filtration on Sephadex G-25 or Bio-Gel P-2 and column chromatography on Dowex-50, according to the procedures previously described (9). Alternatively, the peptides were eluted from peptide maps run in high voltage ionophoresis at pH 4.7 (2 hr at 5 kv) and in ascending chromatography as previously described (10). The peptides were eluted with 0.1 N acetic acid and hydrolyzed in 6 N HCl for 16 hr at 110°C. The hydrolysates were analyzed in a Beckman 120 C amino acid analyzer.

RESULTS

Including the proteins analyzed earlier (3), 102 Bence Jones proteins have now been studied. These proteins were purified by gel filtration on Sephadex G-100 after aminoethylation and were then analyzed by peptide mapping after tryptic digestion. The gel filtration analysis separated in many cases Bence Jones protein "fragments" corresponding to the variable part of the Bence Jones proteins (11).

The primary purpose of the present investigation was to determine the extent of variability in human Bence Jones proteins. For this reason, a simple analytical method, peptide mapping, was chosen over other more precise but more

time-consuming methods of analysis. Visual inspection of the peptide maps of two proteins allows one to establish whether the two proteins are different in amino acid sequence or not. If the two peptide maps are identical, no decision as to whether the amino acid sequences of the two proteins are identical can be taken, but at least isolation and amino acid analysis of the tryptic peptides is required for this purpose.

In practice, at least four peptide maps of each protein analyzed were run. These peptide maps were stained with ninhydrin and then with the reagents specific for either tryptophan, tyrosine, arginine, or reduced sulfur and histidine (8). A composite map containing all the information obtained from the different peptide maps was then drawn; the reactions of specific amino acids were color-coded to facilitate comparison of peptide maps. Each composite peptide map was then compared with the peptide maps of all the other Bence Jones proteins previously analyzed to establish identity or nonidentity.

Whenever peptide maps of two proteins looked similar, the tryptic digests were again analyzed by running the peptide maps side by side. The original method of Ingram (12), a modification of which has been used (13), is particularly suited to compare proteins pairwise, since two peptide maps can be run simultaneously in electrophoresis and chromatography.

102 Bence Jones proteins, 52 of K type and 50 of L type, have been analyzed by peptide mapping in the present study. No two identical proteins have been observed by this method of analysis. This has confirmed previous observations (14–15) that light chains of immunoglobulins show remarkable heterogeneity in the amino acid sequence of the variable region.

Analysis of K Type Bence Jones Proteins.—In analyzing the Bence Jones proteins by peptide mapping, several observations were made which contributed significantly to our knowledge of the variability of these proteins. The tryptic peptides of the invariable C-terminal half were compared in all the peptide maps. No variation of any of the peptides was observed beside that previously reported for peptide K9N (9). The presence of valine in this peptide corresponds to Inv(−*2*) proteins, whereas leucine corresponds to Inv(*2*) proteins.

Of the 52 K type proteins studied, 6 showed the peptide K9N, which contains leucine, characteristic of Inv(*2*) proteins (Table I). The frequency of Inv(*2*) proteins observed (11.5%) corresponds fairly well with that of the Inv(*2*) gene in European populations (16–17) , from which most of our samples were obtained.

Six of the urinary K type proteins studied separated into a Bence Jones protein and a fragment when analyzed by gel filtration on Sephadex G-100 (Table I). The peptide maps of these fragments showed only the tryptic peptides of the variable N-terminal half, which were observed in the peptide maps of the corresponding proteins.

TABLE I

Type K Bence Jones Proteins

Protein	Subtype	Inv(2)	Fragment	Observations
3	II	−	−	
4	I	+	−	*1–18* and *19–23* analyzed*
5	I	−	−	
6	III	−	−	
9	?	−	−	
11	?	−	−	
14	I	−	−	
15	III	−	−	
16	II	−	−	
19	I	−	+	*1–18* and *19–23* analyzed
21	II	−	−	
26	I	−	−	*1–18* and *19–23* analyzed
31	I	−	−	
33	I	−	−	
36	I	−	+	Only *19–23* present
39	II	−	−	
41	I	−	−	
43	I	−	−	Only *19–23* present
45	I	−	−	
48	I	−	+	*1–18* and *19–23* analyzed
50	II	−	−	
53	I	−	−	
54	II	−	−	
56	I	−	−	Only *19–23* present
63	?	−	−	
66	I	−	−	
68	I	−	−	*1–18* and *19–23* analyzed
73	I	−	−	Only *19–23* present
86	I	+	−	
92	I	−	+	
94	I	+	+	
95	I	−	−	Only *19–23* present
96	?			
103	III	−	−	
104	III	−	−	
106	?	−	−	
112	?	−	−	
114	I	+	−	Only *1–18* present
115	I	−	−	Only *1–18* present
116	I	−	−	*1–18* and *19–23* analyzed
118	?	−	−	
119	I	−	−	Only *1–18* present
120	?			
123	I	−	−	
124	III	−	−	
126	I	−	−	Only *1–18* present
130	?	+	−	
131	II	−	−	
136	?	−	−	
137	?	−	−	
139	?	−	+	
140	?	+	−	

* Numbers refer to the position of peptides in the amino acid sequence of light chains: i.e., peptide *1–18* contains the first 18 amino acids from the N-terminus.

One of the proteins from which a fragment was isolated (BJ 94) was an Inv(*1*,*2*) protein. This BJ 94 fragment was used in an attempt to localize the Inv(*1*) serological specificity. Bence Jones proteins are in general Inv(*1*,*2*) or Inv(−*1*,−*2*), but some myeloma immunoglobulins have been reported, which are Inv(*1*,−*2*) (18–19). It has thus been suggested that either the Inv(*1*) or the Inv(*2*) serological specificity may be localized in the variable part of the K type light chains and that it is in some way linked with the other Inv allele. However, the fragment isolated from the Bence Jones protein which was typed as Inv(*1*,*2*) failed to inhibit the Inv(*1*) and Inv(*2*) serological reactions[1]. From these experiments, it was concluded that the Inv serological specificities are probably not localized in the variable region. It cannot be excluded, however, that an Inv determinant is localized in an amino acid sequence which overlaps the variable and invariable region or that this determinant is expressed in the intact light chains only and not in a fragment.

It has been possible (Table I) to assign most K type Bence Jones proteins to each of the three subtypes described by Milstein (20). These K subtypes are defined by a characteristic sequence of the first 23 residues. Aminoethylated proteins of subtype I show in peptide maps two tryptic peptides corresponding to the first 23 residues of the variable half. The first peptide of 18 amino acids, *1–18*, contains methionine in the fourth position and arginine as the COOH-terminal amino acid (20). This peptide has been identified on peptide maps of K type Bence Jones proteins; it is often divided characteristically into two spots because of the oxidation of part of a methionine residue to methionine sulfoxide (3). The second peptide, *19–23*, contains five amino acids and aminoethylcysteine is the COOH-terminal; this peptide also has been identified in peptide maps of Bence Jones proteins (3). Very little variability is observed in these two peptides and only few amino acid interchanges have been described (1).

Proteins of subtype II have proline instead of arginine at position 18, besides having other interchanges, and thus a single peptide of 23 amino acids with COOH-terminal aminoethylcysteine is present in the tryptic digest of these proteins; this peptide has not been identified in peptide maps run at pH 6.4 and it is possible that it precipitates with the insoluble "core" material which is formed at the end of the tryptic digestion (8). In peptide maps of proteins of this subtype, both peptides *1–18* and *19–23* observed in peptide maps of proteins of subtype I are absent.

Proteins of subtype III have arginine at position 18, but no methionine at position 4; peptide *1–18* can thus be easily distinguished from that of proteins of subtype I, since it does not give the reaction typical of reduced sulfur. Peptide *19–23*, moreover, has a different sequence (alanine instead of valine) and a different R_f in chromatography from peptide *19–23* of proteins of subtype I.

[1] Cioli, D., and A. Carbonara. Unpublished material.

By these criteria, most K type proteins analyzed could be assigned to one of the subtypes described. Of 52 proteins analyzed, 27 were of subtype I, 7 of subtype II and 5 of subtype III. 13 proteins were not assigned to any subtype, either because of incomplete information available or because of lack of any precise indications. Most of these proteins do not show any of the peptides characteristic of subtype I or III and were thought in some cases to belong to subtype II.

The relative frequency of proteins of the three subtypes was 69%, 18%, and 13% in the sample analyzed, compared to the frequency of 55%, 10%, and 35% calculated on a smaller sample of proteins which have been partially or totally sequenced (1). We have no explanation for the inversion in the frequency of subtype II and III that we have observed, except for the fact that the calculations are based on small numbers and that our method of subtype assignment is not as reliable as sequencing. It seems, however, clear that subtype I is by far the most frequent. The tryptic peptides *1–18* and *19–23* of six proteins of subtype I have been isolated and analyzed. Peptide *1–18* has amino acid composition identical to that of protein Ag (21) in three proteins (BJ 4, 26, and 48), whereas it shows a methionine-isoleucine interchange in BJ 19; a valine-to-leucine and serine-to-alanine interchange in BJ 68; and a leucine-to-valine interchange in BJ 116. Sequence studies on these and other proteins will be reported in detail elsewhere.[2] Peptide *19–23* of these six proteins has amino acid composition identical to that of protein Ag (21). It can thus be concluded that the results of the analysis of these proteins confirm the assignment to subtype I, based on peptide map analysis.

In five proteins that have been assigned to subtype I (BJ 36, 43, 56, 73, and 95) peptide *1–18* could not be identified on the peptide maps, whereas peptide *19–23* was clearly identified in the expected position. These proteins have thus been tentatively assigned to subtype I, although no explanation has been provided for the absence of peptide *1–18*. Amino acid interchanges present in this peptide may cause a change in its ionophoretic or chromatographic behavior which prevents its identification.

Similarly, in four other proteins, only peptide *1–18* was clearly identified, whereas peptide *19–23* was not (Table I).

Peptide Map Variability and Amino Acid Interchanges.—We have previously attempted to establish the minimal number of amino acid interchanges between Bence Jones proteins by comparing their peptide maps (3). We concluded that there are an average of nine peptide differences between K type proteins and that the minimal number of different peptides observed is six (3). At least one amino acid interchange may correspond to each peptide difference. In examining a larger sample of Bence Jones proteins, we have noticed that some proteins

[2] Alescio-Zonta, L., and C. Baglioni. 1969. Manuscript in preparation.

give very similar peptide maps. Proteins BJ 39 and BJ 50 show for instance several peptides in identical position in their peptide maps (3). Even more striking, however, was the similarity observed for proteins BJ 68 and BJ 116

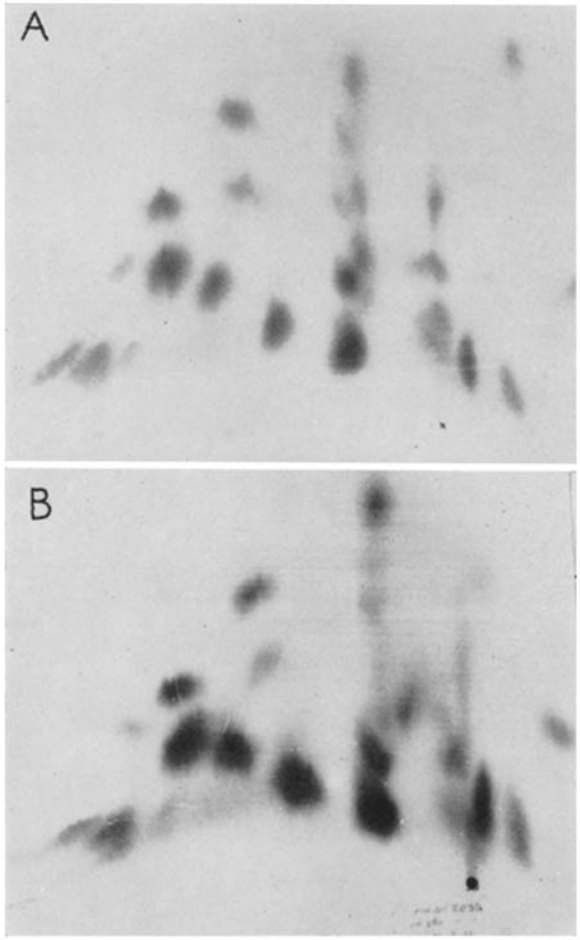

FIG. 1. Peptide maps of protein BJ 68 (A) and BJ 116 (B)

(Fig. 1). Most of the distinctive peptides of these two proteins occupy identical positions in peptide maps, although some minor differences are clearly visible.

It was thus decided to isolate and analyze the tryptic peptides of proteins BJ 68 and BJ 116 to evaluate how similar these two proteins are in amino acid sequence. The tryptic digests of these proteins were separated by high voltage ionophoresis at pH 4.7 followed by chromatography (10). The peptides were

eluted, hydrolyzed, and analyzed. A summary of the amino acid analysis of distinctive peptides is reported in Table II.

The analysis of the tryptic peptides showed that these proteins differ in at least six positions. The differences are caused by amino acid interchanges in peptides which have been localized in the amino acid sequence by homology with the sequence of protein Ag (21) and by the presence or absence of peptides

TABLE II

Analysis of the Tryptic Peptides of Proteins BJ 68 and BJ 116

Peptide	BJ 68	Position in sequence*	BJ 116
2	Asp_2 Ile Met *Ser*$_4$ Glu_2 Thr Pro *Leu*$_2$ *Ala*$_2$ Gly Arg	1–18	Asp_2 Ile Met *Ser*$_5$ Glu_2 Pro *Ala Val*$_3$ Gly Arg
4	Asp Ser Glu_2 Pro_2 Ile Leu Phe Lys	?	Asp Ser Glu_2 Pro_2 Ile Leu Phe Lys
7	Gly Ala_2 Ser Glu_2 Lys	?	Gly Ala_2 Ser Glu_2 Lys
8	Val Ile Glu Lys	104–107	Val Ile Glu Lys
9	Not found	25–41?	Gly_3 Ala Ser Leu_2 Thr_2 Glu_4 Pro_2 Phe Lys
10	Not found	?	Gly_2 Ser Leu Thr Glu_2 Lys
16	Ser Leu Thr Lys	?	Ser Leu Thr Lys
19	Not found	97–103	Val Phe Gly_3 Thr Lys
24	Ser Leu Lys	?	Ser Leu Lys
29	Gly Pro Glu_2 Lys	?	Gly Pro Glu_2 Lys
30	Ala Pro Lys	43–45	Ala Pro Lys
31	Val Thr Ile AECys‡ Arg	19–24	Val Thr Ile Thr AECys‡ Arg
33	Arg	108	Arg

* The position in amino acid sequence is assigned on the basis of the homology with known sequences of Bence Jones proteins (1). The composition of peptides expressed in integral molar residues of amino acids is indicated. Residues italicized are those which have been found to differ between the two proteins.

‡ AECys indicates aminoethylcysteine.

in one protein and not in the other. Moreover, not all the amino acid sequence of the variable half is accounted for by the peptides analyzed. This is easily explained by the precipitation of some of the tryptic peptides at the end of the tryptic digestion (8). It is thus possible that more amino acid interchanges are present in these peptides.

The analysis of these two proteins shows that the variability in amino acid sequence between Bence Jones proteins of the same subtype is not limited to one or very few amino acid residues, but that, in general, many interchanges are present even in those proteins that give very similar peptide maps.

Analysis of L Type Bence Jones Proteins.—50 L type proteins were analyzed by peptide mapping (Table III). In addition, 47 proteins were analyzed by

TABLE III

Type L Bence Jones proteins

Protein	Oz	Fragment	Observations
1	+	−	
7	−	−	
8	−	−	
10	+	−	
12	−	−	
13	−	−	
17	−	−	
18	−	−	
20	+	−	
22	−	−	
25	+	−	
27	−	+	Subtype I*
29	−	−	
30	−	+	
32	−		
35	?	−	Oz peptide not found
37	−	−	
38	−	+	Subtype I*
44	−	−	
46	−	−	
49	−	−	
51	−	−	
52	−	−	
55	−	+	
57	−	−	
61	−	−	
69	+	−	
70	−	−	
74	−	−	
76	−	−	
79	−	+	
87	−	−	
89	Not examined	−	
93	Not examined	−	
97	−	−	
98	−	−	Subtype I*
99	−	−	Serologically Oz(+)
107	−	−	
108	−	−	
110	−	−	
111	+	−	
113	−	−	
117	−	−	
125	−	−	
129	Not examined	−	
132	−	−	
133	−	−	
134	−	−	
135	−	−	
142	+	−	

* Assignment to subtype I (25) is based on analysis of tryptic peptides.[2]

high voltage ionophoresis to identify clearly the Oz peptide (22). 7 of these proteins showed the peptide characteristic of Oz(+) proteins, whereas 39 showed the peptide characteristic of Oz(−) proteins (23). The Oz peptide was not detected in one protein (BJ 35), which was not further investigated. The frequency of Oz(+) proteins determined on the sample of proteins analyzed was found to be 18%. This frequency is considerably lower than that observed by Ein and Fahey (24) on a smaller sample of Bence Jones and myeloma proteins. In order to compare our results with those of Ein and Fahey (24), a sample of several of the proteins analyzed by us was sent to Dr. Daniel Ein for serological typing. For all proteins except one, the serological typing confirmed the chemical typing. The only exception was protein BJ 99, which typed serologically as an Oz(+) protein, whereas by high voltage ionophoresis showed the Oz(−) peptide.

Protein BJ 99 was further analyzed by Dr. Daniel Ein and by ourselves. The Oz peptide was isolated by high voltage ionophoresis and analyzed; it showed the amino acid composition characteristic of the Oz(−) peptide. It was thus concluded that the Oz serological test exceptionally does not type correctly Bence Jones proteins. It was thought possible that this may be caused by an extreme similarity of protein Oz and BJ 99 in the variable half. However, comparison of peptide maps of these two proteins failed to reveal any striking similarity. No explanation could thus be provided for the abnormal typing of protein BJ 99.

The peptide maps of the proteins analyzed were found to differ in several peptides. In no instance did we observe two proteins very similar in their peptide maps, although some variable peptides were present in identical position in several proteins. These similarities have already been pointed out in a previous publication. They are presumably related to the existence of four or five subtypes for L type proteins (25, 26). It has, however, been impossible so far to assign L type proteins to the proposed subtypes on the basis of their peptide maps. The only assignments have been made on the basis of amino acid analysis of tryptic peptides (Table III).

DISCUSSION

Considerable variability has been observed in our sample of Bence Jones proteins analyzed by peptide mapping. In one case in which the peptide maps of two proteins looked almost identical, it was found by analyzing the tryptic peptides that these proteins differed in at least six positions. This observation rules out genetic polymorphism as an important factor in the explanation of variability of amino acid sequence of light chains of immunoglobulins. It has been suggested that genetic polymorphism of human populations may in part account for variability and that the analysis of light chains produced by an inbred strain of mice may show less variability (1). However, genetic polymorphism has

been found to correspond in all cases examined in detail to single amino acid interchanges in peptide chains synthesized under the control of allelic genes.

The variability of immunoglobulin peptide chains can be explained by multiple genes in the germ line or by a somatic mechanism(s) which generates diversity (1). It seems relevant to evaluate thus the different sequences which may be presented in a species. This estimate can be obtained by a mathematical

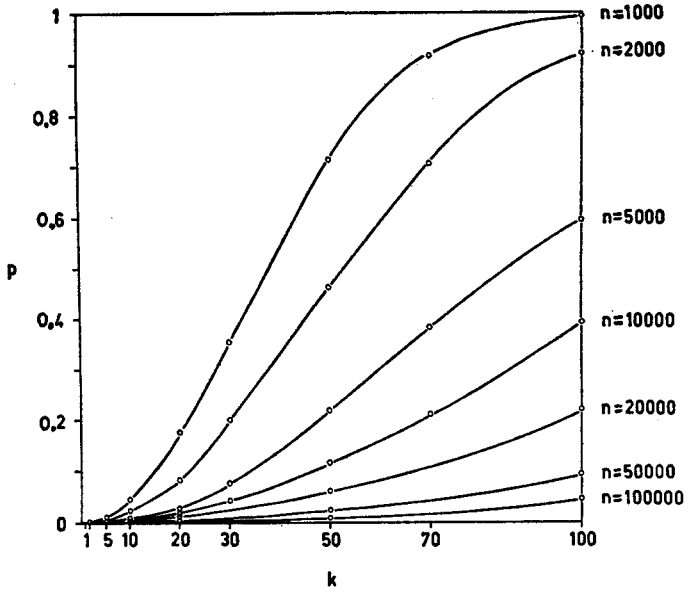

FIG. 2. Probability curves of finding two identical Bence Jones proteins by analyzing a sample of k proteins out of a population of n different proteins. The curves have been calculated for different values of P, k, and n as described in the text.

treatment of the data obtained in the present investigation similar to that previously proposed (2–27). If one assumes that n light chain sequences exist in man (whether the corresponding genes are present in germ line or in somatic cells is not relevant at this point), the probability P of finding two identical light chains by examining a sample of k light chains is given by the following equation:

$$P = \frac{n!/(n - k)!}{n^k}$$

A series of probability curves calculated for different values of n and k are shown in Fig. 2. From these calculations, it seems clear that if the number of different light chain sequences were of the order of a few hundred, we should have found two identical light chains in our survey with a good probability. Since

we have failed to find two identical light chains, we can only set a lower estimate of the probable number n of different light chains that exist in man. We favor a rather large value of n, at least a few thousand. The estimate of n may eventually increase when more sequences are analyzed or more proteins are compared by peptide mapping.

The probable existence of a rather large number of light chain sequences is an argument against germ line theories of genetic control of antibody structure. If the relationship, one gene: one polypeptide chain, holds true for the immunoglobulins, several thousand genes may be necessary to code for all the immunoglobulin chains. The presence of these genes with closely related nucleotide sequences and functions in the genome of vertebrates would pose several problems. A sizeable part of the genome would be occupied by the immunoglobulin genes and rather complicated control mechanisms would be necessary to induce two genes (one for the variable region of the light chain and one for the variable region on the heavy chain) of a very large population of similar genes. In addition, loss of genes similar in nucleotide sequence may be caused by unequal crossing over (28), and this may result in marked instability of the immunoglobulin genes.

The existence of subtypes of light chains (20, 25-26) is suggestive on the other side of multiple genes for these proteins. It seems quite evident that at least as many genes exist in the germ line as there are light chains subtypes. The differences in amino acid sequence between proteins of one subtype and those of a different subtype are so many, including differences in chain length, that they cannot be explained by any of the proposed mechanism(s) by which somatic variability has been suggested to be generated (1).

The relative frequency of proteins of different subtypes seems to be quite characteristic. Among K type proteins, those of subtype I are by far the most frequent both in the proteins that we have analyzed and in those that have been sequenced (1). Sequence studies of pooled serum immunoglobulin light chains suggest that K type light chains of subtype I are also predominant among the normal antibody population (29). The higher frequency of proteins of one subtype is somewhat difficult to reconcile with somatic theories of antibody variability, unless one assumes that some basic sequences are used more frequently because they function better in antibodies. Alternatively, in germ line theories the relative frequency of proteins of different subtype can be correlated with the presence in the genome of a widely different number of structural genes for each subtype. A similar explanation may account for the different proportion of heavy chain subclasses observed in human pooled serum immunoglobulins (2).

The possibility that the subtypes of light chains contribute to the immunoglobulin population to a different degree in different species may be used to confute an argument which has been used against the germ line theories (29).

Characteristic interspecies differences in amino acid sequence have been observed when pooled serum light chains have been examined; this has suggested that only a few structural genes for immunoglobulin chains exist in each species. It would seem impossible otherwise that all the immunoglobulin genes would mutate accordingly to give sequences, characterized by the presence of seemingly invariable residues in some positions of the variable half of light chains (29). However, if one considers the possibility that in one species one subtype is predominant (i.e., subtype KI in man) and that a different subtype is predominant in a different species, by sequencing pooled immunoglobulins, one obtains the sequence of the subtype which is predominant in each species. It is then impossible to draw conclusions on the origin of the immunoglobulin genes, since evolutionary divergence of the structural genes for the different subtypes might have preceded the extensive gene duplication postulated by the germ line theories.

The data obtained in this investigation cannot be used to prove or disprove any of the current theories on the genetic control of antibody structure. They indicate only that the primary structure of immunoglobulin light chains is extremely variable and that possibly more subtypes of light chains exist than have so far been recognized.

SUMMARY

102 human Bence Jones proteins have been purified by gel filtration, digested with trypsin, and analyzed by peptide mapping. In several cases Bence Jones "fragments", corresponding to the variable half of the corresponding proteins, were observed. The peptide maps of the proteins were compared to establish whether any identical proteins were present in the sample analyzed. No Bence Jones protein showed a peptide map identical to that of any other protein, although remarkable similarities in the peptide maps were observed for some proteins. Two proteins that gave very similar peptide maps were then examined in detail, by purifying and analyzing the tryptic peptides. It was then found that these two proteins differ in amino acid sequence in at least six positions.

The probability of not finding two identical sequences by examining a sample extracted from populations of light chains of different sizes has been calculated. This has led to an estimate of the minimal size of the population of light chain sequences in humans. The number of light chain sequences appears to be at least a few thousand.

Information on the frequency of Inv and Oz antigenic determinants and on the relative frequency of subtypes of K chains has been obtained. Proteins of KI subtype are found most frequently. The possibility that different subtypes may be predominant in different species is discussed in relation to the evolutionary arguments used in favor of the somatic theories on the origin of variability of immunoglobulin chains.

We are grateful to Mr. G. Sansone and Mr. G. Di Matteo for skillful technical assistance and to Dr. P. F. Periti for his help with the statistical analysis.

BIBLIOGRAPHY

1. Cohn, M. 1968. The molecular biology of expectation. *In* Nucleic Acids in Immunology. O. J. Plescia and W. Braun, editors. Springer-Verlag New York Inc. New York. 671.
2. Cohen, S., and C. Milstein. 1967. Structure and biological properties of immunoglobulins. *Advan. Immunol.* **7**:1.
3. Baglioni, C., and D. Cioli. 1966. A study of immunoglobulin structure. II. The comparison of Bence Jones proteins by peptide mapping. *J. Exp. Med.* **124**:307.
4. Eisen, H. N., J. R. Little, C. K. Osterland, and E. S. Simms. 1967. A myeloma protein with antibody activity. *Cold Spring Harbor Symp. Quant. Biol.* **32**:75.
5. Eisen, H. N., E. S. Simms, and M. Potter. 1968. Mouse myeloma proteins with antihapten antibody activity. The protein produced by plasma cell tumor MOPC-315. *Biochemistry.* **7**:4126.
6. Stone, M. J., and H. Metzger. 1967. The valence of a Waldenstrom macroglobulin antibody and further thoughts on the significance of paraprotein antibodies. *Cold Spring Harbor Symp. Quant. Biol.* **32**:83.
7. Schubert, D., A. Jobe, and M. Cohn. 1968. Mouse myelomas producing precipitating antibody to nucleic acid bases and/or nitrophenyl derivatives. *Nature (London).* **220**:882.
8. Baglioni, C., M. La Via, and V. Ventruto. 1965. A study of immunoglobulin structure. I. Fingerprinting of aminoethylated Bence Jones proteins. *Biochim. Biophys. Acta.* **111**:479.
9. Baglioni, C., L. Alescio-Zonta, D. Cioli, and A. Carbonara. 1966. Allelic antigenic factor Inv(a) of the light chains of human immunoglobulins: chemical basis. *Science (Washington).* **152**:1517.
10. Baglioni, C., D. Cioli, G. Gorini, A. Ruffilli, and L. Alescio-Zonta. 1967. Studies on fragments of light chains of human immunoglobulins: genetic and biochemical implications. *Cold Spring Harbor Symp. Quant. Biol.* **32**:147.
11. Cioli, D., and C. Baglioni. 1966. Origin of structural variation in Bence Jones proteins. *J. Mol. Biol.* **15**:385.
12. Ingram, V. M. 1958. Abnormal human hemoglobins. I. The comparison of normal human and sickle-cell hemoglobins by "fingerprinting". *Biochim. Biophys. Acta.* **28**:539.
13. Baglioni, C. 1961. An improved method for the fingerprinting of human hemoglobin. *Biochim. Biophys. Acta.* **48**:392.
14. Easley, C. W., and F. W. Putnam. 1966. Structural studies of the immunoglobulins. III. Aminoethylated type K Bence-Jones proteins. *J. Biol. Chem.* **241**:3671.
15. Potter, M., W. J. Dreyer, E. L. Kuff, and K. R. McIntire. 1964. Heritable variation in Bence Jones protein structure in an inbred strain of mice. *J. Mol. Biol.* **8**:814.
16. Salzano, F. M., and A. G. Steinberg. 1965. The Gm and Inv groups of Indians from Santa Catarina Brazil. *Amer. J. Hum. Genet.* **17**:273.

17. Lopéz, V., and R. Butler. 1965. The Inv groups in Switzerland. *Vox. Sang.* **10**:314.
18. Ropartz, C., L. Rivat, and P. Y. Rousseau. 1964. Deux nouveaux facteurs dans les systemes hereditaires de gamma-globuline: le Gm(e) et l'Inv(l). *Proc. Ninth Congr. Int. Soc. Blood Transfus.* 455.
19. Litwin, S. D., and H. G. Kunkel. 1967. The relationship between the Inv(1) and (2) genetic antigens of K human light chains. *J. Immunol.* **99**:603.
20. Milstein, C. 1967. Linked groups of residues in immunoglobulin K chains. *Nature (London).* **216**:330.
21. Titani, K., E. Whitley, Jr., and F. W. Putnam. 1966. Immunoglobulin structure: variation in the sequence of Bence Jones proteins. *Science (Washington).* **152**:1513.
22. Appella, E., and D. Ein. 1967. Two types of lambda polypeptide chains in human immunoglobulins based on an amino acid substitution at position 190. *Proc. Nat. Acad. Sci. U.S.A.* **57**:1449.
23. Quattrocchi, R., D. Cioli, and C. Baglioni. 1967. Distribution of an arginine-lysine interchange in the invariable half of human L-type Bence-Jones proteins. *Nature (London).* **216**:56.
24. Ein, D., and J. L. Fahey. 1967. Two types of Lambda polypeptide chains in human immunoglobulins. *Science (Washington).* **156**:947.
25. Von Bernd, L., M. Steimetz-Kayne, and N. Hilschmann. 1968. Die vollständige Aminosäuresequenz des Bence-Jones-Proteins New (-Typ) -Subgruppen in varmiablen Teil bei Immunoglobulin-L-Ketten vom-Typ. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:945.
26. Hood, L., and D. Ein. 1968. Immunoglobulin Lambda chain structure: two genes, one polypeptide chain. *Nature (London).* **220**:764.
27. Cioli, D., and C. Baglioni. 1968. The structure of human immunoglobulins. *Bull. Soc. Chim. Biol.* **50**:949.
28. Baglioni, C. 1962. The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. *Proc. Nat. Acad. Sci. U.S.A.* **48**:1880.
29. Doolittle, R. F. 1966. The amino-terminal amino acid sequences of rabbit immunoglobulin light chains. *Proc. Nat. Acad. Sci. U.S.A.* **55**:1195.