Proceedings of the National Academy of Sciences
Vol. 68, No. 3, pp. 590–594, March 1971

Amino Acid Sequences of Two Mouse Immunoglobulin Lambda Chains

ETTORE APPELLA

Laboratory of Cell Biology, National Cancer Institute, Bethesda, Maryland 20014

Communicated by Herman N. Eisen, December 30, 1970

ABSTRACT The amino acid sequences of the lambda chains of two mouse immunoglobulins have been determined. No variations were found between these proteins. The results are compatible with earlier fingerprinting analysis of the two chains, where no differences in the tryptic peptides could be demonstrated. The lack of sequence variations could be the result either of a peculiarly restricted expression of the lambda variable-region genome in the BALB/c myelomas or a small number of germ-line genes involved in the biosynthesis of light chains of the lambda type.

Studies of the amino acid sequences of kappa and lambda light chains in man and of kappa chains in mice have provided interesting observations about the distribution of sequence variability and have stimulated new thinking about the genetic basis of antibody diversity (1–3). Both types of chains are found to consist of an NH_2-terminal half (about residues 1–107), which is highly variable in amino acid sequence, and a COOH-terminal half (about residues 108–214), which is essentially invariant. No two chains have so far proved to be identical.

A lambda type of myeloma light chain has been described for the BALB/c mouse strain (4). Two such chains produced by plasma cell tumors, MOPC 104E and RPC 20, have been analyzed for some of their chemical and immunological characteristics (5). Tryptic peptide maps were identical both for the arginine-positive peptides and ninhydrin-positive peptides. The amino-terminal peptides were also identical in the sequence of the first ten amino acids (6). These striking results encouraged further studies. This article reports the complete amino acid sequence of one mouse lambda chain (from MOPC 104E) and the essentially complete sequence of another (from RPC 20) of the same type. These sequences permit a more detailed exploration of the nature of the evolutionary relationship between the lambda and kappa chains of man and mouse.

MATERIALS AND METHODS

The light-chain proteins were obtained as described (6). *S*-aminoethyl (AEC-Lc) and *S*-carboxymethylcysteinyl (CMC-Lc) derivatives were prepared as described (5). The peptide fragments were obtained by digestion of various derivatives with trypsin, thermolysin, chymotrypsin, and pepsin. Digestion with TPCK-treated trypsin (EC 3.4.4.4) was in 0.5% NH_4HCO_3, pH 8.0 (enzyme/substrate ratio 1:100, w/w) for 5 hr at 37°C, and was stopped by freeze-drying the solution. The digest was chromatographed on a Dowex 1-X2 column (0.9 × 130 cm) at 40°C using pyridine–acetic acid buffer for development (pH gradient from 8.5 to about 3.0). Each separate peak, or a cut therefrom, was recovered by rotary evaporation and rechromatographed on a Dowex 50-X2 column (0.9 × 27 cm) at 52°C with a linear gradient of pyridine–acetic acid buffer, from 0.2 M at pH 3.1 to 2.0 M at pH 5.0. Further purification, when required, was by chromatography or electrophoresis on paper and by chromatography on Sephadex gel.

Thermolysin digestion was in 0.05 M Tris buffer, pH 8.0–2 mM $CaCl_2$ (enzyme/substrate ratio 1:200, w/w) for 2 hr at 37°C. The peptides were lyophilized, dissolved in 0.05 M ammonia, and applied to a Sephadex G-25 column (0.9 × 150 cm). The fractions obtained were further purified by Dowex 1 or Dowex 50 chromatography as described above.

Chymotryptic (EC 3.4.4.5) digestion was in 1% NH_4HCO_3, pH 8.0 (enzyme/substrate ratio 1:100, w/w) for 6 hr at 37°C. Peptic (EC 3.4.4.1) digestion was in 0.5% formic acid (enzyme/substrate ratio 1:50, w/w) for 5 hr at 37°C. Peptides were separated on Dowex 1 or Dowex 50. The resultant peptide fractions were tested for purity by high-voltage paper electrophoresis and by paper chromatography, if necessary.

Amino acid analyses of peptides were performed on a Beckman/Spinco analyzer, model 644, after hydrolysis with 6 N HCl in an evacuated tube for 20–24 hr at 105°C. Sequential Edman degradation was performed by the subtractive procedure (7), and by direct identification of the amino acid phenylthiohydantoin derivatives by gas chromatography (8). The COOH-terminal sequence was determined by degradation with carboxypeptidases A (EC 3.4.2.1; ref. 9) and B (EC 3.4.2.2; ref. 10), and by hydrazinolysis (11). Complete enzymatic digestion of peptides with leucine aminopeptidase (EC 3.4.1.1), with or without prolidase (EC 3.4.3.7), was a useful adjunct to acid hydrolysis to evaluate the glutamine, asparagine, and tryptophan contents. Large peptides were degraded to smaller fragments by hydrolysis with thermolysin, chymotrypsin, pepsin, and subtilopeptidase A (EC 3.4.4.16). The fragments obtained were purified by column chromatography or paper electrophoresis, and their sequences were determined as above.

RESULTS

Lambda protein MOPC 104E

Complete digestion with trypsin gave the 19 expected peptides, which included a total of 215 amino acid residues. There were two fewer glycines, and one fewer alanine and leucine, in the sequence than in the direct amino acid analysis. This discrepancy might have resulted from impurities in the preparation. Table 1 summarizes the results and gives com-

Abbreviation: TPCK, L-(1-tosylamido-2-phenyl) ethyl chloromethyl ketone.

plete details of the sequence of each tryptic peptide. In the left-hand column, the peptide number has been arbitrarily assigned and does not correspond to the sequence of elution from the column.

The COOH-terminal amino acid in each peptide was lysine, arginine, or aminoethylcysteine, as would be expected from the specificity of trypsin, except in peptide T19, which had serine as the COOH-terminus. The COOH-terminal peptide bond, aminoethylcysteine-serine, was not split by trypsin. The complete sequence of all tryptic peptides was determined, except for T17 where only a partial sequence was obtained. A chymotryptic peptide of 26 amino acid residues (positions 38–63) was isolated from the tryptic digest and, together with the thermolysin peptides, gave a complete sequence for T3. The bonds Trp 37—Val 38 and Tyr 192—Ser 193 were split in about 5% yield, perhaps as a result of chymotrypsin-like activity still present in our trypsin preparation. The Lys 43—His 44 bond was resistant to trypsin. Tryptic digestion produced a core peptide that was not studied in detail.

To establish the sequential order of the tryptic peptides, *S*-carboxymethylcysteinyl light chains were digested with thermolysin. Table 2 lists those peptides that have been isolated with satisfactory purity. Completely overlapping sequences were obtained, and the sequences of various short regions present in tryptic peptides were confirmed by comparison with the thermolysin peptides obtained from the same region. The peptides obtained had terminal amino acids consistent with the specificity of the enzyme. The complete structure of the lambda light chain, MOPC 104E, is shown in Fig. 1. The alignment of peptides is based on the independent evaluation of the amino-terminal and carboxyl-terminal sequences. Peptides containing overlapping sequences from two or more digests have permitted a unique alignment of all known peptides.

Lambda protein RPC 20

In order to demonstrate differences in the amino acid sequence of another lambda chain, protein RPC 20 was studied simultaneously. This study, although not yet completed, has shown no differences. All of the expected tryptic peptides have been isolated and characterized in terms of amino acid composition and partial amino acid sequence; all of these peptides have their counterparts in MOPC 104E. Moreover, evidence has been obtained from thermolysin digests that the order of the tryptic peptides is the same as in MOPC 104E.

DISCUSSION

Nineteen unique tryptic peptides from lambda chains MOPC 104E and RPC 20 of the mouse were isolated and characterized. The two proteins appear to contain about 215 amino acid residues, corresponding to a molecular weight of 23,000. High tryptic specificity was observed only when digestion was limited to 5 hr at 37°C. Under these conditions, arginine, aminoethylcysteine, and lysine bonds were almost completely cleaved, except for the Lys 43—His 44 bond which was not hydrolyzed. It is well known (12) that the rate of tryptic cleavage at arginine or lysine is decreased when the basic residue is preceded in the sequence by an acidic amino

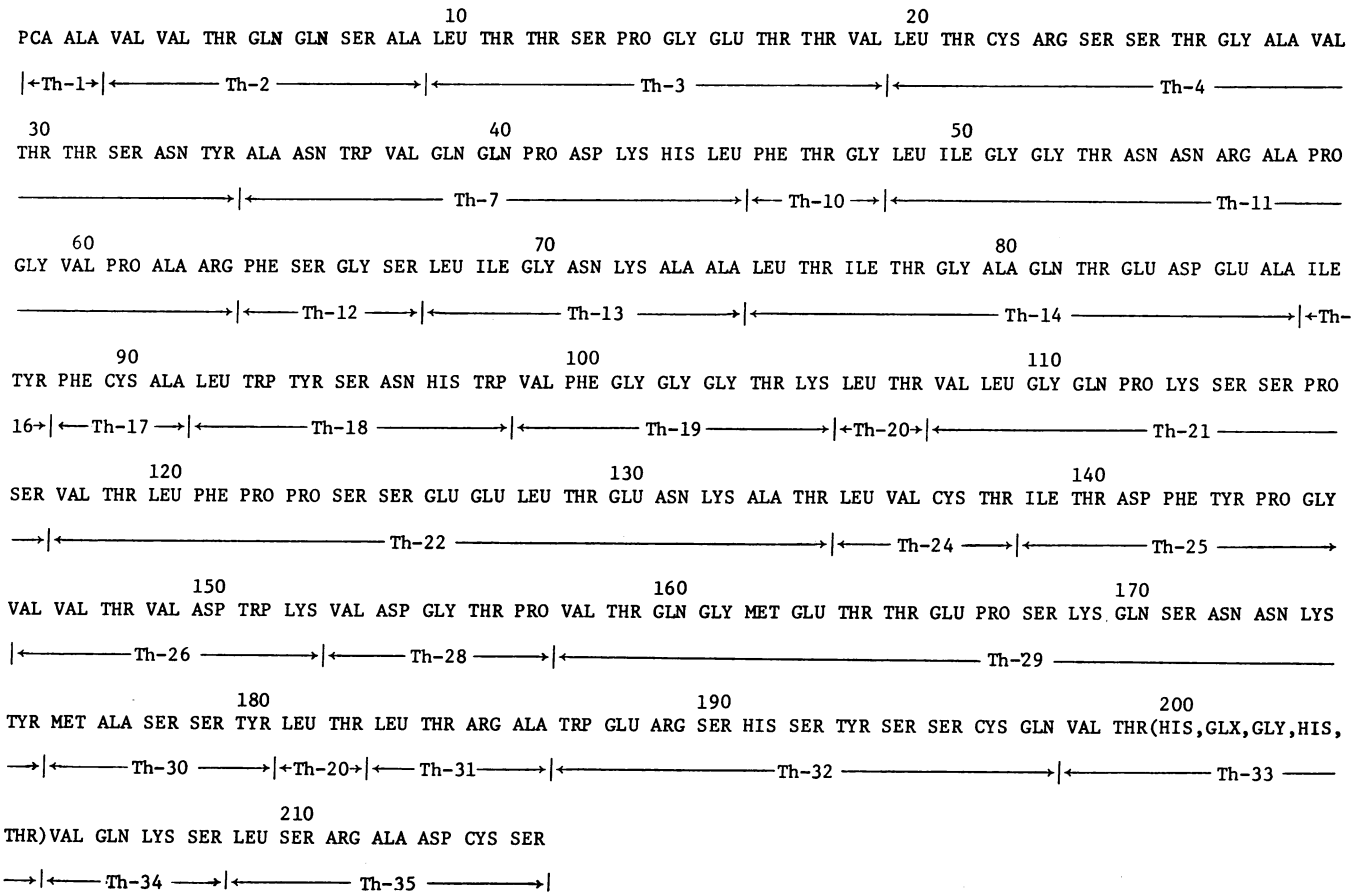

FIG. 1. The amino acid sequence of mouse lambda chain MOPC 104E. The tryptic and thermolysin peptides are arranged in order.

acid residue (e.g., Asp 42). It was not possible, however, to maintain absolute specificity for trypsin even under these limited conditions. The Trp 37—Val 38 and Tyr 192—Ser 193 bonds were highly susceptible to chymotrypsin-like activity still present in the TPCK-treated trypsin used. The complete structure of peptides T14 and T19 has been established only after direct Edman degradation. Peptide T14 contained one more tyrosine and threonine, and peptide T19 one more

Table 1. *Amino acid sequence of tryptic peptides of S-aminoethyl light chain MOPC 104E*

Peptide	Residue no.	Amino acid sequence*	Yield %
T-1	1–22	PCA Ala Val Val Thr Gln Gln Ser Ala Leu Thr Thr Ser Pro Gly Glu Thr Thr Val Leu Thr Cys	21
T-2	1–23	PCA Ala Val Val(Thr,Glx,Glx,Ser,Ala,Leu,Thr,Thr,Ser,Pro,Gly,Glx,Thr,Thr,Val,Leu,Thr,Cys)Arg	16
T-3	24–56	Ser Ser Thr Gly Ala Val Thr Thr Ser Asp Tyr Ala Asn Trp Val Gln Gln Pro Asp Lys His Leu Phe Thr Gly Leu Ile Gly Gly Thr Asn Asn Arg	7
T-4	57–63	Ala Pro Gly Val Pro Ala Arg	33
T-5	64–72	Phe Ser Gly Ser Leu Ile Gly Asn Lys	42
T-6	73–90	Ala Ala Leu Thr Ile Thr Gly Ala Gln Thr Glu Asp Glu Ala Ile Tyr Phe Cys	15
T-7	91–105	Ala Leu Trp Tyr Ser Asp His Trp Val Phe Gly Gly Gly Thr Lys	10
T-8	106–113	Leu Thr Val Leu Gly Gln Pro Lys	8
T-9	114–132	Ser Ser Pro Ser Val Thr Leu Phe Pro Pro Ser Ser Glu Glu Leu Thr Glu Asn Lys	49
T-10	133–137	Ala Thr Leu Val Cys	8
T-11	138–152	Thr Ile Thr Asp Phe Tyr Pro Gly Val Val Thr Val Asp Trp Lys	39
T-12	153–169	Val Asp Gly Thr Pro Val Thr Gln Gly Met Glu Thr Thr Glu Pro Ser Lys	22
T-13	170–174	Gln Ser Asn Asn Lys	5
T-14	175–185	Tyr Met Ala Ser Ser Tyr Leu Thr Leu Thr Arg	11
T-15	186–189	Ala Trp Glu Arg	6
T-16	190–196	Ser His Ser Tyr Ser Ser Cys	7
T-17	197–207	Gln Val Thr(His,Glx,Gly,His,Thr)Val Gln Lys	5
T-18	208–211	Ser Leu Ser Arg	10
T-19	212–215	Ala Asp Cys Ser	14

* Arrows above the sequence represent its establishment by the Edman degradation (→), carboxypeptidase A or B (←), or by hydrazinolysis (←–), whereas arrows below indicate establishment from the smaller peptides derived from larger peptides. T, C, and S represent tryptic, chymotryptic, and subtilopeptidase A peptides. (·) and (†) represent points of cleavage by thermolysin and cyanogen bromide, respectively. These symbols are also used in Table 2.

Table 2. *Amino acid sequence of thermolysin peptides of S-carboxymethylcysteinyl light chain MOPC 104E*

Peptide	Residue no.	Amino acid sequence	Yield (%)
Th-1	1–2	PCA Ala	15
Th-2	3–9	Val Val Thr Glu Glu Ser Ala	20
Th-3	10–19	Leu Leu Thr Ser Pro Gly (Glu, Thr) Thr Val	4
Th-4	20–34	Leu Thr Cys Arg Ser Ser (Thr$_3$,Gly, Ala, Val, Ser, Asx)Tyr T	11
Th-5	20–28	Leu Thr(Cys, Arg, Ser$_2$,Thr, Gly) Ala	3
Th-6	29–34	Val Thr Thr Ser Asn Tyr	5
Th-7	35–45	Ala Asn(Trp, Val, Glx$_2$,Pro, Asp, Lys, His) Leu	3
Th-8	35–37	Ala Asn Trp	10
Th-9	38–45	Val Gln Gln Pro Asp Lys His Leu	5
Th-10	46–48	Phe Thr Gly	25
Th-11	49–63	Leu Ile GlyGlyThr(Asx$_2$,Arg, Ala$_2$,Pro$_2$,Gly, Val) Arg	6
Th-12	64–67	Phe Ser Gly Ser	7
Th-13	68–74	Leu Ile Gly Asn Lys Ala Ala	14
Th-14	75–86	Leu Thr Ile Thr Gly Ala (Glx$_3$, Thr Asx) Ala	8
Th-15	77–85	Ile Thr Gly Ala (Glx$_3$,Thr, Asx)	2
Th-16	87–88	Ile Tyr	15
Th-17	89–91	Phe Cys Ala	3
Th-18	92–98	Leu Trp Tyr Ser (Asn, His) Trp	5
Th-19	99–105	Val Phe Gly Gly Gly Thr Lys	6
Th-20	106–107	Leu Thr	15
Th-21	108–117	Val Leu Gly (Glx, Pro$_2$,Lys, Ser$_3$)	5
Th-22	118–134	Val Thr Leu Phe Pro Pro Ser Ser(Glx$_3$,Leu, Thr$_2$,Asx,Lys,Ala)	4
Th-23	121–134	Phe Pro Pro Ser Ser Glu Glu Leu(Thr$_2$,Glx, Asx, Lys, Ala)	3
Th-24	135–138	Leu Val Cys Thr	7
Th-25	139–145	Ile Thr Asp Phe (Tyr, Pro, Gly)	3
Th-26	146–152	Val Val Thr Val Asp Trp Lys	3
Th-27	147–152	Val(Thr,Val, Asp) Trp Lys	5
Th-28	153–157	Val Asp Gly Thr Pro	10
Th-29	158–175	Val Thr Gln Gly Met (Glx$_3$,Thr$_2$,Pro Ser$_2$,Lys$_2$,Asx$_2$,Tyr)	5
Th-30	176–180	Met Ala Ser Ser Tyr	7
Th-31	183–186	Leu Thr Arg Ala	4
Th-32	187–197	Trp Glu Arg Ser (His, Ser$_3$,Tyr, Cys)Gln	3
Th-33	198–204	Val Thr(His$_2$,Glx, Gly, Thr)	10
Th-34	205–208	Val Gln Lys Ser	6
Th-35	209–215	Leu(Ser Arg Ala Asp Cys) Ser	7

serine, than was determined by amino acid analysis. Isolation of short chymotryptic fragments from this area confirmed the Edman degradation results.

In order to determine the sequential order of the tryptic peptides, *S*-carboxymethylcysteinyl light chains were digested by thermolysin. This enzyme preferentially hydrolyzes the peptide bonds involving amino groups of leucine, isoleucine, valine, phenylalanine, and alanine (13). In this study, we found the same type of specificity, with a slight tendency to hydrolyze the Tyr 175—Met 176 and Ala 186—Trp 187 bonds. The amino acid sequences of mouse lambda (positions 108–215) and kappa and human lambda and kappa chains were compared. There are 75 identities in the common region between mouse lambda and human lambda, but only 37 between mouse kappa and mouse lambda and about 25 between human kappa and mouse lambda. This strongly indicates that the genes for kappa and lambda chains must have differentiated early in evolutionary history.

The presence of at least three fundamental sequences in human kappa chains has been proposed, based on the concurrences of certain amino acids (linked groups) in the variable positions of the chains (14). Four fundamental sequences have been proposed for human lambda chains (15). The amino-terminal portion of 20 kappa chains from the mouse have been examined recently, and these proteins can be divided into at least nine distinct subgroups (16). Studies of mouse lambda chains MOPC 104E, RPC 20, and two others (HOPC 1 and 511, also from the BALB/c strain) in our laboratory do not as yet reveal a pattern of subgroups.

Pairs of human light chains within a subgroup have shown an average of 13 amino acid interchanges (17). Mouse kappa chains within a subgroup similarly differ from one another, with an average of 12 interchanges. A distinct germ-line gene is required for each *V* (variable) region subgroup (17). A controversy, however, exists as to whether the intrasubgroup differences are generated somatically from a small number of germ-line genes or are encoded in several large sets of germline genes evolved from a small number of ancestoral genes by duplication and subsequent divergence (18, 19). It is difficult to decide between these alternatives as insufficient data are available. The identity of two mouse lambda chains could be interpreted as evidence of a small number of germline genes involved in the biosynthesis of light chains, or as resulting from peculiarly restricted expression of the lambda *V* region genome in the BALB/c plasma cell neoplasias. It is known, however, that the lambda chain occurs in low abundance (about 3% of all light chains) in normal mouse immunoglobulins (K. R. McIntire, personal communication). The interesting question of a restricted lambda *V* genome versus a restricted expression thereof will remain after examination of the primary structure of normal mouse lambda chains if the low abundance correlates with a paucity of sequences.

I wish to thank Dr. John K. Inman for reviewing this manuscript and for many helpful comments.

1. Reviewed in *Cold Spring Harbor Symp. Quant. Biol.*, **32** (1967).
2. *Symposium on Developmental Aspects of Antibody Formation and Structure*, eds., J. Sterzl and H. Rhia (Academic Press, New York, 1969).
3. Milstein, C., and J. Pink, *Prog. Biophys. Mol. Biol.*, **21**, 211 (1970).
4. McIntire, K. R., R. M. Asofsky, M. Potter, and E. L. Kuff, *Science*, **150**, 361 (1965).
5. Appella, E., K. R. McIntire, and R. N. Perham, *J. Mol. Biol.*, **27**, 391 (1967).
6. Appella, E., and R. N. Perham, *J. Mol. Biol.*, **33**, 963 (1968).
7. Konigsberg, W., and R. J. Hill, *J. Biol. Chem.*, **238**, 2547 (1962).
8. Pisano, J., and T. Bronzert, *J. Biol. Chem.*, **244**, 5597 (1969).
9. Fraenkel-Conrat, H., J. I. Harris, and A. L. Levy, *Methods Biochem. Anal.*, **2**, 397 (1956).
10. Gladner, J. A., and J. E. Folk, *J. Biol. Chem.*, **231**, 393 (1958).
11. Fraenkel-Conrat, H., and C. M. Tsung, *Methods Enzymol.*, **11**, 151 (1967).
12. Smyth, D. G., *Methods Enzymol.*, **11**, 214 (1968).
13. Matsubara, H., A. Singer, R. Sasaki, and T. H. Jukes, *Biochem. Biophys. Res. Commun.*, **21**, 24 (1965).
14. Milstein, C., *Nature*, **216**, 330 (1967).
15. Langer, B., M. Steinmetz-Kayne, and N. Hilschmann, *Hoppe-Seyler's Z. Physiol. Chem.*, **349**, 945 (1968).
16. Hood, L., M. Potter, and D. McKean, *Science*, in press.
17. Hood, L., and D. Talmage, *Science*, **168**, 325 (1970).
18. Smithies, O., *Science*, **157**, 267 (1967).
19. Dreyer, W. J., W. R. Gray, and L. Hood, *Cold Spring Harbor Symp. Quant. Biol.*, **32**, 353 (1967).