N-Terminal Amino-Acid Sequence of African Lungfish Immunoglobulin Light Chains

(*Protopterus aethiopicus*/5.8 and 19S/ κ and λ chains)

G. W. LITMAN*, A. C. WANG†, H. H. FUDENBERG†, AND R. A. GOOD*

* Department of Pathology, Pediatrics, and Microbiology, University of Minnesota, Minneapolis, Minn. 55455; and

† Department of Microbiology and Department of Medicine, University of California, San Francisco, Calif. 94122

Contributed by Robert A. Good, July 6, 1971

ABSTRACT The sequence of the N-terminal ten amino acids of the unblocked light chains derived from the low molecular weight immunoglobulin of a dipnoid fish, the African lungfish (*Protopterus aethiopicus*), has been determined. A degree of sequence heterogeneity as extensive as that displayed by pooled mammalian light chains was encountered. The major N-terminal sequence of the amino acids of light chains from lungfish immunoglobulins can be made homologous with that of various Elasmobranchian, Chondrosteian, Avian, and Mammalian species if an internal deletion of two base triplets in the gene coding for light chains from lungfish immunoglobulins is assumed. The existence of a deletion and/or insertion mechanism may have profound biological significance, since it would be ideal for the introduction of variability into immunoglobulin chains without sacrificing certain integral features of the antibody molecule.

Much of our current understanding of antibody structure has been derived directly or indirectly from data obtained by amino acid-sequencing of immunoglobulin heavy and light chains (1, 2). Although these studies have markedly influenced our direction and progress toward the understanding of the genetic mechanisms governing immunoglobulin biosynthesis, the exact nature of this mechanism is still obscure. One of the approaches least used has been the amino acid sequencing of isolated immunoglobulin chains from phylogenetically critical species. The phylogenetic approach, although primarily concerned with the evolution of specific structural forms, permits a view of genetic mechanisms that may have become more complex in more distal phases of evolution. To approach the basic problem of evolution and speciation of the immunoglobulin molecule, we have chosen for study the low molecular weight immunoglobulin of the African lungfish, *Protopterus aethiopicus*‡ (3, 4). This species is of special interest for it represents a modern member of an independent evolutionary line; the most primitive deviation available for

the study of a line of evolution leading to primitive amphibians, and later to mammals.

METHODS AND MATERIALS

The low molecular weight immunoglobulin was purified by a combination of gel filtration and ion-exchange chromatography from the sera of African lungfish, *Protopterus aethiopicus*, netted in Lake Victoria, Kampala, Uganda (3). The homogeneity of the purified product was ascertained by immunoelectrophoresis, polyacrylamide gel electrophoresis, and analytical ultracentrifugation. The purified 5.8S immunoglobulin was reduced with dithiothreitol and alkylated with iodoacetamide in the presence of 6 M guanidine·HCl to ensure complete reduction of all inter- and intrachain disulfide bridges. The heavy and light chains were separated by gel filtration chromatography on a 2.0 × 200 cm column of Sephadex G-100 in 6 M guanidine·HCl (pH 5.5). The isolated chains were examined by acid urea-polyacrylamide gel electrophoresis (5) and no cross-contamination was noted. Desalted and lyophilized chain preparations were examined for N-terminal amino acids by reaction with dansyl chloride in 8 M urea, followed by acid hydrolysis and chromatography on polyamide sheets (6). Pyrrolidone carboxylic acid (PCA) was detected in isolated heavy and light chains by digestion of the chains with Subtilisin (1% w/w) at pH 8.0 in 0.1 M NH_4HCO_3 for 17 hr at 37°C. The digests were chromatographed on AG 50W×2(H^+ form) in H_2O and the acidic peptides eluted from the column were separated by high-voltage electrophoresis at pH 3.5. Marker strips were prepared and stained with ninhydrin and then with Clorox-tolidine (7). PCA-containing acidic peptides were identified by both their staining behavior (ninhydrin-negative and Clorox-tolidine positive), and by acid hydrolysis. Amino-acid analysis showed the presence of at least 1 mol of glutamic acid per mol of peptide. Desiccated light chains were subjected to amino-acid sequence analysis by the Edman degradation method (8), with an automatic Beckman model 890 Protein Sequencer (9). After each degradation cycle, a portion (20%) of the recovered residue was converted to a phenylthiohydantoin (PTH) derivative and identified by gas-liquid chromatography (10). The other portion (80%) of the recovered product was hydrolyzed at 150°C for 16 hr in 6 N HCl that contained phenol and was examined by amino-acid analysis. The latter procedure made possible the positive identification of residues and calculation of per cent recovery.

‡ African lungfish possess two antigenically distinct classes of immunoglobulin, 19 S and 5.8 S. Both classes display antibody activity and differ antigenically in their heavy chains. The 19S molecule possesses equal numbers of heavy and light chains, and is most likely a pentamer similar in overall structure to mammalian IgM. The 5.8S immunoglobulin possesses equal numbers of heavy and light chains, and represents the earliest evolutionary deviation from the IgM type molecules. Both of the African lungfish immunoglobulins seem to resemble the two classes of immunoglobulin previously described for the Australian lungfish, *Neoceratodus forsteri* (4).

TABLE 1. Sequence of the N-terminal 10 amino acids of the unblocked light chains derived from the low molecular weight immunoglobulin of the African lung fish *Protopterus aethiopicus*

	1				5				10	
Major Sequence	Asp	Leu	Thr	Glx	Asx	Ala	Ser	Met	?*	Val
Minor components	Glu	Ile	Asx Gly Lys	Gly	—	Gly	Thr Phe Ala	Leu Val	Gly Ala Ser Thr	Gly

* No predominant amino acid can be assigned to this position; glycine, alanine, serine, and threonine are detected in low quantity.

RESULTS

Attempts to detect an unblocked amino terminus in immunoglobulin heavy chains from the lung fish were unsuccessful, with only minute amounts of dansyl aspartic acid being visualized by the dansyl chloride procedure. Light chains, on the other hand, showed a considerable yield of dansyl aspartic acid and lesser amounts of dansyl glutamic acid. Four electrophoretically distinct, ninhydrin-negative and Chloro-tolidine-positive, acidic peptides were detected in the Subtilisin digests of the heavy chains. The R_f s of the PCA-containing peptides were .47, .55, .59, and .64, relative to free PCA. The major peptide, $R_f = .47$, was isolated in relatively pure form and contained Glu, Ser, Val, Leu, and Gly. The other three PCA-containing peptides were isolated in lesser quantities and proved to be crosscontaminated with one another. Inadequate quantities of purified chains precluded further purification and characterization of these three peptides. Lungfish light chains contained considerably less PCA (70%) as the amino-terminal residue. Both electrophoretic and chromatographic analysis of the PCA-containing peptides of the light chains showed them to be identical to the PCA peptides isolated from the heavy chains. The unblocked light chains were subjected to amino-acid sequence analysis and the first 10 residues from the amino terminal are listed in Table 1. The top line presents the major residue, while the bottom line lists the minor components found at each position. The existence of these minor components emphasizes the heterogeneity found in this region of the light chain from lungfish. Table 2 compares the major sequence of lungfish light chains with the sequences obtained for various elasmobranchian, avian and mammalian light chains.

DISCUSSION

The phylogenetic approach to basic biochemical genetics has often revealed simple operational mechanisms obscured through the passage of evolutionary time. The genetic basis of antibody formation is probably one of the most complex biological processes involving a multigenic system and, thus, has been the subject of several phylogenetic studies. Two such studies have involved the determination of the N-terminal amino-acid sequence of heavy and light chains isolated from elasmobranchian (leopard shark) and chondrosteian (paddlefish) immunoglobulins (18, 19). In both studies, homology between the N-terminal amino acids of the heavy and light chains of the individual species was detected. Furthermore, a close sequence homology was identified between the paddlefish, leopard shark, mouse, and human κ light chains. Thus, a common genetic precursor of both heavy and light chain $V\kappa$ regions was suggested. Although chondrosteians and elas-

mobranchians share a common ancestor with mammals, they diverged and evolved independently of the line that includes the Sarcopterygii and which culminated with the emergence of the terrestrial forms and, ultimately, the mammals. To facilitate the understanding of genetic aspects of antibody formation in higher vertebrates, the most primitive available member of the evolutionary line—Sarcopterygii—was chosen for study.

The amino-acid sequence in the light chains of the lungfish immunoglobulin proved to be unique. The data strongly suggest that deletion of two nucleotide-base triplets has occurred, when we compared the sequence with prototype mammalian light chains. A definite homology of the lungfish light chain with the other light chain references is observed if this deletion is taken into consideration. The unblocked sequence cannot, however, be assigned to either the κ or λ type, since equal amounts of homology exist between it and the sequences of mammalian κ and λ light chains. The presence of a blocked amino terminus in the majority of lungfish light chains more closely aligns this fraction of the light chain population with the λ class of mammalian chains, although any phylogenetic comparison based on one residue is extremely hazardous.

Although the unblocked light-chain sequence from immunoglobulins of lungfish cannot readily be classified as either κ or λ , it does provide useful data with which to examine hypotheses concerning genetic mechanisms involved in generation of antibody variability. Some hypotheses propose the existence of multiple ancestral genes that code for immunoglobulin polypeptide chains, which subsequently gave rise to the present genes. Other hypotheses propose that all immunoglobulin structural genes present today evolved from a single primordial gene (15, 20). The striking similarity between the unblocked light-chain sequence in the lungfish, that of the shark, and the sequences of the human κ and λ chains strongly favors the hypothesis that immunoglobulin structural genes evolved from a single primordial gene.

Examination of the compiled sequences reveals that the "deletions" and/or the "additions" of single or groups of amino acids are not to be found only in light chains from lungfish immunoglobulin; various types of deletions and/or additions involving the amino-terminal residues have been demonstrated in mammalian and avian light chains; for example, rabbit N-terminal addition [0][§], pig N-terminal addition [0] plus N-terminal deletion [1], and chicken N-terminal deletion [1,2]. Lungfish light chains illustrate still

§ [] denotes the positions affected, as represented in Table 2, by homology of the sequences in question with human kappa (κ)-type light chains.

TABLE 2. Sequence of N-terminal amino acids of the light chains derived from elasmobranchian, dipnoid, avian, and mammalian immunoglobulins

Species	N-terminal residues												
	0	1	2	3	4	5	6	7	8	9	10	11	12
African Lungfish		Asp	—	—	Leu	Thr	Glx	Asx	Ala	Ser	Met	*	Val
Leopard Shark (11)		Asp	Ile Pro	Val Ile	Leu Met Val	Thr	Glx	Pro	Ala Gly Pro	Ser	Val		
Chicken (12)		—	—	Ala	Leu	Thr	Gln	Pro	Ala	†	Ala	Val	Gly
Mouse (λ) (13)		PCA	Ala	Val	Val	Thr	Gln	Gln	Ser	Ala	Leu	Thr	Thr
Mouse (κ) (14)		Asp	Ile	Val	Met	Thr	Gln	Ser	Pro	Ala	Ser	Leu	Ser
		Glu	Val	Gln	Leu				Thr	Ser	Leu	Ala	Pro
										Leu	Phe	Gln	Pro
Rabbit (κ) (15)	Ala	Asp Ala	Ile Val Phe	Val Glx	Met Val Leu	Thr	Glx	Thr	Pro	Ser Ala	Ser	Val	Ser Thr Glx
Pig (κ) (16)	Ala	—	Ile	Val Glx	Leu Asx	Thr Ser Glx	Glx	Ser	Pro	Ser Thr Leu	Leu	Leu Ala Gly	Ala Pro Ser
Human (λ) (17)		PCA	Ser Tyr	Val Ala Glu	Leu	Thr Ala	Gln	Pro	Pro Ala	Ser Ala	Val Ala	Ser	Gly Val
Human (κ) (17)		Asp Glu	Ile Val Met	Val Gln Leu	Met Leu	Thr	Gln	Ser Thr	Pro	Ser Thr Leu Gly Ala Asx	Ser Thr Phe	Leu	Ser Pro

The sources of the unblocked sequences are provided by reference notations. Deletions are indicated by (—) and are based on a proposed homology with the amino-terminal sequence of human κ -type light chains. Alternative amino acids at given positions are indicated where adequate information is available. No attempt has been made to associate the various alternatives with established light-chain subclasses e.g., κ I, κ II, κ III; such information can be obtained by consulting the original sources. The order of presentation of the alternatives is that found in the original source or represents the authors' citation of existing data. In the latter case, alternate amino acids are listed in descending frequency of occurrence. Precise estimation of the relative proportion of each alternative at a given position is extremely difficult as Ser, Thr, Trp, and half-cystine are destroyed during the acid hydrolysis of PTH derivatives. The estimation of Ser and Thr in some of the above examples is made by gas-liquid chromatography, and is thus qualitative.

Many of the above light chains are derived from sources as diverse as normal immunoglobulins, immunoglobulins found in chemically induced and naturally occurring plasmacytomas, and homogenous immunoglobulins induced by injection of certain streptococcal vaccines. Where the sequence of normal light chains did not extend to include the first twelve residues, the sequences were completed by utilizing sequences of light chains derived from the above sources.

* See Table 1.

† Deletion at this position was proposed by the authors of (12).

another type of amino-terminal deletion, namely an internal [2,3] deletion. From inspection of compiled sequences of Bence-Jones proteins, it is obvious that the occurrence of "micro" deletions, involving relatively few base triplets, is not limited to the amino-terminal portions of the light chain alone (21, 2). A somewhat similar situation, differing in magnitude from that of the above example, is illustrated in the sequences obtained from study of the characteristic proteins of patients with heavy-chain disease. In these cases "macro" deletions, involving fairly extensive portions of the variable region, have been described (22–25). Although the exact operational level of the addition/deletion process is not yet known, its frequent occurrence in the light chains of immunoglobulins derived from various species underscores its importance. Whether this deletion/addition system is a regular genetic mechanism related to the generation of antibody variability remains open to speculation. A genetic mechanism capable of inserting and removing relatively few amino acids would be an ideal means for introducing variability into immunoglobu-

lin chains without sacrificing important overall structural features of the antibody molecule.

We thank Dr. J. R. L. Pink for his advice concerning sequencing methodology and most thoughtful review of this manuscript and Dr. G. G. B. Klaus and Dr. J. W. Goodman for allowing us to view their manuscript on amino-acid sequencing of leopard-shark immunoglobulin before publication. A. C. W. is a fellow of the National Genetics Foundation, R. A. G. is Professor of the American Legion Memorial Research and Regents' Professor of Pediatrics, Microbiology, and Pathology. Aided by grants from The National Foundation-March of Dimes U.S. Public Health Service (AI-08677, AI-00292, AI-09813) University of Minnesota Graduate School, and National Science Foundation (GB27666).

1. Milstein, C., and J. R. L. Pink, in *Progress in Biophysics and Molecular Biology*, ed. J. A. V. Butler and D. Noble (Pergamon Press, Oxford and New York, 1970), p. 209.
2. Pink, J. R. L., A. C. Wang, and H. H. Fudenberg, *Annu. Rev. Med.*, 22, 145 (1971).
3. Litman, G. W., D. Frommel, S. L. Chartrand, J. Finstad, and R. A. Good, *Immunochemistry*, 8, 345 (1971).

4. Marchalonis, J. J., *Aust. J. Exp. Biol. Med. Sci.*, **47**, 405 (1969).
5. Neville, D. M., *Biochim. Biophys. Acta*, **133**, 168 (1967).
6. Gros, C., and B. Labouesse, *Eur. J. Biochem.*, **7**, 463 (1969).
7. Goodman, J. W., G. G. Klaus, D. E. Nitecki, and A. C. Wang, *J. Immunol.*, **104**, 260 (1970).
8. Edman, P., and G. Begg, *Eur. J. Biochem.*, **1**, 80 (1967).
9. Wang, A. C., J. R. L. Pink, H. H. Fudenberg, and J. Ohms, *Proc. Nat. Acad. Sci. USA*, **66**, 657 (1970).
10. Pisano, J. J., and T. J. Bronzert, *J. Biol. Chem.*, **24**, 5597 (1969).
11. Klaus, G. G. B., D. E. Nitecki, and J. W. Goodman, *J. Immunol.*, 1971, in press.
12. Kubo, R. T., I. Y. Rosenblum, and A. A. Benedict, *J. Immunol.*, **105**, 534 (1970).
13. Apella, E., *Proc. Nat. Acad. Sci. USA*, **68**, 590 (1971).
14. Hood, L. E., M. Potter, and D. J. McKean, *Science*, **170**, 1207 (1971).
15. Hood, L., K. Eichmann, H. Lacklund, R. M. Krause, and J. J. Ohms, *Nature*, **228**, 1040 (1970).
16. Novotny, J., F. Franěk, and F. Šorm, *Eur. J. Biochem.*, **14**, 309 (1970).
17. Dayhoff, M. O., R. V. Eck, and L. T. Hunt, *Atlas of Protein Sequence and Structure*, ed. M. O. Dayhoff (National Biomedical Research Foundation, Silver Spring, Md., 1969).
18. Suran, A. A., and B. W. Papermaster, *Proc. Nat. Acad. Sci. USA*, **58**, 1619 (1967).
19. Pollara, B., A. Suran, J. Finstad, and R. A. Good, *Proc. Nat. Acad. Sci. USA*, **59**, 1307 (1968).
20. Hill, R. L., R. Delaney, R. E. Fellows, and H. E. Lebovitz, *Proc. Nat. Acad. Sci. USA*, **56**, 1762 (1966).
21. Edelman, G. M., and W. E. Gall, *Annu. Rev. Biochem.*, **38**, 415 (1969).
22. Frangione, B., and C. Milstein, *Nature (London)*, **224**, 597 (1969).
23. Terry, W. D., J. Ohms, *Proc. Nat. Acad. Sci. USA*, **66**, 558 (1970).
24. Franklin, E. C., and B. Frangione, *Proc. Nat. Acad. Sci. USA*, **68**, 187 (1971).
25. Smithies, O., D. M. Gibson, E. M. Fanning, M. E. Percy, D. M. Parr, and G. E. Connell, *Science*, **172**, 574 (1971).