*ON THE MECHANISM OF ANTIBODY SYNTHESIS: A SPECIES COMPARISON OF L-CHAINS**

By Leroy E. Hood, William R. Gray, and William J. Dreyer

DIVISION OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated by James Bonner, February 15, 1966

Recent remarkable advances in knowledge of the molecular events underlying gene function have made it possible to translate detailed knowledge of protein structure into genetic terms and vice versa. In this paper we present the results of a series of amino acid sequence studies of immunoglobulin light chains. When these and previous data are translated into genetic terms, they place numerous constraints on theories of antibody formation. The results are tentatively interpreted as supporting a theory which apparently violates the classical one gene-one polypeptide chain rule.

In humans, and in other vertebrates, the organism's main defense against disease lies in its ability to produce antibodies. These are complex protein molecules (immunoglobulins) which react with foreign substances, or disease-producing organisms (antigens), and render them innocuous. The antigen-antibody reaction is highly specific, and an extraordinarily large number of different kinds of antibody molecules are required to provide an effective defense against the multitude of antigens encountered in nature. Because of this heterogeneity it has not yet proved feasible to isolate homogeneous antibodies and to study the nature of the differences between them. Such studies would be of immense value in increasing our understanding of how antibodies are made and of how the antibody-producing system has developed.

This difficulty in obtaining pure antibodies has been circumvented by taking advantage of a neoplastic condition (multiple myeloma) of the cells which are normally responsible for making antibodies. Individual tumors appear to arise from the unrestrained division of a single cell, and may secrete a homogenous protein belonging to any one of the major classes of immunoglobulins. Such proteins are amenable to detailed structural studies of a type which are not possible with the heterogeneous populations of antibodies secreted in response to antigens.

Multiple myeloma has been encountered only in humans and in mice. Plasma cell tumor proteins from these two species have been studied for many years, with particular emphasis on those which resemble the light (*L*-) chains of immunoglobulins (Bence-Jones proteins). In an early study of mouse myeloma *L*-chains, 17 proteins were examined by peptide mapping techniques. It was concluded that each protein contained a large amount of sequence which was common to all, while each also contained much that was unique to itself.[1, 2] Further structural investigations[3] suggested that each of these proteins is a single polypeptide chain, and that the common sequence must be localized in a particular region. Concurrent studies on human proteins revealed a similar situation with the exception that minor sequence changes have been found within the common region of the human light chains.[4, 5] More recent data on the amino acid sequences of mouse light chains permits us to say with certainty that all of the common peptides were derived from the carboxyl terminal 107 residues of these proteins.[6] Consideration of these

findings led to a theory of antibody *L*-chain formation in which many genes code for the different "specificity" regions (the amino terminal halves), while a single gene codes for the "common" region (the carboxyl terminal halves).[7]

In this paper we shall present the results of a study of the amino terminal sequences of 13 different light chains derived from mice and from humans. The data are compatible with the assumption that many "specificity" genes code for residues 1–105 in both species. The sequence comparison suggests that these "specificity" genes exist in the germ line and that they must have evolved long before the divergence of mice and humans on the evolutionary tree.

Materials and Methods.—The *L*-chains studied were obtained from the following sources:

(1) Human myeloma *L*-chain proteins were prepared from the urine of multiple myeloma patients, and purified by dialysis, ammonium sulfate fractionation, and gel-filtration on G-100 Sephadex equilibrated with 0.2 *N* ammonium bicarbonate.

(2) Mouse myeloma *L*-chain proteins were obtained from the urine of mice carrying various strains of transplantable plasma cell tumors[2, 8] and were purified by the same methods used for the human proteins.

(3) Human γG myeloma proteins were obtained from the serum of myeloma patients, and were purified by ammonium sulfate fractionation and chromatography on DEAE-cellulose.[9] The proteins were partially reduced and alkylated, and chains were separated on P-200 acrylamide gel, equilibrated with 1.0 *M* acetic acid.[10]

(4) Normal *L*-chains were obtained from pooled human γG and were purified by a procedure similar to that used for γG myeloma proteins.

Each of the *L*-chains was oxidized with performic acid by the method of Hirs[11] before analysis by the phenylisothiocyanate procedure.

Sequential degradation of the proteins from the amino terminal end was carried out using a modified 3-cycle form of the phenylisothiocyanate (PITC) procedure,[12] and terminal amino acids were identified as their phenylthiohydantoin (PTH) derivatives. In addition, samples of the PTH derivatives were hydrolyzed with base, and the liberated amino acids were further characterized as their dimethylaminonaphthalene sulphonyl ("Dansyl") derivatives.[13]

Results.—The sequences determined by the PITC procedure are shown in Table 1. *L*-chains from the three tumor sources (mouse myeloma *L*-chains, human myeloma *L*-chains, and human γG myeloma proteins) show only limited variations in sequence. It is remarkable that the differences found among the proteins from mice mirror those found among the human proteins. All proteins studied possess the same amino acid at the second, fifth, and sixth positions. At the first and third positions there are two alternatives and at the fourth position there are three. The single terminal sequence reported by Titani *et al.*[4] also fits this pattern. Observations on pooled human *L*-chains are in complete accord with the findings among the myeloma proteins: that is, two alternatives are found at the first and third positions and the same amino acid is found in the second position in all cases (the reported *N*-terminal residues found for pooled human *L*-chains represent 90–95% of the end groups found at each step[14]). Thus, if all the homogeneous *L*-chains derived from the tumors were mixed together and a "pooled analysis" done, the results would be essentially identical to those shown for the normal human *L*-chains.

Discussion.—These results in conjunction with other experimental observations

TABLE 1

AMINO TERMINAL SEQUENCES OF *L*-CHAINS

Source of light chains		Residue Number 1	2	3	4	5	6
Mouse Bence-Jones proteins	MBJ 41	Asp	Ilu	Gln	Met	Thr	Gln
	MBJ 70	Asp	Ilu	Val	Leu	Thr	Gln
	MBJ 6	Asp	Ilu	Val	Val	Thr	Gln
	MBJ 9	Asp	Ilu	(Gln)			
Human Bence-Jones proteins	HBJ 3	Asp	Ilu	Val	Leu	Thr	Gln
	HBJ 12	Glu	Ilu	Val	Val	Thr	Gln
	HBJ 10	Asp	Ilu	Gln	Met	Thr	Gln
	HBJ 6	Asp	Ilu	Gln	Met	Thr	Gln
	HBJ 1	Asp	Ilu	(−)	Met	Thr	Gln
	HBJ 5	Glu	Ilu	Val	(Leu)		
	HBJ 4	Asp	Ilu	Val			
	Ag*	Asp	Ilu	Gln	Met	Thr	Gln
Human 7*S* myeloma proteins	HS 4	Glu	Ilu	Val	Leu	Thr	Gln
	HS 6	Glu	Ilu	Val	Leu	Thr	Gln
Pooled γ-globulin (normal human)		Asp + Glu	Ilu	Val + Gln			

* Taken from Titani *et al.*[4]

permit us to make the following conclusions: (*a*) the amino terminal sequences are remarkably similar in all of the chains studied here; (*b*) the differences at any one position in the sequence are restricted to a small number of amino acids; (*c*) each of the light chains studied to date which has been derived from different plasma cell tumors of the mouse contains a different amino acid sequence;[1, 2] (*d*) these molecules are single polypeptide chains;[3] (*e*) the variations in amino acid sequence occur from position 1 through 105 in the sequence of the light chains;[4–6] (*f*) there appears to be no sequence variation in the entire carboxyl terminal region of these mouse light chain molecules (positions 106–212); (*g*) small changes noted in the carboxyl terminal region of light chains from humans probably represent genetic heterogeneity in the human population;[4, 5] and (*h*) genetic and structural studies indicate that at least some portion of the light chains, presumably the carboxyl terminal half, is coded by a single gene.[4–6, 15]

Numerous theories have been advanced over past decades in an effort to explain the extraordinary phenomenon of antibody production.[7, 16–22] We shall discuss some of these theories in general terms.

(1) *A single gene carried in the germ line undergoes hypermutation during somatic differentiation:* It is unlikely that this hypothesis is correct, since the *L*-chain sequence differences are highly restricted even in pooled human *L*-chains. To account for differences of this type it would be necessary to assume either that the mutation mechanism is such as to allow only two or three alternatives, or that intense selection during somatic differentiation eliminates all other mutant forms. No single mutation mechanism yet encountered could give rise to the observed differences without also producing many other alternatives. It is pertinent to remember that an average triplet code word can generate five to seven amino acid replacements by single base changes alone. For example, the triplet AAC (coding for asparagine) can generate triplets coding for lysine, aspartic acid, histidine, tyro-

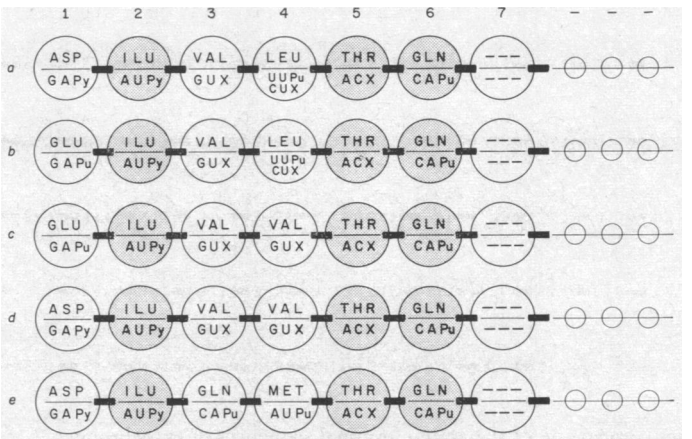

FIG. 1.—Base sequences of RNA coding for amino terminal sequences of five *L*-chain variants (see Table 1).[26] Shaded residues are invariant. Abbreviations: *Pu*, A or G; *Py*, U or C; and *X*, any one of the four bases.

sine, serine, threonine, or isoleucine by this mechanism. Similar objections arise when considering other mutational processes such as inversion of base sequences, multiple crossing over, or reading-frame shifts (see Fig. 1).

We are unable to conceive of a really intense mutation-selection system operating in the absence of an antigen. It is now known that differentiation of the plasma cells, including specification of the *L*-chain, does occur prior to the introduction of the antigen.[23]

In addition to these objections, we find it difficult to envision either a hypermutability process, or a selection mechanism which operates exclusively on one half of the *L*-chain molecule.

(2) *A single gene remains constant throughout differentiation, while a specialized translation mechanism develops in such a way as to cause the changes in sequence:* It has been suggested that special code words might occur in the "specificity" region of a single gene.[22] These could then be translated differently, depending upon which of a small number of alternative activating enzymes had been selected during differentiation. The main objection to this theory is that all such ambiguous code words would have to be strictly limited to one region of this particular gene, since the carboxyl terminal half of these proteins, and the complete sequences of other proteins, are always expressed in one form only. Other objections arise when we consider what the special code words might be, whether we assume that they are some of the redundant code words of the normal dictionary, or a supplementary set of words employing unusual nucleotide bases. We feel that both possibilities are very remote, but are amenable to experimental test.

(3) *Multiple germ line genes code for the "specificity" region and a single gene codes for the "common" region:* The data reported herein, taken together with previously known facts, seem to support the hypothesis that any one of many different genes may code for the first 105 amino acid residues. Only one of these "specificity" genes would normally be active in a given plasma cell. Residues 106–212 are assumed to be encoded by a single "common" gene, thus leading to the same sequence

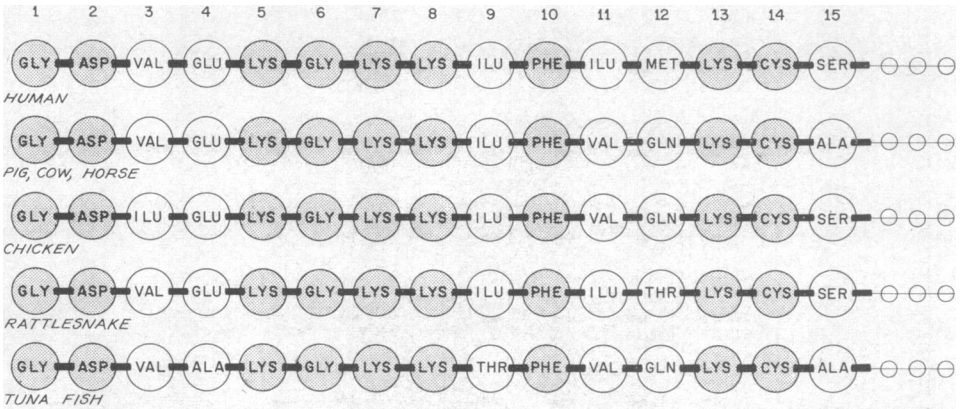

FIG. 2.—Species comparison of the amino terminal sequences of cytochrome c.[27] Shaded residues are invariant.

in each of the mouse light chains of the κ class.[8] According to this theory, the specificity genes arose by the slow process of chemical evolution in a manner not unlike that which has led to the species variations which occur in the amino acid sequences of other proteins. A species comparison of the first 15 residues of cytochrome c is presented in Figure 2. Qualitatively speaking, the same general kinds of restrictions on sequence variation that are seen in cytochromes are also seen in the amino terminal region of the light chains.

Either of two mechanisms which are known in microbial systems might be utilized to unite the information from two separate genes, so as to give a single L-chain product.[24, 25] The first possibility is that the two genes operate independently to produce separate polypeptide chains which are then joined enzymatically. Such a mechanism would be plausible using an enzyme with a double trypsin-like specificity for lysine and arginine. All human and mouse L-chains for which we have sequence data have Lys or Arg at position 105 and Arg at 106.[4-6] (Note that residue 106 is the start of the common region.) This general class of enzymatic mechanisms is utilized for the synthesis of glutathione and cyclic antibiotics.[24] Experiments designed to test the possibility of this enzymatic mechanism are in progress.

A second possible mechanism is that the fusion of informational macromolecules (DNA or RNA) occurs during differentiation of the specific plasma cells. This process would resemble in many respects the incorporation of a lambda bacterial virus into an *E. coli* bacterial genome. The result would be a new combined gene (or genes) which would be relatively stable and heritable during subsequent cell divisions in a manner analogous to the genome of *E. coli* K12 λ.[7, 25]

Thus, we favor the hypothesis that, following differentiation of a plasma cell precursor, two originally distinct genes, coding separately for the N-terminal and C-terminal halves of a light chain, become expressed as a single, continuous polypeptide chain. The genetic mechanisms involved in this unusual phenomenon are clearly of interest in themselves, and may well be of more general significance in terms of other types of cellular differentiation and organogenesis.

Summary.—We present the results of sequence analysis of the N-terminal regions of a number of human and mouse L-chains. Several theories of antibody

formation are discussed in the light of these and earlier structural findings. The most reasonable theory of *L*-chain formation seems to be one in which many genes code for the amino terminal halves of the molecules, while a single gene codes for the remainder. Mechanisms for synthesizing the complete *L*-chain from the information contained in a "specificity" gene and the "common" gene are discussed.

Note added in proof: Since this paper was submitted, Milstein's results on human Bence-Jones proteins have become available for comparison [*Nature*, **209**, 370 (1966)]. From a study of the amino acid sequences around cysteine residues in a number of proteins, he concluded that the variations observed could not be accounted for by any one of several simple mutational mechanisms examined. This is one of the conclusions reached in this work also. The amino acid sequences are known around the cysteines in two mouse proteins,[6] and fit closely the pattern observed in the human proteins; in both cases the cysteine residues occupy the same positions in the two species (residues 23 and 86 according to the numbering of Hilschmann[5]).

The authors thank Dr. R. Doolittle for extensive discussion and helpful advice on the use of the modified 3-cycle PITC procedure. They also thank Drs. W. Buckley and J. Gally for the human light chains used in these studies.

* This work was supported by U.S. Public Health Service research grant GM-0-6965 and U.S. Public Health Service postdoctoral fellowship (1-F2-GM-20,388-01) (LH).

[1] Potter, M., W. J. Dreyer, E. L. Kuff, and K. R. McIntire, *Federation Proc.*, **22**, 649 (1963).

[2] Potter, M., W. J. Dreyer, E. L. Kuff, and K. R. McIntire, *J. Mol. Biol.*, **8**, 814 (1964).

[3] Bennett, J. C., L. Hood, W. J. Dreyer, and M. Potter, *J. Mol. Biol.*, **12**, 81 (1965).

[4] Titani, K., E. Whitley, Jr., L. Avogardo, and F. W. Putnam, *Science*, **149**, 1090 (1965).

[5] Hilschmann, N., and L. C. Craig, these PROCEEDINGS, **53**, 1403 (1965).

[6] Status of amino acid sequence work: Sequence studies on one mouse *L*-chain are almost complete. Approximately 95 per cent of the molecule is accounted for by two long stretches of polypeptide chain (1–70 and approximately 86–212). Details of positions 71–85 are not yet complete. Within these regions we have deduced a unique and fully independent ordering of peptides, and 75 per cent of the amino acids are assigned to exact positions. Other proteins are being studied concurrently, and many of the "common" peptides have been analyzed. The peptides designated as common in the peptide-map survey are derived from residues 141–212, and completely account for this part of the molecule. Residues 106–140 are contained in a single large tryptic peptide which is not identifiable on peptide maps. Peptides defining this region have been obtained from two proteins, and are found to be identical in composition. There are very many differences between the mouse and human sequences in the carboxyl terminal halves of the molecules, with about 40 amino acid replacements. Several amino acids of particular structural significance (Pro, Tyr, Cys, His) probably occupy identical positions in both species. A report of this aspect of the project will be published elsewhere (Gray, W. R., W. J. Dreyer, and L. Hood, in preparation).

[7] Dreyer, W. J., and J. C. Bennett, these PROCEEDINGS, **54**, 864 (1965).

[8] The mouse proteins were all taken from the same one of the two distinct antigenic types of *L*-chains [designated as λ by McIntire, K. R., R. M. Asofsky, M. Potter, and E. L. Kuff, *Science*, **150**, 361 (1965)]. The human proteins were all of the κ type. The structural studies indicate striking similarities between these two classes of proteins at both the *N*-terminal and *C*-terminal ends. In contrast, there is only very slight resemblance between the mouse λ type and the human λ type [Milstein, C., *Nature*, **205**, 1171 (1965); Hood, L., unpublished results]. It would thus seem appropriate to reverse the proposed nomenclature of mouse *L*-chains to correspond with the well established nomenclature for the human proteins.

[9] Sober, H. A., and E. A. Peterson, *Federation Proc.*, **17**, 1116 (1958).

[10] Fleischman, J. B., R. H. Pain, and R. R. Porter, *Arch. Biochem. Biophys.*, Suppl. 1, 174 (1962).

[11] Hirs, C. H. W., *J. Biol. Chem.*, **219**, 611 (1956).

[12] Edman, P., *Ann. N. Y. Acad. Sci.*, **88**, 602 (1960).

[13] Gray, W. R., and B. S. Hartley, *Biochem. J.*, **89**, 59P (1963).

[14] Three human λ *L*-chains were blocked in the *N*-terminal position; hence in the *N*-terminal analysis of pooled human γG *L*-chains one is probably looking only at κ chain residues.

[15] Dray, S., G. O. Young, and L. Gerald, *J. Immunol.*, **91**, 403 (1963).

[16] Pauling, L., *J. Am. Chem. Soc.*, **62**, 2643 (1940).

[17] Karush, F., *Trans. N. Y. Acad. Sci.* (Series II), **20**, 581 (1958).

[18] Lederberg, J., *Science*, **129**, 1649 (1959).

[19] Smithies, O., *Nature*, **199**, 1231 (1963).

[20] Burnet, M., *Nature*, **203**, 451 (1964).

[21] Smithies, O., *Science*, **149**, 151 (1965).

[22] Potter, M., E. Apella, and S. Geisser, *J. Mol. Biol.*, **14**, 361 (1965).

[23] Jerne, N. K., A. A. Nordin, and C. Henry, in *Cell Bound Antibodies*, ed. B. Amos and H. Koprowski (Philadelphia: Wistar Inst. Press, 1963), p. 109.

[24] Mach, B., and E. L. Tatum, these PROCEEDINGS, **52**, 876 (1964).

[25] Campbell, A. M., *Advan. Genet.*, **11**, 90 (1962).

[26] Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, **53**, 1161 (1965).

[27] Dayhoff, M. O., R. V. Eck, M. A. Chang, and M. R. Sochard, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, 1965).

INHIBITION OF LYSOZYME SYNTHESIS BY ACTINOMYCIN D IN BACTERIOPHAGE T4-INFECTED CELLS OF ESCHERICHIA COLI

BY JAY J. PROTASS AND DAVID KORN

LABORATORY OF BIOCHEMICAL PHARMACOLOGY, NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, NATIONAL INSTITUTES OF HEALTH

Communicated by C. B. Anfinsen, February 16, 1966

There is considerable evidence that in T-even bacteriophage infection of cells of *E. coli* there is a temporal sequence of phage-directed protein synthesis.[1-4] However, the mechanisms controlling "early" and "late" protein formation are not well understood. Two opposing theories are as follows: (*a*) The input phage genome transcribes only "early" messenger RNA (mRNA), while "late" mRNA is transcribed from newly synthesized genome replicas.[5, 6] This theory thus proposes that regulation operates at the level of *DNA transcription*. (*b*) Both "early" and "late" mRNA species can be transcribed from the input phage genome, but "late" mRNA is not expressed during the early phase of the infectious cycle. This theory, then, assigns a major role in regulation to *mRNA translation*. Indirect evidence supporting this latter hypothesis has recently been published.[7]

In this paper we present some initial results of our study of factors controlling phage-directed protein synthesis. Using T4-infected, EDTA-treated[8] cells of *E. coli*, we have investigated the effect of actinomycin D on the production of the "late" phage-directed enzyme, lysozyme. The data indicate that the formation of mRNA capable of directing lysozyme synthesis does *not* significantly precede the time at which lysozyme activity becomes demonstrable in the infected cells.

Materials and Methods.—The experiments were performed with bacteriophage T4 and *E. coli* CR34. Conditions of infection and the method of sensitization of *E. coli* to actinomycin D by EDTA treatment were as previously described.[9] Crude extracts were prepared from infected cells by rapidly chilling aliquots of the culture