IMMUNOGLOBULINS: STRUCTURE, GENETICS, AND EVOLUTION

by

Leroy E. Hood

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1968

(Submitted November 29, 1967)

This thesis is dedicated to the Grand Canyon country and the Sierra Nevada for the constant renewal and release they have provided these past four years — to my wife, Valerie, who shared my wilderness pleasures — and to all who fight to preserve the mountains and the canyons for future generations.

ACKNOWLEDGMENTS

Individuals in the Caltech community who made my four years exciting and rewarding are too numerous to name, but special thanks must go to

W.J. Dreyer for the broad horizons which he provided;

R.D. Owen for his warm enthusiasm and ready counsel;

W.R. Gray for his fruitful suggestions and penetrating insights;

R.F. Doolittle for his warm friendship and constant encouragement;

B.G. Sanders for his thoughtful advice and able collaboration;

and J.C. Bennett for a stimulating and enjoyable introduction to science.

ABSTRACT

The immune system is capable of generating an immense number of different antibody molecules. The nature of the genetic machinery responsible for this diversity has been studied by selective amino acid sequence analysis of homogeneous immunoglobulin light chains (derived from myeloma tumors). The evolution of the immune system has also been examined through chemical studies of normal pooled light chains derived from various mammalian and avian species. These studies place constraints on proposed genetic mechanisms for antibody diversity. The theories, the structural constraints, and the evolutionary implications of these observations are discussed.

TABLE OF CONTENTS

CHAPTER I

THE NATURE OF THE PROBLEM

DEFINITIONS

The immune response, found only in vertebrates, is the organism's main line of defense against infection. This homeostatic mechanism may, for the purpose of discussion, be broken down into three components — a stimulus, antigen; the effector system, the immune system; and the response, antibody. A great deal of information has accrued in each of these areas which places constraints on the mechanistic nature of the immune response. A terse summary of the salient features of each will be useful at this point.

Antigen (1,2): The input to the immune system, antigen, is any substance (protein, carbohydrate, nucleic acid, lipid, or any mixture) capable of initiating an antibody response. Generally antigens are foreign to the organism. The size of the antigen is important, for commonly molecules of less than 4,000 molecular weight are not antigenic. This limitation is concerned with a failure to process small antigens properly and not with an intrinsic inability to make antibodies to small molecules, as is demonstrated by the ability of very small (and relatively simple) groups (e.g. dinitrophenol or DNP) to initiate a good antibody response after coupling to a larger molecule (the carrier). Such small molecules are termed haptens. Most antigens, being large molecules compared to haptens, have many antigen determinants, that is, discrete sites which can evoke the synthesis of specific antibody. It should be pointed out that almost any molecule can be made antigenic

by coupling to an appropriate carrier or through chemical modification of its tertiary structure.

The antigen is apparently "processed" in some fashion by macrophages soon after entry into the organism's body (see article by E. Lennox and M. Cohn (3) for a review of this processing). The nature of this processing is a poorly understood phenomenon. Among many puzzling points are the following: 1) Much antibody appears to be directed against surface antigenic determinants. When these surface determinents are modified (by denaturation through reduction and alkylation), then antibody is produced against the denatured protein which frequently does not react with the native material (4). Whatever "processing" occurs (i.e. presumably partial hydrolysis of the antigen), the organism must preserve detailed surface information for activation of the antibody producing system. 2) Fishman et al. (5,6) have evidence which suggests that antigenic "information" is coupled to RNA in the macrophage and that it is this antigenic determinant – RNA complex which is responsible for initiating specific antibody synthesis in the appropriate lymphoid cells. 3) The experiments of Campbell and Garvey (7) indicate that antigenic determinants can remain sequestered in the organism for years, possibly coupled to RNA. The roles of long-stored antigenic fragments and coupling of antigen to RNA remain unclear.

Immune System (8,9): The immune system is comprised of lymphoid organs (bone marrow, thymus, spleen, lymph nodes) which make up about 0.5% of vertebrate body weight. The interrelationships of these organs

are not well understood although the bone marrow is thought to generate immune stem cells, the thymus is thought to initiate a maturation process (either through hormonal factors or through direct cellular contact) to produce competent cells (i.e. cells capable of participating in immune reactions), and finally the spleen and lymph nodes are thought to provide the architectural framework for antigen processing (via macrophages) and activation of mature immunologically competent cells to effect the final immune response (e.g. antibody production).

Cell lineage of the immune system is another area of great confusion, particularly at the earlier stages of the development of competent cells (9). It has been established that lymphocytes of all sizes (small, medium and large) as well as the classic plasma cells are antibody producers (10). Furthermore, there appear to be at least two populations of small lymphocytes, those with an extremely long half-life (probably years) and those with a much shorter life span (11).

The paucity of knowledge in all of these areas is due in large part to the necessity of dealing with the immune system in vivo at least until very recently (12). Hence the features described are average properties of large populations of cells of mixed lineage. There are, however, certain general properties of this system worth considering (13): 1. Primary response — upon initial exposure to a given pathogen (e.g. staphylococcus) the immune system generally takes several days to synthesize antibodies which are capable of combining with the pathogen and rendering it innocuous. 2. Secondary response — upon subsequent exposure to this pathogen the response is more rapid and the antibody

produced more effective; that is, there appears to be an immunological memory. 3. The immune system does not generally make antibodies against the individual's own tissues; it is tolerant to them. 4. Extremely high doses of antigen produce immunologic paralysis, a condition in which no effective antibody is made. For any unified theory of antibody formation these features must be explained in terms of the responses of single cells. Perhaps the most fruitful single approach toward understanding the molecular workings of the immune system has been an analysis of antibody structure.

Antibody: The hallmark of the immune response is the specificity of the antibody – antigen reaction. The entire population of antibody molecules is given the designation immunoglobulins. Presumably most molecular species of immunoglobulin are the product of a specific immune response directed against a particular antigen. Hence the molecular heterogeneity of the immunoglobulin population is, in part, a consequence of antigenic diversity in the environment.

To return for a moment to the problem of antibody specificity, let us consider this phenomenon at the molecular level. A simple chemical group such as dinitrophenol (DNP) when linked to an appropriate carrier, can stimulate the production of complementary antibody. How can a foreign stimulus (DNP) evoke this highly specific response at the molecular level? Does the DNP molecule somehow provide the organism with information which is translated into complementary antibody, or on the other hand, does this antigenic determinant only combine with

a preformed receptor to trigger cell division and antibody synthesis (i.e. the immune response)? It was this particular aspect of the immune response which directed thinking about the genetic and molecular basis of antibody formation for more than half a century. It will be profitable to consider the evolution of this thought for the historical insight which it gives into the two major theories of antibody formation, the instructionistic and selectionistic postulates, one of which (selectionistic) still serves as a framework for the formulation of modern theories.

HISTORY

The initial attempt to formulate a general theory of antibody formation was made by Ehrlich in 1900 (14). He postulated that all cells had many receptors (side chains) which allowed them to seize a wide variety of molecules for food or other purposes. Many specificities were present in these side chains which permitted combination with a broad spectrum of different molecules, including antigens. Specific combination of a particular side chain with a molecule of complementary configuration was postulated to stimulate the synthesis of large numbers of identical "side chains", many of which broke away to become soluble antibody. The antigen played a selective role in that it reacted with a pre-existent complementary side chain to initiate a specific antibody response. Implicit in this concept was the hypothesis that every antibody pattern was pre-existent in the cell and that the antigen

served merely as a trigger to initiate the immune response. This was tne first formulation of the selectionistic theory.

After the turn of the century immunological studies seemed to rule out the possibility of preformed patterns, for enormous numbers of substances, both artificial and natural, were shown to be antigenic (15,16). Very few of these antigens were cross reactive; that is, each antigen seemed to provoke a fairly unique antibody response. It appeared unlikely that an organism could have such a wide range of preformed patterns. Rather, each antigen was presumed to direct in some way specific antibody production. Subsequent thought was directed toward understanding the nature of this process. The idea that antigen served as a template for antibody formation was clearly enunciated by Alexander (17), Mudd (18), and Breinl and Haurowitz (19) in the early 1930's and was elaborated and modified by Pauling (20) in the 1940's.

In considering the perplexing problem of how a prodigious number of different antibodies can possibly exist, Pauling suggested that antibody diversity could result from folding of the same peptide chain in different ways. In Pauling's words: "It is assumed that antibodies differ from normal serum globulins only in the way in which the two end parts of the globulin polypeptide chain are coiled, these parts, as a result of their amino acid composition and order, having accessible a very great many configurations with nearly the same stability; under the influence of an antigen molecule they assume configurations complementary to surface regions of the antigen, thus forming two active ends" (20). The direct template theory required the persistence of

antigen throughout the duration of the antibody response and raised difficult questions about the molecular basis of immunologic memory.

An "indirect template" theory was proposed by Burnet and Fenner in 1949 (21) and expanded by Burnet in 1956 (22). The suggestion was that the antigen might induce a specific and heritable change in the cellular mechanism responsible for antibody synthesis. The implication was that antigenic determinants could modify the genetic material of the cell so as to cause the production of complementary antibody. These changes were permanent and could be passed on to subsequent generations of cells. This theory did not require long term persistence of antigenic fragments in antibody producing cells in order to explain the phenomenon of immunologic memory.

The direct and indirect template theories came to be known as instructionistic theories of antibody formation.

A radical departure from instructionistic thought was taken by Jerne in 1955, when he proposed (as had Ehrlich) a selective role for the antigen (23,24).

The essence of Jerne's theory was as follows:

1) The gamma globulin molecules of normal plasma represent a population of specificities which can combine with all of the antigenic determinants the organism is likely to meet.

2) The function of the antigen is to act as a carrier of complementary antibody to a system of cells which can reproduce this antibody.

3) It is presumed that once antibody is taken into this antibody producing system, replicas of this natural antibody are produced.

4) This process is potentiated by further experience with the same antigen because the level and degree of specificity of circulating antibody is increased by the primary response.

5) Those antibodies reacting to the organism's tissue and cells ("self") are removed from the circulation before the antibody producing system matures. The organism thereby becomes tolerant to its own tissues.

Thus Jerne's theory was an attempt to explain many different facets of the immune response (e.g. tolerance, the enhanced secondary response, etc.) quite in contrast to the earlier theories which focused on explaining antibody specificity. Jerne failed, however, to provide a convincing mechanism for the synthesis of specific antibody. It was not clear how antibody producing cells could be stimulated to produce a specific antibody merely by the presence of that antibody molecule in combination with complementary antigen.

The clonal selection hypothesis of Burnet (25,26) was a logical extension of Jerne's theory. In this theory, Jerne's population of natural antibodies was replaced by clones of cells; each clone was capable of synthesizing one genetically determined type of antibody. Thus the "clonal selection theory" postulated the existence of "clones of mesenchymal cells each carrying immunologically reactive sites corresponding in appropriately complementary fashion to one (or possibly a small number of) potential antigenic determinants" (26). Every antigenic determinant that we are likely to encounter is presumably represented in the body by a clone of cells capable of making complementary antibody. When the antigen is introduced, it makes contact with

the receptor sites on a complementary clone, initiates synthesis of specific antibody and stimulates clonal expansion through cell division.

Lederberg used the clonal theory as a point of departure in postulating a specific genetic model for antibody synthesis. He formulated nine propositions including the following: "A2. The cell making a given antibody has a corresponding unique sequence of nucleotides in a segment of its chromosomal DNA: 'its gene for globulin synthesis.' A3. The genetic diversity of the precursors of antibody producing cells arises from a high rate of spontaneous mutation during their lifelong proliferation. A4. This hypermutability consists of random assembly of the DNA of the globulin gene during certain stages of cellular proliferation" (27). Hypermutation implied that the information necessary for making all antibody specificities was not contained in the zygote, rather it accumulated by a random process during somatic differentiation (somatic theory). This was a departure from Ehrlich (14) and Jerne (23) who had contended that each antibody producing cell had all the genetic information necessary for making all the antibodies (germ-line theory). (An attempt to distinguish between these possibilities, the somatic and the germ-line theories of antibody formation, is the focus of Chapter 3 in this thesis.)

When I commenced my studies in 1963, the selectionistic theories appeared to have much the stronger case (although there was a great deal of controversy among proponents of the rival theories):

1) It was difficult to imagine a molecular model for the template process. All of the available information on protein synthesis suggested

that DNA makes RNA which in turn serves as a template for protein synthesis. Instructionistic theories required that one protein (antigen) instruct a second (antibody). Furthermore, it appeared evident that the primary amino acid sequence governed the tertiary configuration of the protein molecules. A template process would require that one protein (antigen) direct the synthesis of a second (antibody) and furthermore, that antigen confer complementary specificity on the newly synthesized antibody. Clearly such a proposal stretched the imaginations of even the most vocal instructionists. The selective theories of Burnet and Lederberg were quite compatible with the accepted machinery for protein synthesis.

2) General features of the immune system were difficult to explain in terms of a template process. Tolerance (recognition of self), immune paralysis, and the enhanced secondary response were readily explained by the selectionistic proposals (see Jerne's theory).

3) The general analysis of immunoglobulin structure, which is the subject of this thesis, also clearly pointed toward the selectionistic hypothesis as we shall see in subsequent discussion.

GENERAL ANTIBODY STRUCTURE

Theories of antibody formation reviewed in the previous section were based entirely on general features of the immune system (antibody specificity, immunologic memory, tolerance, etc.). As work progressed

on the chemical structure of immunoglobulins (antibodies), an exciting new possibility arose for analysis of the immune response at the molecular level (28,29). Presumably a detailed chemical analysis of antibody structure would shed light on the nature of the genetic machinery underlying this fascinating process. However, this analysis was complicated by the heterogeneity of immunoglobulins (30).

The heterogeneity of antibodies is extremely easy to demonstrate at the functional level. A simple hapten (DNP) can stimulate the synthesis of specific antibodies with binding constants which vary over a range of 10^6, demonstrating an overwhelming degree of active site heterogeneity (31). On the other hand, all immunoglobulins can be divided into discrete types or families of molecules (e.g. IgG, IgA, IgM) by physical, chemical and immunologic criteria (molecular weight, carbohydrate content, antigenic classification)(28,29). Specific antibody to a particular antigen can generally be found in all of the major types (32). Furthermore, all types of immunoglobulins appear to share a common subunit structure. Perhaps this is best illustrated by considering in detail the general structure of the most common antibody type, IgG, which comprises 85-90% of the normal serum immunoglobulin.

This class can be isolated from the remainder of immunoglobulins by simple ion exchange chromatography (33). The IgG molecule is roughly ellipsoid in shape (240Å x 57Å x 19Å) with an active site at either end (34,35). It has a molecular weight of about 160,000 and is constructed of four polypeptide chains, two identical light chains (molecular weight approximately 23,000) and two identical heavy chains (molecular weight

approximately 50,000). Each light chain is joined through a disulfide bond to one heavy chain to form the basic immunoglobulin subunit, a light-heavy chain pair (36). The IgG molecule is made up of two such identical subunits probably joined by disulfide bonds between the heavy chains (37,38,39). Each of the other immunoglobulin types (e.g. IgA, IgM) has a similar light-heavy chain subunit structure.

A majority of the studies which have attempted to localize antibody specificity lead to the conclusion that both light and heavy chains are directly involved in the active site (40), hence structural studies of either chain can provide information on the genetic and structural basis of antibody diversity.

In brief summary, immunoglobulins exhibit 1) extreme molecular heterogeneity, 2) common chemical and physical properties which suggest extensive regions of structural similarity (identity), and 3) a common subunit structure, the light-heavy chain pair. The difficulty of isolating a homogeneous antibody from the serum of an animal is obvious. The techniques of protein chemistry were not equal to the isolation of a single molecular species of immunoglobulin from a serum containing many other closely related antibody species. The isolation of homogeneous antibody for detailed chemical analysis seemed virtually impossible. However, nature once again provided an ingenious solution for this apparent dilemma — there was one simple system which could provide homogeneous immunoglobulins.

MULTIPLE MYELOMA AND HOMOGENEOUS IMMUNOGLOBULINS

Multiple myeloma, a cancer of antibody producing cells (plasma cells), produces homogeneous immunoglobulins (41,42,43,44,45). Presumably individual myeloma tumors are initiated by the unrestrained division of a single cell which in time generates a large clone of cells, each synthesizing the same protein. The immune system of the myelomatous animal at later stages of the disease produces essentially a homogeneous immunoglobulin as the cancerous clone comes to effectively replace the normal immune cells.

The myelomatous process can affect cells capable of synthesizing any one of the different types of immunoglobulin (IgG, IgA, IgM)(43). Frequently, myeloma cells produce only a light chain, which is small enough to pass through the kidney filter (43). These subunits can be collected from the urine relatively free of other proteins and are termed Bence-Jones proteins (after the British investigator who first described them).

Structural and immunologic observations provide compelling evidence for the supposition that myeloma proteins are in fact members of the normal pool of immunoglobulins for 1) both share a wide range of antigenic specificities (46), 2) the type ratio of immunoglobulin proteins is similar in the normal and the myeloma pools (46,47) 3) peptide maps and sequence studies confirm the apparent similarity of these two populations (48,49) and 4) recently, myeloma proteins have been found with antibody activity (50,51).

Myeloma probably occurs in many different animals, but this disease has been studied most thoroughly in humans and a single inbred strain of mouse (BALB/c). Tumors arise spontaneously in humans; they can be induced by intraperitoneal injection of mineral oil in mice (only in the BALB/c strain) (44). The mouse system permits serial transfer of individual tumors. A single myeloma tumor, passed through more than one hundred generations, was found to synthesize the same homogeneous immunoglobulin throughout (52). The genetic mechanism responsible for immunoglobulin synthesis in this case appears to be stable and heritable. This is convincing support for the hypothesis that plasma cell genes (i.e. stable and heritable units) direct the synthesis of immunoglobulins.

The concern as to whether or not the myeloma process selects cells from a special subset of the normal immunoglobulin producing population, particularly in the mouse system where the tumors are artifically induced, is unanswerable at this time. Nevertheless, the system appears to have exciting possibilities. Myeloma tumors synthesize homogeneous immunoglobulin (antibody) which is easily purified. Detailed comparisons of different tumor proteins (i.e. individual antibodies) seemed to provide a system which would permit molecular and genetic constraints to be placed on the mechanism responsible for antibody diversity.

LIGHT CHAIN STRUCTURE

Early structural studies of myeloma proteins quite naturally centered on light chains because light chains are smaller in size and

easily purified. These chains can be divided into two distinct classes by antigenic classification (46). These two classes, termed L and K, share few if any peptide spots (45,48), suggesting that they are encoded by separate genes. In contrast, when peptide map comparisons are made among different myeloma light chains of a given class and species, each protein shares ten or eleven tryptic peptides (common peptides) with all of the others (48,45). In addition, each protein has a unique set of tryptic peptides (distinguishing peptides) which are not shared with any of the other proteins. A simple explanation for this puzzling observation comes to mind if we consider the peptide maps of fetal and adult hemoglobulins; that is, a common set of peptides is shared by both proteins whereas a second set is unique to each protein. The fetal hemoglobin molecule is composed of two distinct polypeptide chains, α and γ; adult hemoglobin has α and β chains. It seemed likely that the common and distinguishing regions of light chains have a similar molecular explanation and that two separate polypeptide chains were joined either by an unknown covalent bond or by noncovalent association.

POINT OF DEPARTURE

This was the point of departure for my studies. To summarize, it appeared:

1) Central to an understanding of immune function is the genetic mechanism which is responsible for generating an antibody diversity of

immense proportions. Theoretically this problem can be approached through a study of any of the parts of the immune response (antigen, the system or antibody). Practically, however, not all of these approaches are equally fruitful.

2) Studies of antigen and the constraints it places on antibody mechanisms are hopelessly confused by our lack of knowledge in the area of antigen processing.

3) The immune system is an extremely complex and sophisticated biologic entity. Our understanding of its component parts is at a primitive level — due necessarily to the lack of in vitro studies which permit the analysis of individual components.

4) The analysis of antibody seemed an extremely promising approach. A system was available for producing many different homogeneous immunoglobulins (antibodies). By analysis and comparison of different myeloma proteins one could compare different antibodies at the protein structural level. Furthermore, this analysis could readily be extended to the level of nucleic acid sequence comparisons through the availability of the genetic code.

5) Since light and heavy chains both appeared to be involved in the antibody site, elucidation of the molecular nature of either chain would probably provide clues as to the nature of the genetic mechanism responsible for generating antibody diversity. The light chain seemed the obvious candidate because of its smaller size and ease of purification.

6) The first step in defining the nature of light chain variability was elucidation of the structural linkage between the common and the distinguishing regions. I began with an investigation of the nature of this linkage.

CHAPTER II

LIGHT CHAINS: ONE POLYPEPTIDE OR TWO?

INTRODUCTION

Light chain structure presents an intriguing problem at the molecular level. On the one hand, light chains can be separated into two distinct classes, L and K, with different properties by immunologic and peptide map criteria (45,46,48) — and are presumably synthesized by separate genes. On the other hand, peptide map comparisons of proteins within a single class demonstrate an amino acid sequence which is shared by all chains (common peptides) and at the same time, a sequence unique to each tumor protein (distinguishing peptides)(45,48).

These observations immediately bring to mind a number of relevant questions:

1. How are the common and variable regions distributed in the light chain molecule? One might argue that common and distinguishing sequences are interspersed. Such a postulate appears unreasonable both from the lack of precedent for such a structure and in the difficulty in understanding how the gene for such a protein might work. Rather it appears likely that the common peptides represent a single contiguous stretch of polypeptide chain common to each protein and that the distinguishing peptides likewise represent a single contiguous stretch of sequence unique to each light chain.

2. How are these common and distinguishing sequences linked? Three general mechanisms can be suggested: a) Two separate polypeptide chains are linked through one or more covalent bonds. b) Two separate polypeptide chains are noncovalently associated. c) The two regions

are adjacent parts of a single polypeptide chain. There are precedents at the molecular level for a) and b). For example, the two chains of insulin are covalently linked through disulfide bonds whereas the four chains of hemoglobin are associated through noncovalent bonds. Implicit in mechanisms a) and b) is the suggestion of multiple polypeptide chains. The third mechanism only requires a single polypeptide chain, but it would suggest "unusual" events at the level of transcription or translation in order to a) generate many different light chain sequences from a single gene or to b) join two separate genes (those encoding the common and distinguishing regions) or their peptide products together to produce a single peptide chain. The most reasonable assumption at this time seemed to be that light chains were comprised of two separate polypeptide chains.

A literature survey revealed two preliminary reports (53,54) which suggested that myeloma light chains could be converted into two separate polypeptide chains by complete reduction of disulfide bridges. In neither case were the separated polypeptides characterized, and the possibility of nonspecific proteolysis has to be considered. (The urine and kidney have many proteolytic enzymes.) A perusal of the literature revealed that other light chain studies had been carried out under conditions that would not have cleaved all disulfide bonds. Hence the existence of disulfide linked common and distinguishing polypeptide chains appeared to be a possibility worth investigating more rigorously.

Mouse myeloma light chains were submitted to a variety of procedures which destroyed all disulfide bonds. Attempts were then made to fractionate the resulting polypeptide chains in a number of different systems.

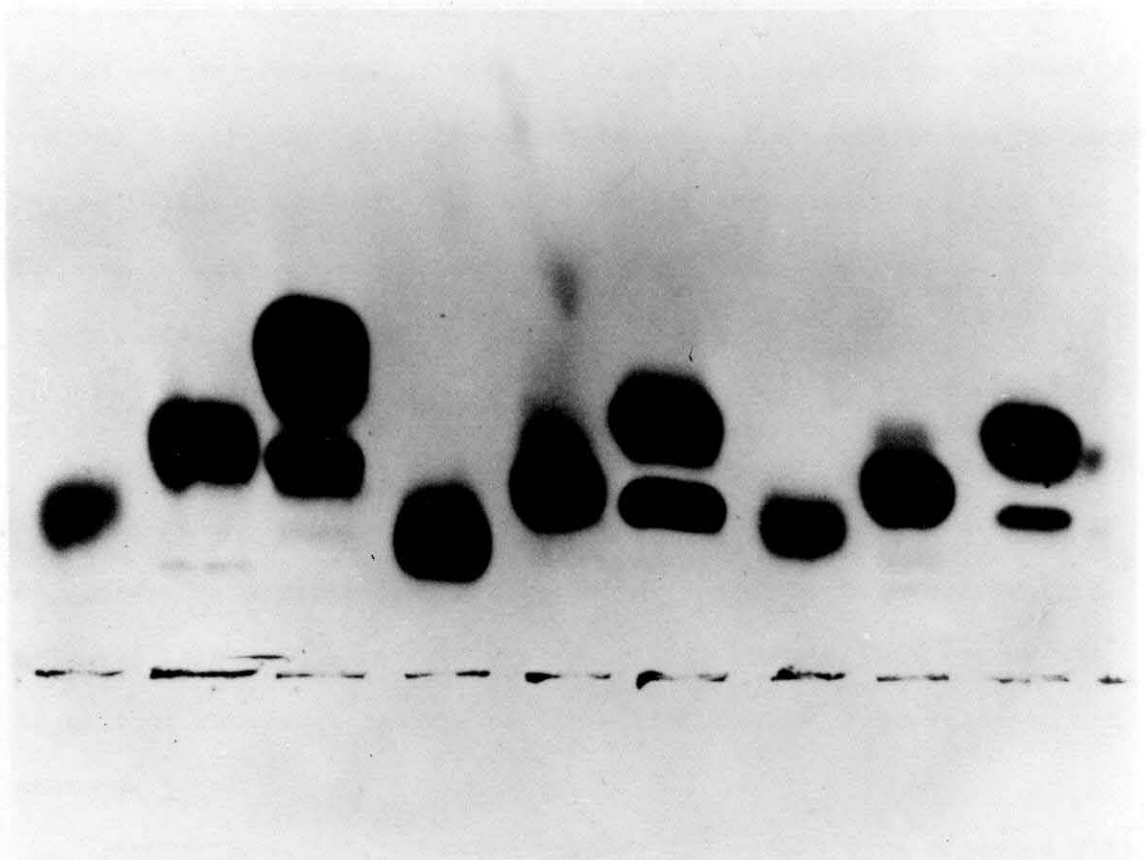
MATERIALS AND METHODS

Nomenclature: Myeloma urinary light chains will be designated MBJ (Mouse Bence-Jones protein) or HBJ (Human Bence-Jones proteins).

Isolation and Purification: Myeloma light chains were isolated from the urine of BALB/c mice as described by Potter et al. (45). Dialyzed and lyophilized samples of the urinary proteins were eluted through G-100 Sephadex equilibrated with 0.2 M-ammonium bicarbonate.

Electrophoresis: A method of horizontal acrylamide gel electrophoresis was adapted from the procedure of Sogami and Foster (55). Gelling of 100 ml of 7% acrylamide and 0.4% N,N'-methylene bis acrylamide was initiated with 1 ml of fresh 10% ammonium persulfate and 0.1 ml N,N,N,N'-tetramethylenediamine. This solution was poured into plastic trays (½" x 3¼" x 4"), covered with a glass plate to exclude air and allowed to gel for 1 hour. Gels were removed from the trays and equilibrated overnight in 8 M-urea and 0.05 M-acetic acid. The proteins were applied by insertion of cellulose acetate strips which had been soaked in 2% solutions of the samples. Electrophoresis was carried out for 3 hours at 75 volts. The gels were stained 30 minutes in 1% amido

Figure 1. Acrylamide gel electrophoresis. See text for description of method using electrophoresis at pH 3 in 8-M urea. The anode is near the origin, the cathode at the top of the figure. Samples from left to right: reduced-alkylated MBJ 47; reduced MBJ 47; untreated MBJ 47; reduced-alkylated MBJ 41; reduced MBJ 41; untreated MBJ 41; reduced-alkylated MBJ 9; reduced MBJ 9; untreated MBJ 9.



black 10B (Hartman-Ledden Co.) in 7% acetic acid and destained in 4% acetic acid for several days.

Reduction and Alkylation: Light chains (10 mg/ml) were reduced in 10 M-urea, 0.5 M-mercaptoethanol and 0.1 M-NH_4HCO_3 at 37°C for 18 hours. Urea, which had been rendered cyanate-free by passage over a mixed-bed ion exchange resin, was freshly prepared each time. Following reduction, iodoacetic acid (recrystallized from ether and petroleum ether) was added in tenfold molar excess over mercaptoethanol. The pH was adjusted to 8.5 with NH_4OH and the reaction was allowed to proceed for 15 minutes at room temperature. The reaction was terminated by the addition of tenfold molar excess of mercaptoethanol over the iodoacetic acid. Samples were exhaustively dialyzed against water and lyophilized.

Gel Filtration of Reduced Proteins: Samples of MBJ 70 and MBJ 9 were reduced (as above) without alkylation. These proteins were then eluted through G-100 Sephadex equilibrated with 1.0 M-acetic acid in an attempt to separate any components or differing molecular weight released through the reduction of disulfide bonds. This column was capable of resolving gamma globulin, light chains, ribonuclease, and acetone.

Peptide Maps(56): Protein samples (10 mg/ml) were digested with trichloroacetic acid-precipitated trypsin (Worthington, lot no. 591) at 37°C for 18 hours in 0.1 M-NH_4HCO_3 (pH 8.2) with an enzyme to substrate ratio of 1 to 20. Approximately 1.5 mg of each sample was applied to Whatman no. 3 chromatographic paper and subjected to 20

hours of descending chromatography followed by electrophoresis at right angles to the chromatographic dimension. A phenol red marker was used to follow chromatographic development. The chromatographic solvent was the organic phase (upper phase) of a butanol-acetic acid-water (2/:8:40) mixture. Electrophoresis was done in pyridine acetate buffer (acetic acid-pyridine-water, 10:1:289) pH 3.6 at 3200v (55v/cm) for 65 minutes in a Gilson Medical Electronics model D electrophorator with the origin near the anode. The papers were dried 20 minutes in an oven at 80°C before staining.

Stains: Ninhydrin (57). 600 ml ethanol, 200 ml of glacial acetic acid, 30 ml of 2,4,6-trimethylpyridine and 1 g of ninhydrin produce a stain which gives color differentiation of certain peptides. Amino terminal glycine, serine, proline and asparagine produce a yellow color whereas most other amino terminal residues stain various hues of purple. The peptide map was dipped in this solution and developed for 7 minutes at 80°C.

Pauly stain for histidine and tyrosine. 50 mg. of β-diazosulfanilic acid was dissolved in 50 ml of 10% Na_2CO_3, and the resulting mixture was sprayed lightly onto the peptide paper. Histidine containing peptides take on a cherry red while tyrosine residues stain a reddish-brown color.

Ehrlich stain for tryptophan (58). 2 g of p-dimethylaminobenzaldehyde were dissolved in a solution of 180 ml of acetone and 20 ml of concentrated HCl. The peptide paper was dipped immediately into this solution and allowed to air dry for a few minutes. The tryptophan-containing peptides develop a deep blue-purple color.

Order of staining. Peptide maps were made in duplicate. One paper was sprayed lightly with Pauly solution, then dipped in the ninhydrin reagent to locate all peptides precisely and finally dipped in the Ehrlich solution. The second paper, which was dipped only in ninhydrin, was photographed and information from the multiple staining procedure was added.

Immunology: Immunodiffusion and immunoelectrophoresis were carried out on microscope slides containing 3.0 ml of 1% agar made up in 0.075 M-vernal buffer, pH 8.6 (B2 Buffer, Beckman Instrument Co.) as described in Ouchterlony (59) and Scheidegger (60). Diffusion and electrophoresis slides were developed for 24 hours at room temperature in a humid chamber. Antigen concentrations for both procedures were approximately 10 mg/ml for all samples except serum which was used undiluted.

Antiserum to light chains was developed in rabbits with subcutaneous toe pad injections of 10 mg of protein in 2 ml of saline/complete Freund's emulsion (1:1). One month later two 5 mg I.V. injections in saline were given two days apart, and a week later the animals were bled. In some cases it was necessary to give a second I.V. booster.

In later experiments it was necessary to check the purity of IgG proteins from many different species. To do this, rabbit antiserum was prepared against whole serum from each of these animals and IgG purity was examined by immunoelectrophoresis. IgG proteins migrate to a characteristic position which serves to differentiate them from many other serum contaminants. These rabbit antisera were prepared as

described above except that 1 ml of serum was used for the primary injection and ½ ml for each secondary I.V. injection.

Amino Acid Analysis: Protein samples were hydrolized for 20 hours with constant boiling (5.7 N) HCl at 105°C in evacuated ampules. After hydrolysis, the HCl was removed in vacuo over P_2O_5 and NaOH. The method for automatic amino acid analysis used in this laboratory is a micro-version of the procedures described by Piez and Morris (61). High resolution resin was obtained from the Spinco Division of Beckman Instrument Co. Hydrolysate corresponding to approximately 0.1-0.2 mg protein was applied to the column.

Oxidation and Thin Layer Electrophoresis: 1 mg samples of MBJ 70 and MBJ 9 were performic acid oxidized (62), lyophilized and electrophoresed on a thin (0.5 mm) layer of silical gel (80% formic acid, 15°C, for 90 minutes at 200v and 10 ma.). The thin layer plates were prewashed with 80% formic acid by descending chromotography in order to remove impurities from the silica gel. The center of each plate was masked while the edges were sprayed with ninhydrin to locate peptides. The unstained center portions of the protein zones were scraped from the plate, eluted with 80% formic acid and lyophilized after six-fold dilution with distilled water.

Titration of Sulfhydryl Groups: Proteins were treated with 5-5' dithiobis (2-nitrobenzoic acid) at pH 8.0 according to the method of Ellman (63), and the free sulfhydryl content was calculated from the optical density at 412 mμ.

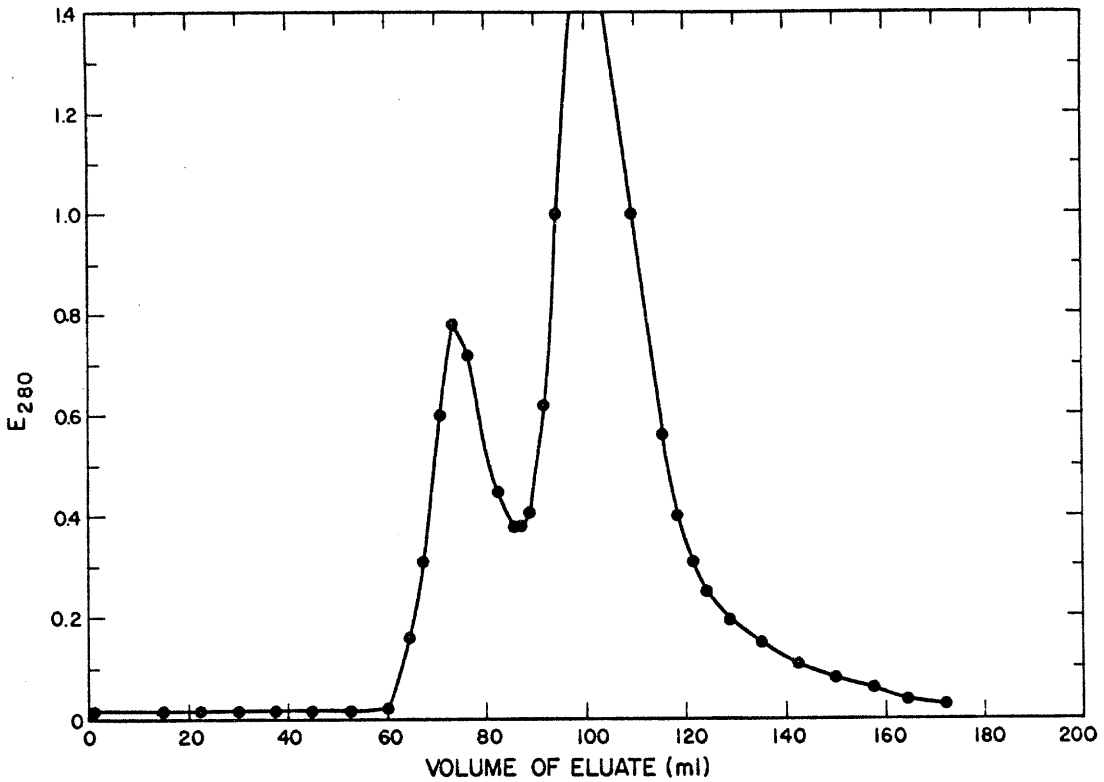
Carbohydrate Stain for Light Chains: MBJ 9, 41, and 70 were submitted to electrophoresis using the disc acrylamide gel system of Davis (64). A modification of the periodic acid-Schiff technique was used to stain the acrylamide gels for the presence of carbohydrate (65).

RESULTS

Electrophoresis: Electrophoretic patterns of 3 light chains in acrylamide gel are shown in Figure 1. Each protein is shown in three molecular states: 1) native protein dissolved in 8 M-urea, 2) reduced protein and 3) reduced and alkylated protein. All proteins have two forms in the unreduced state which, upon reduction with or without alkylation, are converted to a single major protein band. It should be pointed out that the staining technique used would probably not detect small peptides as they would not be acid-fixed in the gel. Nevertheless, these results imply, but do not prove, that each light chain exists in a monomer and a dimer form — the dimer being formed through the disulfide linkage of two monomer units.

Gel Filtration: The gel filtration pattern of MBJ 9 (Figure 2) shows the separation of two subunits when the protein is the native form. Peptide maps of tryptic digests of both forms appear to be identical. Reaction with Ellman's reagent reveals 0.8 free sulfhydryl groups per mole of monomer (mol. wt. assumed to be 24,000) and 0.0 in the dimer. This is strong evidence for the existence of monomers and dimers in light chains purified from urine. Further studies have shown that the

Figure 2. Fractionation of MBJ 9 on a column (1.75 cm x 125 cm) of Sephadex G-100 in 0.2 N-NH_4CHO_3 at 2°C. The small peak is dimer, the larger peak monomer.


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
E_{280}
0
20
40
60
80
100
120
140
160
180
200
VOLUME OF ELUATE (ml)

monomer is the component with the greatest electrophoretic mobility in the acrylamide procedure (Figure 1).

Immunological Reactivity: Figure 3 shows the immunological reactivity of five different myeloma light chains, including the three studied. The antiserum in the center well was produced against MBJ 63. It can be seen that these proteins have numerous antigenic sites in common. It should also be stressed that spur formation seen with MBJ 63 indicates that it has serologically reactive sites which are not found in any of the other proteins tested. A similar pattern is observed when antiserum formed against a different light chain is used.

The existence of serologically reactive sites unique to MBJ 63 implies that this protein has antigenic determinants and hence unique amino acid sequence which is not shared by any of the other light chains. The fact that similar spurs can be noted in antisera prepared against other light chains suggests that each light chain has a segment of unique amino acid sequence. On the other hand, each of these proteins also shares antigenic determinants with all the others, so a segment of amino acid sequence must be common to all light chains.

Peptide Maps: Tryptic peptide maps of three light chains are shown in Figure 4. On the left of the figure are shown the photographs of the actual fingerprints while on the right tracings are shown to help relate peptide similarities and differences. There is a set of peptides common to all light chains examined; however, in each case clear cut peptide differences are also demonstrated. The specific stains, noted on the tracings, support these observations.

Figure 3. Ouchterlony plate: reaction of MBJ 41, MBJ 70C, MBJ 30, MBJ 63 and MBJ 9 against a light chain antiserum (anti-MBJ 63) to show immunologic cross-activity (numbers on Figure).

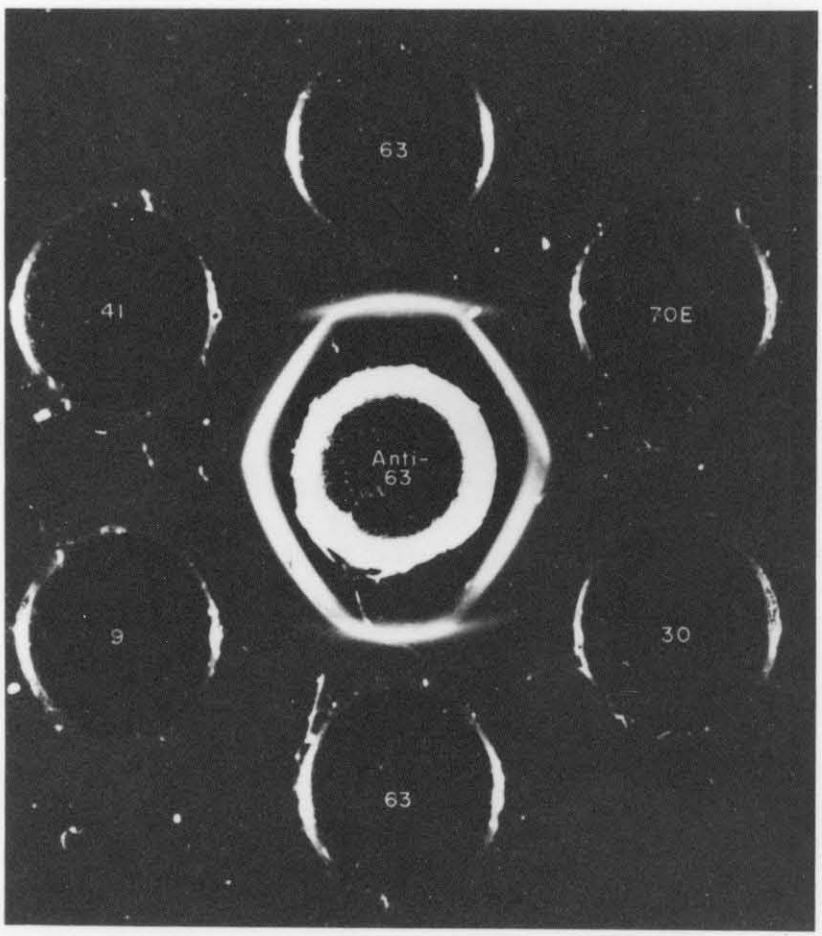
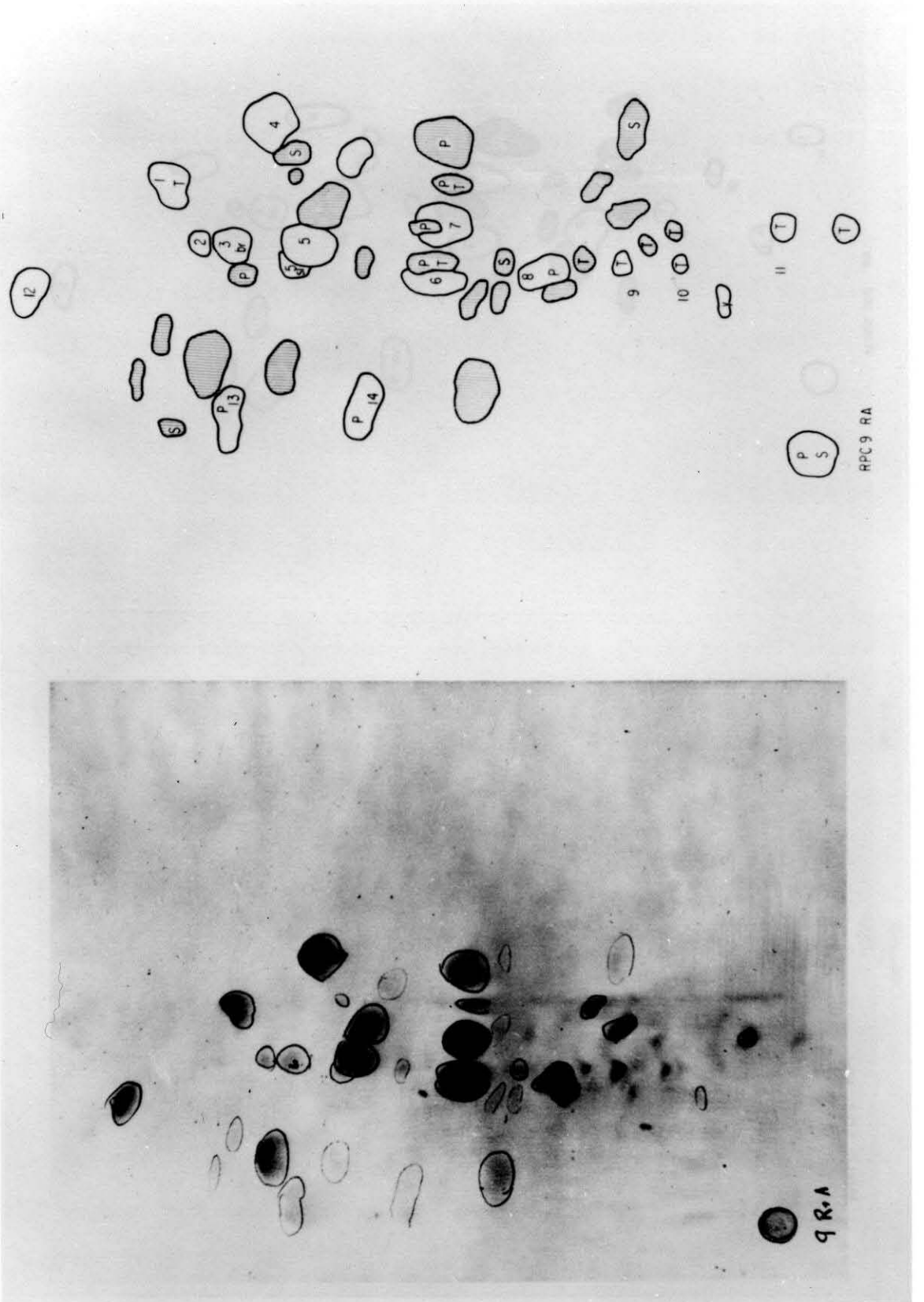
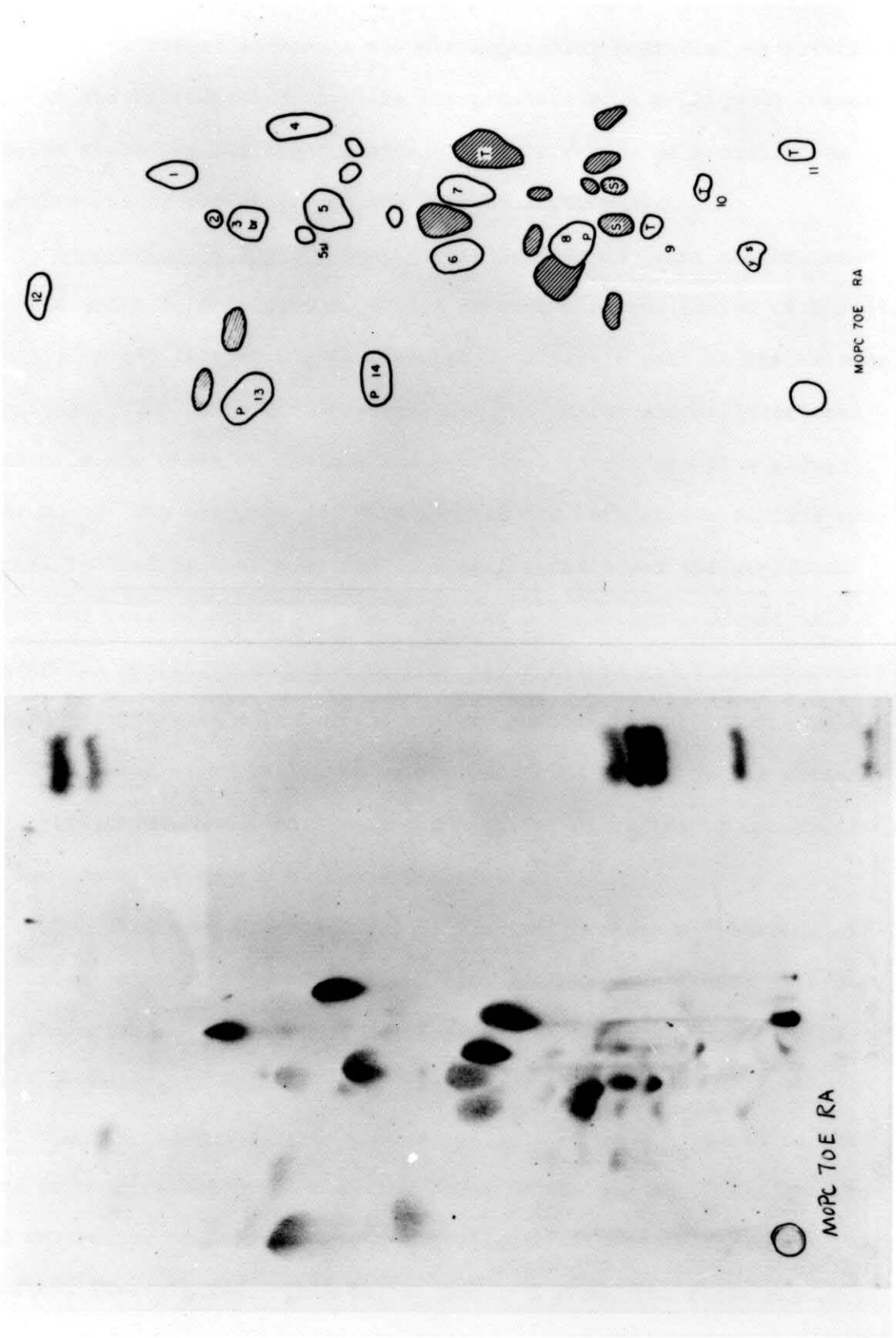

63
41
70E
Anti-
63
9
30
63

Figure 4a,b,c. Peptide maps: see text for description of method. Photographs and tracings of the peptide maps are shown: (a) MBJ 70E; (b) MBJ 41; (c) MBJ 9. Electrophoresis was carried out in the long (vertical) dimension and chromatography in the short (horizontal) dimension. The origin is in the lower left-hand corner. The positive electrode was on the origin side during electrophoresis, and peptides moved toward the negative side. Open circles on the tracings represent the peptides common to all light chains; shaded ones are unique to the specific molecule studied. The following letters indicate reaction with specific residue stains: E = Ehrlich for tryptophan, P = Pauly for tyrosine or histidine, Br = peptide staining brown with ninhydrin color dip.

MOPC 41 RA
41 Rat


RPC9 RA
9 R.A


MOPC 70E RA
MOPC 70E RA

The existence of common and distinguishing peptides, as pointed out in the introduction, implies the presence of a contiguous common region shared by all light chains and the presence of a continuous distinguishing region (s) unique to each light chain.

Separation of Peptide Chains: Fully reduced light chains, separated under acid conditions (which prevent the reoxidation of disulfide bonds) by gel filtration, were eluted as a single peak at the monomer position. The amino acid analyses for the native and fully reduced proteins are given in Table 1 and expressed in numbers of residues per mole. The analyses are very similar for both states of each protein, indicating that a second polypeptide chain was not separated from the reduced material. Performic acid oxidation produced only 1 ninhydrin positive component on thin layer silica gel electrophoresis in 80% formic acid which, when examined on the automatic amino acid analyzer, had essentially the same composition as the native protein. No evidence was obtained from the removal of disulfide or noncovalently linked material by any of these techniques.

In comparing MBJ 9 and MBJ 70 (Table 1), there are obvious differences in amino acid composition far exceeding the limits of error of the method. This further confirms the presence of unique amino acid sequences in each of these proteins.

Carbohydrate Analysis: MBJ 9, 41, and 70 each appeared to have one or more carbohydrate moieties bound to the protein. The question of whether or not this moiety was covalently linked to the light chain was not investigated. This is, as far as I know, the first report of carbohydrate association with light chains.

TABLE 1

Amino acid compositions of mouse light chains. An approximate molecular weight of 27,000 was assumed for MBJ 70E on the basis of preliminary sedimentation studies. For purposes of comparison, MBJ 9 was assumed to have the same number of valines (12) as MBJ 70E. The molecular weight of MBJ 9 is unknown, and the amino acid differences listed above are strictly relative. Analyses of unreduced MBJ 70E and MBJ 9 were carried out on two independent preparations.

TABLE 1

Amino Acid Compositions of Mouse Light Chains

Amino acids	MBJ 70E	MBJ 70E	Reduced MBJ 70E+gel filtration	MBJ 9	MBJ 9	Reduced MBJ 9+gel filtration	Average residues MBJ 70E	Average residues MBJ 9	Difference MBJ 9 MBJ 70E	Amino acids
Asp	28.5	28.2	28.0	24.1	23.9	23.8	28.2	23.9	-4.4	Asp
Thr	19.7	19.2	20.0	25.8	25.3	25.4	19.6	25.5	+5.8	Thr
Ser	32.8	32.3	29.8	41.8	41.4	42.7	32.3	41.6	+10.4	Ser
Glu	25.0	24.6	24.0	25.8	25.5	24.4	24.5	25.2	-0.1	Glu
Pro	16.2	16.4	15.1	15.7	15.1	15.0	15.9	15.3	-0.9	Pro
Gly	18.0	17.7	17.1	16.4	16.0	16.0	17.6	16.1	-1.6	Gly
Ala	15.7	18.2	13.9	17.7	17.3	17.0	15.9	17.3	+1.1	Ala
Val	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	0.0	Val
Cys	4.6	4.3	4.9	4.1	4.2	2.3	4.6	3.5	-2.3	Cys
Met	4.7	4.7	5.2	6.4	6.6	6.3	4.9	6.4	+1.4	Met
Ile	11.0	10.7	9.8	11.3	11.3	10.8	10.5	11.5	+0.3	Ile
Leu	14.9	14.7	12.6	12.9	12.9	11.7	14.1	12.5	-1.6	Leu
Tyr	7.1	6.9	7.7	11.0	11.2	11.3	7.2	11.2	+4.0	Tyr
Phe	11.6	11.6	11.1	10.5	10.5	10.5	11.4	10.5	-0.9	Phe
Lys	15.3	14.2	15.5	16.4	16.8	16.9	15.0	16.7	+1.7	Lys
His	3.3	3.6	4.2	2.7	2.7	2.6	3.7	2.7	-1.0	His
Arg	8.2	8.2	8.1	9.9	10.5	9.1	8.2	9.8	+1.6	Arg

DISCUSSION

This detailed analysis of three mouse myeloma light chains presents convincing evidence for the presence of a polypeptide sequence common to the three proteins (common tryptic peptides and immunologic cross reactivity) which is combined in each case with a unique stretch of polypeptide chain (immunologic spur formation, differences in amino acid analysis and distinguishing tryptic peptides). These results confirm and extend earlier light chain studies (45,48).

A likely explanation for these observations is that each light chain is comprised of two polypeptide chains, covalently linked or noncovalently associated (see introduction).

Since Adams-Mayne et al. (53) and Van Eijk et al. (54) had presented preliminary evidence that disulfide cleavage of human myeloma light chains produces two dissimilar polypeptide chains, the following attempts were made to separate such chains: a) both fully reduced and fully alkylated samples were electrophoresed in acrylamide gels; b) reduced samples were eluted through G-100 Sephadex in 1 N-acetic acid (acid conditions prevent the reoxidation of disulfide bonds) and subjected to quantitative amino acid analysis; c) samples were oxidized (under conditions which should cleave all disulfide bonds), dissolved in 80% formic acid (an excellent polypeptide solvent), and then electrophoresed on thin layer silica gel plates and stained with ninhydrin, a reagent capable of detecting even small peptides. In each case we failed to demonstrate the presence of multiple polypeptide

chains linked by disulfide bonds. These same tests should also have detected the existence of two distinct polypeptide chains noncovalently associated. Note that these procedures were designed to separate peptide chains by size as well as charge. Furthermore, Bennett had ruled out the possibility of covalent ester and phospho-ester bonds in these proteins (66).

The two polypeptide chain hypotheses (covalent linkage or non-covalent association) appear most unlikely. Rather we are forced to the conclusion that the two light chain regions (common and distinguishing) are parts of a continuous polypeptide chain.

Our understanding of light chain structure was enhanced by the partial amino acid sequence studies of three human K light chains published in 1965 and 1966 (67,68,69). Each light chain is a single polypeptide chain somewhat larger than 200 amino acids. The carboxy terminal halves (approximately 100 amino acids) of these three proteins are identical except for the presence of a single amino acid difference in one chain. The amino terminal halves display numerous amino acid interchanges although a remarkable amino acid homology is evident throughout this region. The structural basis for the peptide map, immunologic, and amino acid analysis comparisons of light chains is obvious. The common region is the carboxy terminal half of the light chain molecule whereas the distinguishing region is the amino terminal portion.

It seems valid to draw some general conclusions about light chain structure at this point: 1) the amino terminal half of the light chain

is concerned with antibody specificity, for this is the only region of the molecule in which changes occur from one antibody light chain to the next — hence we will designate it the specificity region (S); and 2) the carboxy terminal half is probably necessary for more general light chain functions (e.g. binding to the heavy chain) — it will be designated the common region (C).

Now that light chain variability was localized, specific experiments could be designed to elucidate the nature of this variability which, hopefully, would yield clues as to the nature of the genetic mechanism responsible for antibody diversity.

CHAPTER III

TWO GENES — ONE POLYPEPTIDE CHAIN

INTRODUCTION

The antibody molecule exhibits an exquisite molecular complementarity for the antigen which initiates its synthesis. This complementarity is similar to that which enzymes display toward their respective substrates. The enzyme-antibody comparison can be extended to highlight some of the central questions regarding genetic mechanisms for generating antibody diversity. Both enzymes and antibodies are linear polypeptide chains folded in a complex three dimensional array. In both, the primary amino acid sequence directs the folding of these linear chains and thereby controls the nature of the active sites. A separate gene controls the synthesis of each different type of enzyme molecule, and presumably the same is true of each type of antibody molecule. This picture is complicated, at least in the case of antibody light chains, by the fact that each myeloma protein (of a given class) has a common carboxy terminal sequence and a unique amino terminal sequence. These considerations raise a number of important questions regarding antibody variability. 1) How many antibody genes are present in each immunologically mature individual? A lower estimate of the antibody light chain pool size can be obtained by the following calculation: Nineteen mouse K chains were analyzed by peptide map analysis (45); each had a unique set of distinguishing peptides. The size of the light chain pool from which 19 random proteins may be drawn, such that each sequence is unique, is approximately 100 at the 90% confidence level (70). Since this is only a lower limit to the

pool size, each individual must be capable of generating a very large number of antibody light chain genes. 2) What is the relationship between genes encoding the specificity region and those encoding the common region? 3) What genetic mechanism is responsible for generating this large number of antibody genes? Are these genes inherited through the germ line (germ-line theory) or are only a few genes inherited and the diversity generated from these during differentiation of the immune system (somatic theory)?

A study of the nature of amino acid variability in the specificity region of light chains provides an experimentally feasible approach to these questions. One can determine the amino acid sequences of a number of light chains and compare them with one another to look for meaningful patterns of variation which might reflect the underlying genetic mechanism responsible for antibody diversity. As the genetic code dictionary is available, these light chain sequences can be translated into partial nucleotide sequences, and thereby the search for meaningful patterns can be extended to the DNA level.

There are two general approaches to such a comparative amino acid sequence study. On the one hand, the entire amino acid sequence of a limited number of proteins can be determined; such complete determinations are necessary to provide a general framework for analyzing light chain variability. On the other hand, it is an analysis of the variability in the specificity region which is going to provide more specific insights into the genetic mechanism of antibody variability. Therefore, a method was developed to compare large numbers of proteins at their

amino termini (15-20 residues). Fifteen light chains are compared in this region.

NOMENCLATURE

The variable region of light chains (approximately residues 1-107) will be designated as the "specificity" (S) region since it presumably confers antibody specificity. The shared sequence will be termed the "common" (C) region (approximately residues 108-214). The specificity region of the K class will be designated SK and the common region, CK. Different families of sequences within the specificity region of the K class (subclasses) will be designated SK_I, SK_{II}, and SK_{III}. Occasionally there will be a need to denote a particular position in the light chain sequence. This will be indicated in parenthesis after the region and class designation, e.g. SK(1) would denote the amino terminal residue in the specificity region of a K light chain.

METHODS

The intent, in part, of this work was to develop new and improved methods to facilitate sequence comparisons of large numbers of proteins; therefore some of the new procedures will be described in detail.

Purification: Human myeloma light chains were prepared from the urine of patients with multiple myeloma. Urines were dialyzed exhaustively against 0.2 N-ammonium bicarbonate and the proteins precipitated

by adding equal volumes of saturated ammonium sulfate solution to the dialysis residue. The precipitates were washed, redissolved in distilled water and fractionated on G-100 Sephadex equilibrated with 0.2 N-ammonium bicarbonate at 4°C.

Human IgG myeloma proteins were obtained from the serum of myeloma patients. An equal volume of saturated ammonium sulfate was added to the serum, the precipitate was spun down, redissolved in distilled water and reprecipitated twice in a similar fashion. The precipitate was exhaustively dialyzed against 0.01 M-sodium phosphate buffer at pH 6.8 and eluted in the same buffer from DEAE Cellulose or DEAE Sephadex G-50 (33). The eluate was dialyzed against distilled water and lyophilized.

Chain Separation: The IgG proteins were dissolved (30 mg/ml) in 0.5 M-tris buffer titrated to pH 8.6 with concentrated HCl and were reduced in 0.1 M-beta mercaptoethanol at room temperature for 4 hours. A 10% molar excess of iodocetamide (over mercaptoethanol) was added and allowed to react for 1 hour at 4°C. This solution was dialyzed overnight against 1 N-acetic acid and the chains separated on P-200 Biogel (3.5 cm x 150 cm) or G-100 Sephadex (2.5 cm x 125 cm) equilibrated against 1 N-acetic acid. Normal light chains were obtained from pooled human IgG by a similar procedure. This procedure reduces and alkylates only the interchain disulfate bonds. Iodoacetamide was generally used for the alkylation procedure because light chains so treated were more soluble at pH 8 than those treated with iodoacetic acid.

In later experiments IgG was reduced without alkylation. In this procedure, chains were reduced, dialyzed directly against 1.0 N acetic acid and applied to a gel filtration column equilibrated with 1.0 N acetic acid.

Enzymes: Trypsin (Worthington, 3 x recrystallized, lot T3RL 6258) was reacted with L- (1-tosylamide - 2 phenyl) ethyl chloromethyl ketone (TPCK) to produce TPCK-trypsin as described by Kostka et al. (71). This TPCK treatment blocks chymotryptic activity. Chymotrypsin (Worthington 3 x recrystallized, lot CDl 6077), subtilisin (Nutritional Biochemicals Corporation, lot 238RG), and the TPCK-trypsin were stored at -20°C in 0.001 N-HCl at 10 mg/ml. Enzyme digestions were carried out with 1/50 ratio (w/w) of enzyme to peptide in 0.1-N ammonium bicarbonate for 4 hours at 37° C.

Pepsin (Worthington, 2 x recrystallized) was stored at -20°C in 0.01 N-HCl and 0.5 M-NaCl at 2 mg/ml. Digestions were carried out with a 1/50 (w/w) ratio of enzyme to peptide in 5% formic acid at 23°C for one hour (72).

100 mg of carboxypeptidase A (Worthington, 2 x recrystallized, lot COA 690) was washed with 2 ml of water three times, dissolved in 10 ml of 2.0 M-ammonium bicarbonate and reacted with 50 µl of diisopropylflurophosphate (DFP). A slight precipitate was spun down. 20 µl aliquots of this solution were placed in small pyrex tubes (0.5 cm x 5 cm), frozen, sealed under nitrogen and stored at -20°C until just prior to use. Enzymes digestions were done in 0.2 N-ammonium bicarbonate a 1/1 (w/w) ratio (peptide to enzyme) in 28°C water bath.

For kinetic studies, aliquots were taken at 1,3,9,27,60,180,540 and 1080 minute intervals, and the reaction was stopped by immersing each aliquot in boiling water for 45 seconds. Free amino acids were then analyzed by high voltage electrophoresis on paper.

Carboxypeptidase B (Worthington, 20 mg/ml lot COB 6 ED) was treated with 5 µl of DFP per ml and portioned out in 20 µl aliquots as described for carboxypeptidase A. Digestions were carried out in 0.1 N-ammonium bicarbonate at 1/50 (w/w) ratio of enzyme to peptide.

Hydrazinolysis (73,74): This procedure is used to determine the carboxy terminal amino acid of proteins. Non-carboxy terminal amino acids are converted to hydrazides. The free carboxy terminal amino acid is then separated from the hydrazides by ion exchange chromatography. With peptides the ion exchange step can be eliminated as the hydrazides are not present in sufficient quantities to hinder direct analysis of the free amino acids.

Anhydrous hydrazine (95%, Matheson-Coleman-Bell Division of Matheson Company) was redistilled in vacuo over calcium oxide. The constant boiling fraction came off at 25°C. 2 ml fractions were collected into glass vials which had previously been flushed with dry prepurified nitrogen. The vials were flushed again with nitrogen after being filled, sealed in a flame, frozen and stored at -20°C until just prior to use.

Peptide samples (20-40 mµ moles) were placed in small pyrex tubes and dried over P_2O_5 and NaOH in vacuo. 50 µl of hydrazine was added to each sample, the tubes were flushed with nitrogen and sealed

over a flame. The samples were heated at 70°C for 48 hours, dried down at 60°C over P_2O_5 and NaOH and the free amino acids examined by high voltage electrophoresis on paper.

Protein samples (20-40 mμ moles) were placed in pyrex hydrolysis tubes (15 cm x 0.5 cm) and 200 μl of hydrazine was added to each sample. The tubes were evacuated and flushed with nitrogen 3 times, sealed in vacuo and incubated at 70°C for 48 hours. Samples were dried in vacuo over P_2O_5 and NaOH and then suspended in about 200 μl of 0.5 N-ammonium acetate titrated to pH 5.6 with concentrated ammonium hydroxide. These solutions were placed on an XE 64 column (8 cm x 0.5 cm) and eluted with the acetate buffer. Free amino acids passed through the column unhindered whereas the hydrazides were retained. The free amino acids were analyzed by high voltage electrophoresis on paper.

Preparative Peptide Maps: The peptide map technique already described was used to isolate small fragments from large peptides. Large peptides (100-500 mμ moles) were digested with appropriate enzymes (chymotrypsin or subtilisin), and the resulting fragments were separated on peptide maps as described in Chapter 2 except that 1) the original application was a larger spot (about 2½" in diameter), 2) the chromatographic dimension was developed for fifteen hours (to prevent the loss of small hydrophobic peptides), and 3) maps were dried at room temperature for all steps (this presumably improves the yield). Detection ninhydrin spray was used to locate peptides which were then cut out, eluted from paper with water and analyzed (Figure 3).

Amino Acid Analysis: Samples of peptides were sealed in small pyrex tubes with 50 µl of constant boiling HCl (5.7N) and hydrolyzed for 12-16 hours in a 105°C oven. The HCl was removed in vacuo at 60°C over P_2O_5 and NaOH. The samples were then analyzed by high voltage paper electrophoresis and/or column chromatography.

High voltage one dimensional paper electrophoresis (75) was used to screen the amino acid compositions of all peptides. This method can resolve all of the common amino acids found in protein hydrolysates. Hydrolysates were applied to Whatman 3MM chromatographic paper and run at 7800 volts (50v/cm) for 105 minutes at pH 1.67 in a formic acid buffer (water-formic acid 45:3 v/v). The papers were dipped in cadmium-ninhydrin stain and developed over night at 37°C. Amino acid spots were then cut out, eluted in 100% methanol and quantitated by spectrophotometric readings at 500 mu ± 10% accuracy. Permanent records were obtained by photographing the papers with a Polaroid camera, using transmitted light and a 500 mµ interference filter.

Stains

Detection Ninhydrin: This dilute ninhydrin stain was used to detect peptides which were to be eluted from paper for subsequent analysis. A solution was made up of 50 mg ninhydrin, 75 ml ethanol and 25 ml 2 N-acetic acid. The peptide papers were sprayed lightly 4 to 5 times and developed for 5 minutes at 80°C. This procedure destroys approximately 10-15% of the free amino groups present (α amino and lysine ε amino).

Cadmium Ninhydrin Stain: This stain was used for high voltage amino acid papers. A stock solution was made of 400 ml of water, 80 ml of concentrated acetic acid and 4 gm of cadmium acetate. Just prior to the dipping of each paper, 36 ml of stock solution was mixed with 300 ml of acetone and 3 gm of ninhydrin. Amino acid papers were dipped and dried at 37°C for 12 hours or 80°C for 8 minutes.

Peptide Bond Stain (76): This stain colors peptides with an intensity directly proportional to the size of the peptide (i.e. number of peptide bonds). The peptide paper, previously dried for 5 minutes at 80°C, was sprayed heavily with 2½% chlorox, air dried for 5 minutes, sprayed heavily with 95% ethanol and air dried again for 2½ minutes. It was then sprayed (or dipped) with a filtered solution made up of equal volumes of 2% acetic acid freshly saturated with O-tolidine and 0.05 M-KI. Blue spots develop rapidly at room temperature.

Mild Acid Hydrolysis (77): The following procedure (73) cleaves certain acid-labile peptide bonds (e.g. serine, glycine and threonine bonds cleave on the N-terminal side whereas aspartic acid and glycine cleave on the C-terminal side). 100 mμ moles of peptide was dissolved in 100 μl of 2 N-HCl and placed in boiling water for 5 minutes. The sample was dried down and the peptide fragments purified by flat plate electrophoresis in a pyridine acetate buffer at pH 3.6.

Aspartic acid residues were cleaved from peptides (proteins) by the following procedure (79): 100 mμ of peptide was dissolved in 400 μl of 0.1 N-acetic acid (pH 2.0). The sample was evacuated, sealed in a flame and hydrolyzed for 12 hours at 105°C.

Amide Determinations: This procedure was simplified by the fact that none of the peptides investigated had more than one acidic or amide group. Each peptide was electrophoresed at 3000 volts (50 v/cm) for 1½ hours at pH 6.5 (10% pyridine buffer titrated to pH 6.5 with acetic acid). Free amino acids were used as standards. Papers were developed with a collidine/ninhydrin stain. Neutral peptides migrated slightly toward the anode due to electroendosmosis whereas acidic peptides moved toward the cathode.

Peptide Sequencing: The procedure described in detail by Gray (80) was used. Briefly this consisted of 1) determining the peptide end group by the extremely sensitive dansyl chloride technique, 2) using the phenylisothiocyanate reaction (Edman procedure) to remove the amino terminal residue and 3) taking an aliquot of the degraded peptide for the dansyl chloride procedure. This process was repeated until the peptide was sequenced. The dansyl amino acids were separated by flat plate electrophoresis in various buffer systems and identified by inspection under ultraviolet light. A modified version of the dansyl-Edman's procedure permits one to complete four cycles in a single day (81).

N-Terminal Analysis of Intact Proteins (82,83): The phenylisothiocyanate (PITC) reaction was carried out in its modified three cycle form. Light chains were performic acid oxidized as native and alkylated proteins were much less soluble in the coupling buffer. Approximately 10 mg of protein was dissolved in 1.0 ml of coupling buffer (15.0 ml pyridine, 10.0 ml water and 1.18 ml dimethyallylamine) which was titrated to pH 9.0 with trifloroacetic acid. Insolubility

at this stage, particularly in later steps of degradation, prompted us to use 12 ml glass tubes with conical ground glass bottoms tapered to receive a glass homogenizer. Insoluble proteins could then be ground up and finely dispersed in the coupling buffer. Then 50 ml of redistilled phenylisothiocynate was added, the tubes flushed for a few seconds with nitrogen, stoppered with glass and incubated at 40°C for 1 hour. The mixture was extracted once with 4 ml of benzene and 4 times with 4 ml of butyl acetate, the organic (upper) phases being discarded. The final traces of butyl acetate were removed with a stream of nitrogen. About 0.5 ml of water was added, the sample shell frozen and evacuated over NaOH and P_2O_5 for 2 hours. The dried material was washed 3 times with 1 ml portions of ethyl acetate. Final traces of ethyl acetate were removed by gentle aspiration and the sample dried for 15 minutes.

The phenylthiocarbamyl derivatives were dissolved in 200 µl of trifloroacetic acid, gently flushed with nitrogen and incubated in glass stoppered tubes at 40°C for 15 minutes. This promotes a cyclization reaction which cleaves 2-anilino-5-thiazolinone derivatives from the peptide chain. Thiazolinones were then extracted with successive 2,2 and 1 ml portions of dichloroethane. The peptide chains were dried in vacuo and ready for a second round of coupling at 23°C. The thiazolinones were converted into the corresponding phenylthiohydantoins (PTH amino-acids) by adding 0.3 ml of conversion buffer (30% ethanol titrated to pH 1.0 with 0.15 M-HCl) and incubated at 80°C for 1 hour under nitrogen. The PTH amino acids were extracted with three 1 ml

portions of ethyl acetate, and the organic phase was dried down with a gentle stream of nitrogen at 40°C.

The PTH amino acids were dissolved in 50 µl dichloroethane, samples were removed and diluted with ethanol (10 µl in 1 ml) and spectra were recorded between 245 mu and 320 mu. The PTH derivatives have a maximum at 269 mµ which is used for yield calculations.

PTH amino acids were identified by thin layer chromatography on Eastman TLC fluorescent sheets using two solvent systems: a) m-xylene and b) heptaine-75% formic acid-dichloroethane (1:2:2) systems D and F or Sjoquist (84) . Just prior to chromotography in the D system, the thin layer sheets were treated with foramide-acetone (1:4) and dried briefly.

Identifications were confirmed by the dansyl chloride procedure. Aliquots of each sample were taken at the thiazolinone stage and converted to the free amino acid by 12 hour hydrolysis in 30 µl of 0.1 N-NaOH at 105°C in a sealed tube. The pH of the hydrolysate was reduced to about 8 by 3 minute exposure to a CO_2 atmosphere (dry ice). 30 µl of dansyl chloride (3 mg/ml) was added and the mixture incubated at 40°C for 90 minutes. The acetone was evaporated in a stream of nitrogen and the remaining solution extracted twice with 60 ml of ethyl acetate (water saturated). The pH of the mixture was lowered to about 4 with a citrate buffer and a final ethyl acetate extraction (100 µl) removed the dansyl amino acid. This extraction was dried, redissolved in 1 N-ammonia, and applied to paper and electrophoresed as described by Gray (80). Yields are very low with this procedure, but it is useful

for detecting certain PTH amino acids which are difficult to determine with chromatography (e.g. threonine, serine, and methonine sulfone).

Isolation of Large Amino Terminal Peptides: A procedure was developed for the isolation of large amino terminal peptides using the following rational: 1) α and ϵ amino groups were blocked by acetylation; 2) disulfide bridges were reduced and the resulting cysteines alkylated with iodoacetamide; 3) the protein was cleaved at arginines by trypsin; 4) arginine residues were removed from the C-terminus of all peptides with carboxypeptidase B (CP-B); and 5) the acidic amino terminal peptides were separated from all others by passage through Dowex 50 (H^+) column (only the amino terminal peptides are free of positive charge). The same procedure was used for proteins with a blocked α amino group except for omission of the acetylation step. With unalkylated proteins CP-B also removes C-terminal lysine. Pepsin or chymotrypsin was also used as an alternative to the trypsin-carboxypeptidase enzymatic digestion procedure.

1) Acetylation: 50 mg of light chain were dissolved in 2 ml of freshly deionized 10 M-urea, 3 ml of pyridine was added slowly, and the mixture was cooled to $4^{\circ}C$. Three aliquots of 100 ul of acetic anhydride were added to the sample at 5 minute intervals with vigorous stirring. This mixture was dialyzed against 0.2 N-ammonium bicarbonate for 12 hours at $4^{\circ}C$ and then against 10 M-urea-1 M-tris-HCl (pH 8.6) for 6 hours at $23^{\circ}C$. This procedure was carefully designed to keep the protein in solution or finely suspended.

2) Reduction and Alkylation: The solution was made 0.1 M in mercaptoethanol and placed at 37°C for 16 hours. A 5 times molar excess of iodoacetamide (over mercaptoethanol) was added, and the mixture reacted with vigorous stirring at 23°C for 15 minutes. The mixture was swamped with a 5 times molar excess of mercaptoethanol (over iodoacetamide) and dialyzed against repeated changes of 0.1 M-ammonium bicarbonate at 4°C. Most proteins were insoluble at this stage.

3) Enzyme digestion: 2% (w/w) trypsin was added directly to the dialysate and after a 4 hour reaction at 37°C, 2% (w/w) carboxypeptidase B was added to the same mixture and reacted for 1 hour at 23°C. The sample was lyophilized, redissolved in about 5 ml water and any precipitate spun down.

4) Column chromatography: The supernatant was added to a Dowex 50 (Technicon Peptide resin A) (H+) column (8 cm x 1 cm) and eluted with water under 10 lbs/in^2 nitrogen pressure. 10 ml of eluate was collected and lyophilized.

5) Paper electrophoresis: Aliquots were electrophoresed at pH 6.5 (10% pyridine acetate buffer) on the flat plate to determine homogeneity and charge of the isolated peptide (s).

Since it would appear that this procedure might be very generally applied to the screening of large numbers of protein samples, cautionary comments are in order. We appeared to have been extremely fortunate in that the CP-B used had very little carboxypeptidase A activity. All of the CP-B used in these experiments came from a single Worthington batch. It would be advisable to test CP-B for carboxypeptidase A contamination

before setting out to screen large numbers of proteins.

We might also point out that this procedure has been carefully designed to keep the protein "in solution" or at least finely suspended throughout. Lyophilization of these proteins after acetylation or after alkylation and prior to enzymatic digestion generally reduced the yield significantly, for such protein tended to remain as clumps during enzymatic digestion.

One should be careful not to collect more than the "break through" volume of the Dowex 50 column, for in some cases, ninhydrin positive peptides do start trailing through the column soon after the blocked peptides are eluted. This calibration can be done by spectral readings at 215 mμ (the peptide bond absorbs here).

Occasional samples precipitated out on the surface of the Dowex 50 resin, forming a thick hard shell which blocked further column elution even under very high nitrogen pressures. Stirring the surface of the resin generally broke up this shell and permitted normal elution.

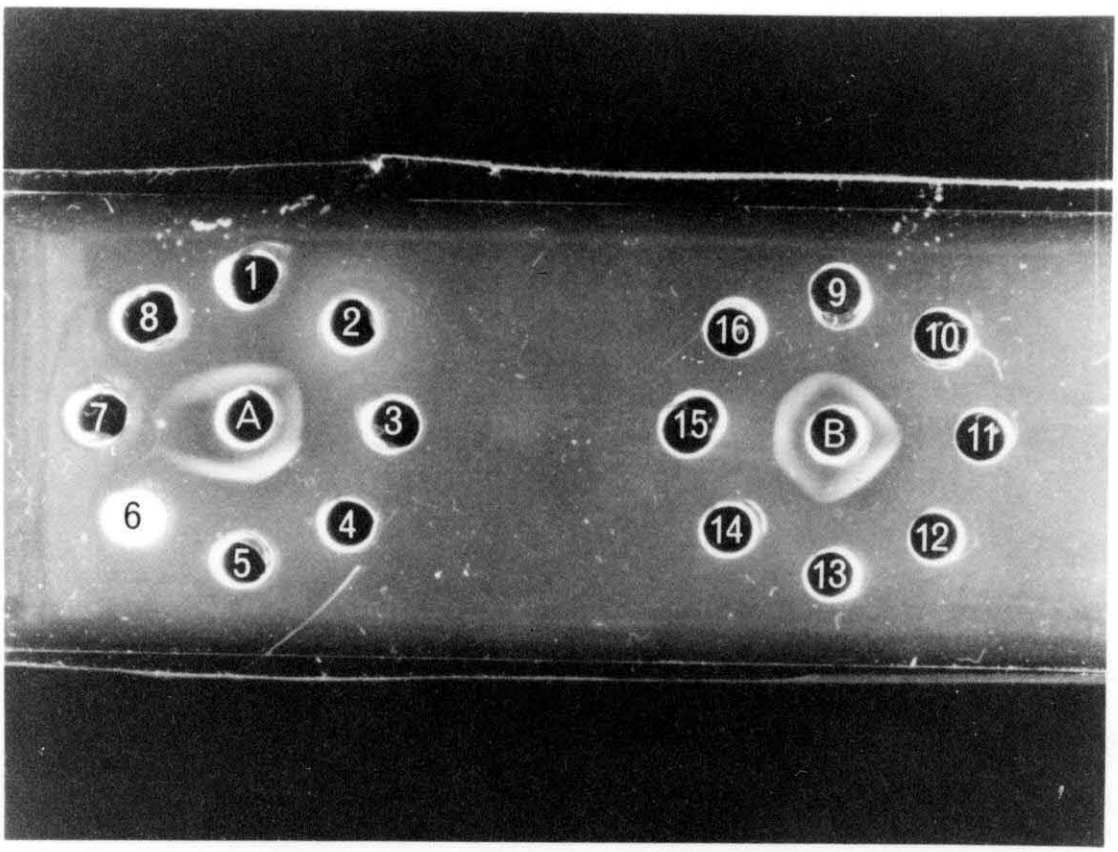
RESULTS

Immunodiffusion: The anti-light chain-sera divided all of the human proteins studied into either the K or the L class (Figure 1). This was also true of the intact human IgG proteins. These results can be summarized as follows:

K Class: HBJ's 1,3,4,5,6,10,12,HS4 and HS6

L Class: HBJ's 2,7,8,9,11, and 15

Figure 1. Immunodiffusion: reaction of human light chains against rabbit anti-kappa serum in A and B wells. Myeloma light chains are indicated by well number, that is HBJ 1, HBJ 2, HBJ 3, ... HBJ 8 are distributed clockwise around the A well. The same is true of the B well except 14 is HS 4, 15 is HS 6 and 16 is normal pooled human light chain.


1
8
2
7
A
3
6
4
5
9
16
10
15
B
11
14
12
13

No spur formation was noted with either antiserum.

PITC Procedure on Intact Light Chains: This procedure gave unambiguous results out to 6 steps with most of the light chains analyzed (Table 1). In each case the results obtained from the silica thin layer chromatograms were confirmed by the dansyl procedure.

The PTH residues and residue yields are given at each stage of degradation in Table 1. The amount of PTH amino acid can be calculated from spectral readings using an average molar extinction coefficient of 16,000 at 269 mμ. The yield in step 1 is calculated by comparing mμ moles PTH amino acid recovered to mμ moles of protein reacted initially (based on protein dry weight). Yields for subsequent steps are expressed as percentages of the PTH derivate found in the preceeding step.

An examination of the yields in Table 1 underscores two important considerations:

1) Protein must be soluble or at least finely dispersed in the coupling buffer in order to get adequate coupling. For example, HBJ 4, although soluble at step 1, became progressively more insoluble through steps 2 and 3. The yields dropped accordingly (Table 1). On the other hand, HBJ 6, although extremely insoluble in the coupling buffer, was finely dispersed with the glass homogenizer at every step after the 1st and, accordingly, gave reasonable yields.

2) Yields of greater than 100% which occur one step after threonine or serine (see step 6 - Table 1) probably reflect the destruction of these labile PTH derivatives (i.e. conversion to the dehydro form).

TABLE 1

Amino Terminal Sequences and Yields of K Light Chains from PITC Procedure

Source of light chains		Residue Number 1	2	3	4	5	6	% Yield+ 1	2	3	4	5	6
Mouse Bence-Jones proteins	MBJ 41	Asp	Ilu	Gln	Met	Thr	Gln	(85,	75,	90,	130,	45,	110)
	MBJ 70	Asp	Ilu	Val	Leu	Thr	Gln	(70,	90,	100,	90,	60,	80)
	MBJ 6	Asp	Ilu	Val	Val	Thr	Gln	(65,	100,	110,	95,	75,	110)
Human Bence-Jones proteins	HBJ 3	Asp	Ilu	Val	Leu	Thr	Gln	(80,	95,	80,	100,	80,	65)
	HBJ 12	Glu	Ilu	Val	Val	Thr	Gln	(100,	70,	110,	95,	65,	100)
	HBJ 10	Asp	Ilu	Gln	Met	Thr	Gln	(100,	70,	110,	90,	80,	125)
	HBJ 6	Asp	Ilu	Gln	Met	Thr	Gln	(35,	70,	100,	100,	55,	160)
	HBJ 1	Asp	Ilu	Leu	Met	Thr	Gln	(100,	90,	85,	90,	105,	110)
	HBJ 5	Glu	Ilu	Val	Leu			(95,	80,	105,	50)		
	HBJ 4	Asp	Ilu	?				(100,	65,	15)			
	Ag*	Asp	Ilu	Gln	Met	Thr	Gln						
Human 7S myeloma proteins	HS 4	Glu	Ilu	Val	Leu	Thr	Gln	(75,	90,	75,	115,	60,	100)
	HS 6	Glu	Ilu	Val	Leu	Thr	Gln	(60,	105,	90,	90,	65,	105)
Pooled γ-globulin (normal human)		Asp + Glu	Ilu	Val + Gln	Met + Leu Val	Thr		(50)					

*Taken from Putnam et al. (94).
+Yield calculations are described in text.

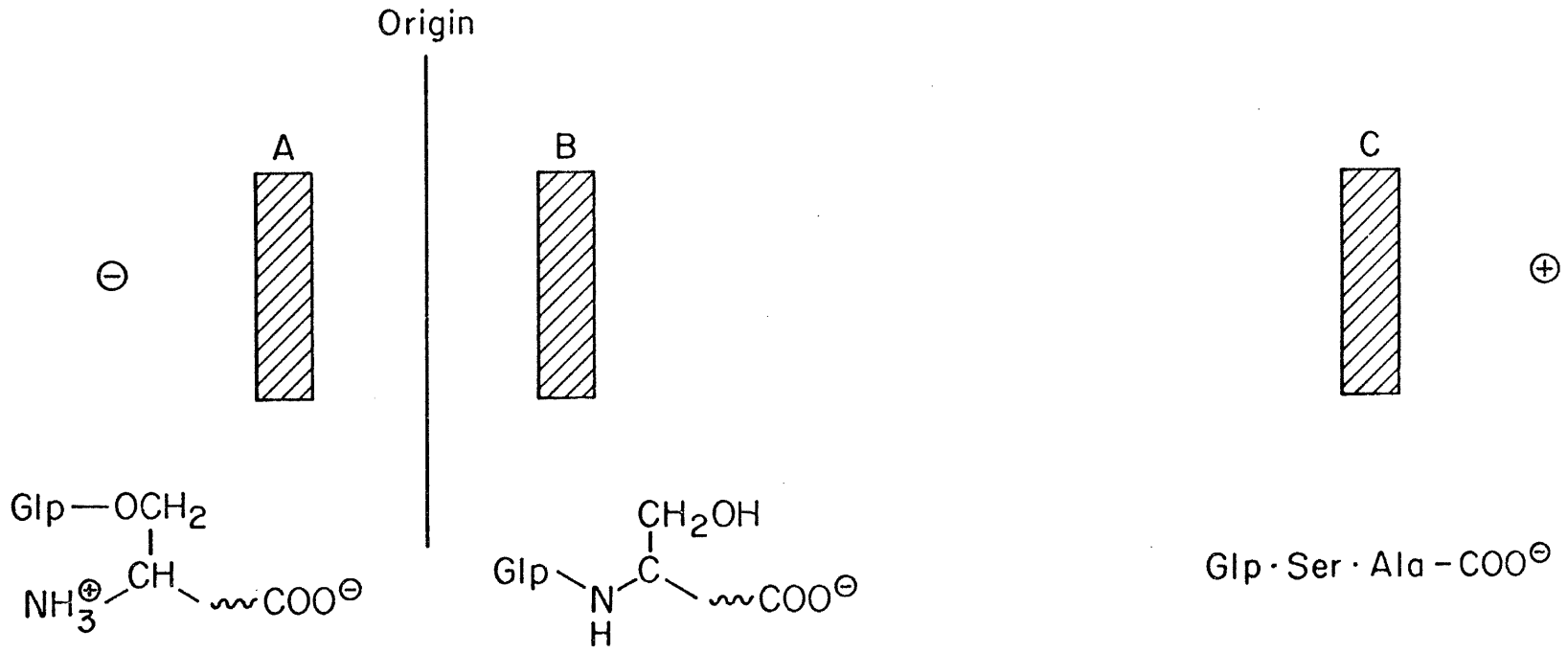
For example, loss of PTH threonine at a particular step would give a low spectral reading, yet the amount of protein capable of yielding PTH derivatives in subsequent steps is unchanged.

Five light chains failed to yield any end groups with the PITC procedure (HBJ's 2,7,8,11, and 15). A second PITC step gave multiple PTH derivatives (at least ten different PTH amino acids in comparable yields), indicating extensive nonspecific hydrolysis. Perhaps this indicates acid cleavage of C-terminal glycine and aspartic acid bonds (N-terminal cleavage should give high concentrations of the residues at which cleavage occurred). These residues are particularly acid labile although it is difficult to explain why similar hydrolysis didn't happen with the unblocked proteins. A sixth light chain (HBJ 9) gave a strong tyrosine, a somewhat weaker glycine and a much lighter serine, threonine and leucine at step 1. It behaved similar to the "blocked" proteins at step 2.

N-terminal Peptides: Terminal peptides were isolated from L and K proteins by the Dowex 50 (H+) column procedure (see methods section). Amino terminal peptides from both classes were generally ninhydrin negative, peptide bond spray positive, and unreactive to the dansyl chloride procedure (i.e. there were no free amino groups). Occasionally ninhydrin positive contaminants were eliminated by a second passage through a regenerated Dowex 50 (H+) column.

When amino terminal peptides from pepsin-digested L proteins were run electrophoretically at pH 6.5, three peptides were generally observed (Figure 2). The most acidic, present in low yield, was the

Figure 2. Paper electrophoresis (pH 6.5) of amino terminal peptides from a pepsin digest of HBJ 2 run at 3000 volts for 1 hour at 23°C. See test for explanation.


Origin
A
B
C
⊖
⊕
Glp—OCH2
NH3⊕
CH
COO⊖
CH2OH
Glp
N
H
C
COO⊖
Glp · Ser · Ala – COO⊖

amino terminal tripeptide (Figures 3 and 4). The other two peptides, one neutral and the other slightly acidic, had identical amino acid compositions. One would expect the blocked N-terminal peptide to be slightly acidic. It is possible that an acyl shift has occurred between a hydroxy amino acid and an adjacent residue to produce an additional positive charge in some of the N-terminal molecules thereby generating two populations of N-terminal peptides (with identical amino acid sequences), one acidic (unshifted), the other neutral (shifted). This is illustrated in Figure 2. Such a shift is acid catalysed and may have occurred during the Dowex 50 (H+) column separation. This reaction reverses itself slowly under neutral conditions(85), and it is probably that the neutral electrophoretic buffer did not have time to reconvert all of the "acyl shifted" peptide to its original form. Many of the acetylated K N-terminal peptides also demonstrate this same phenomenon.

Acid hydrolysates of most amino terminal peptides were examined quantitatively on the amino acid analyzer (Table 2). In all cases the compositions were compatible with the sequence data.

The strategy used to sequence the amino terminal portion of HBJ 2 will be discussed in detail as most of the other amino terminal peptides were characterized in an analogous manner. Initially pepsin was used to produce the amino terminal peptide, and the yields of the three preparations after passage through the Dowex 50 (H+) column ranged between 65 and 80%. Subtilisin and chymotryptic fragments were isolated from this peptide using preparative peptide map techniques, and here the

Table 2

Amino Acid Compositions of N-terminal K and L peptides[+]

Amino Acids	HBJ 2	HBJ 7	HBJ 8	HBJ 11	HBJ 15	HBJ 4	HBJ 5	HBJ 10
Cysteine ‡	0.8[‡]	0.7[‡]	-	-	-	-	-	-
Aspartic Acid	0.4	0.2	-	-	-	1.1	0.8	2.1
Threonine	3.6	2.7	0.8	2.0	2.5	0.9	2.0	1.0
Serine	5.6	5.5	5.0	3.0	3.9	2.8	3.0	5.3
Glutamic Acid	3.4	3.2	3.6	3.2	3.0	2.3	2.9	2.2
Proline	3.2	3.1	2.2	3.0	2.0	1.0	2.3	1.0
Glycine	3.1	3.9	2.0	2.0	2.2	-	0.8	1.0
Alanine	2.1	1.2	3.3	1.2	2.3	-	-	1.0
Valine	1.0	2.1	0.9	1.0	1.1	-	0.9	1.0
Methionine	-	-	-	-	-	0.8	-	1.2
Isoleucine	0.9	0.8	0.7	-	1.0	-	0.8	1.0
Leucine	1.0	1.1	1.0	1.0	1.1	1.0	3.3	1.1
M U M Peptide	50	32	10	27	25	18	25	12

[+]Analysed by column chromatography

‡About 1:1 mixtures of cysteine and carboxymethylcysteine

yields ranged between 15 and 30%. An illustration of the preparative peptide map for HBJ 2 is shown in Figure 3. The peptide fragments obtained from the subtilisin and chymotryptic digests were sequenced using the dansyl-Edman procedure and carboxypeptidase A (Figure 4).

Sequencing the blocked amino terminal peptide fragment presented a problem since the dansyl-Edman procedure could not be used. A chymotryptic digest of the entire light chain yielded a blocked tetrapeptide corresponding in composition to the blocked subtilisin fragment of the pepsin amino terminal peptide (Figure 4). When the chymotryptic peptide was treated with carboxypeptidase A, leucine and then valine were released in that order. The resulting amino terminal dipeptide could be separated from the free amino acids by passage through a Dowex 50 (H^+) column. This dipeptide (glu,ser) was shown by hydrazinolysis to have a C-terminal serine. The nature of the blocked glutamic acid residue was not investigated, but it was presumed by analogy to be a pyrrolidone carboxylic acid since this group occurs in the amino terminal positions of human and rabbit heavy chains (86). Hence the amino terminal blocked peptide has a sequence of Glp-Ser-Val-Leu as shown in Figure 4.

A kinetic study using carboxypeptidase A on the intact pepsin fragment gave the following results:

Time (minutes):	3	9	27	60
mμ mole threonine:	4	7	16	23
mμ mole valine:	0	2	11	22

Clearly the C-terminal sequence of the peptic fragment is -Val-Thr (Figure 4).

Figure 3. Preparative peptide map separation of a subtilisin digest of the pepsin N-terminal fragment from HBJ 2. This operation is described in the text.

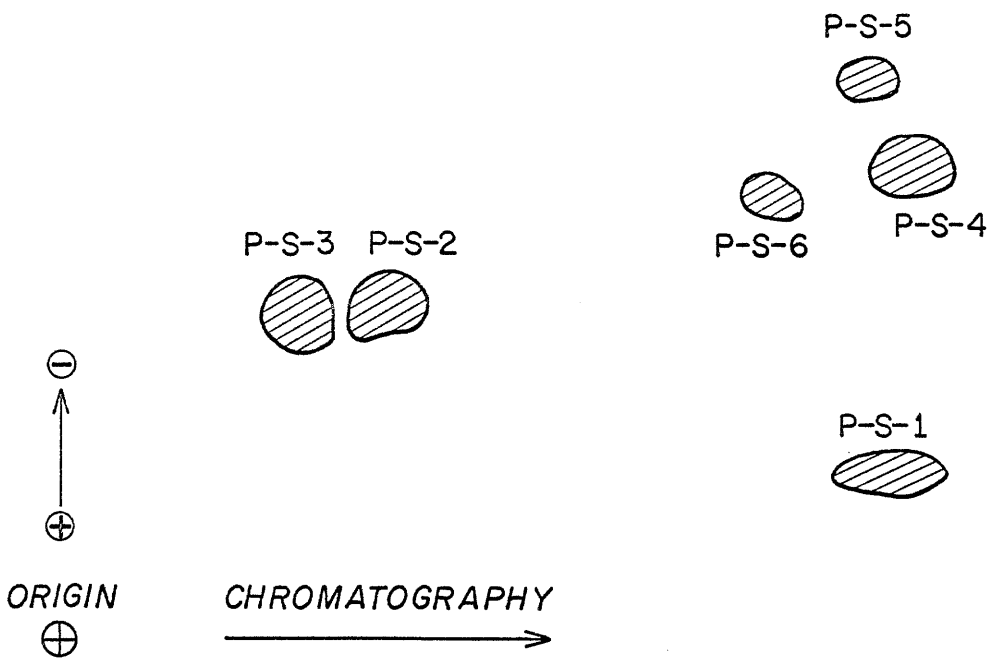

P-S-5
P-S-4
P-S-6
P-S-3
P-S-2
P-S-1
ORIGIN
CHROMATOGRAPHY

Figure 4. Determination of the structure of the blocked trypsin carboxypeptidase B (T-CPB) fragment from human L chain HBJ 2. Peptides designated T-CPB, pepsin and chymotrypsin are those isolated from digests of the whole protein; those designated P-subtilisin 1, etc., are fragments of the original pepsin peptide produced by further digestion with a second enzyme. Sequences established by dansyl-Edman procedure are indicated $\rightharpoonup$; sequences established by carboxypeptidase A or hydrazinolysis are indicated $\leftharpoondown$. "Glp" indicates a blocked glutamic acid residue, presumably a pyrrolidone carboxylic acid.

RESIDUE NUMBER

Enzyme	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
T-CPB	(Glp	Ser	Ala	Leu	Thr	Gln	Pro	Pro	Ser	Ala	Ser	Gly	Ser	Pro	Gly	Gln	Ser	Val	Thr	Ile	Ser	Cys	Thr	Gly	Thr)
Pepsin-1	(Glp	Ser	Ala	Leu	Thr	Gln	Pro	Pro	Ser	Ala	Ser	Gly	Ser	Pro	Gly	Gln	Ser)	Val	.	Thr	.				
Pepsin-2	(Glp	Ser	Ala)																						
Chymotrypsin	(Glp	Ser)	Ala	.	Leu																				
C-CPA-Hyd	Glp	.	Ser																						
C-acid		(Ser	Ala	Leu)																					
P-subtilisin 1	(Glp	Ser	Ala	Leu)																					
P-subtilisin 2					Thr	Gln	Pro	Pro	Ser	Ala	Ser														
P-subtilisin 3												Gly	Ser	Pro	Gly	Gln	Ser								
P-subtilisin 4																		Val	Thr						
P subtilisin 5					Thr	Gln	Pro	Pro																	
P subtilisin 6																	Ser	Val	Thr						
P chymotrypsin 1	(Glp	Ser	Ala	Leu)																					
P chymotrypsin 2					Thr	Gln	Pro	Pro	Ser	Ala	Ser	Gly	(Ser	Pro	Gly)	Gln	Ser	Val	Thr						
T-CPB-pepsin 1				Leu	Thr	Gln	Pro	Pro	(Ser	Ala	Ser	Gly	Ser	Pro	Gly	Gln	Ser	Val	Thr)						
T-CPB-pepsin 2																				Ile	Ser	Cys	Thr	Gly	Thr
Sequence	Glp	Ser	Ala	Leu	Thr	Gln	Pro	Pro	Ser	Ala	Ser	Gly	Ser	Pro	Gly	Gln	Ser	Val	Thr	Ile	Ser	Cys	Thr	Gly	Thr

A trypsin-carboxypeptidase B amino terminal peptide was also prepared and fragmented with pepsin and chymotrypsin. These fragments were purified and sequenced as described above. This procedure gave a completely unambiguous sequence for the first 25 residues in this L chain (Figure 4).

A similar approach (Figure 5) was taken to determine unambiguous sequences for all the amino terminal peptides from HBJ 1, HBJ 5, HBJ 7, HBJ 8, HBJ 10, HBJ 11, HBJ 15 and HS 4. It should be pointed out, however, that the yields of K N-terminal peptides were much lower than their L counterparts, ranging between 15 and 40%. It was, of course, unnecessary to determine the sequence of residues 1-6 for the K class proteins as these had been determined using the PITC procedure on intact protein.

The amino terminal peptide of HBJ 4 was prepared using chymotrypsin. The specific aspartic acid cleavage procedure was used to remove the acetylated N-terminal aspartic acid, leaving the remainder of this chymotryptic fragment intact. Electrophoresis of this peptide at pH 4.4 gave a neutral, ninhydrin positive peptide plus aspartic acid. The peptide was sequenced using the dansyl-Edman procedure (Figure 6).

Sequence Comparisons: Although most proteins which have been sequenced for at least 18 amino terminal residues are individually unique, the variation within a light chain class at a given residue is extremely limited (Figure 6). Generally, only 1, 2 or occasionally 3 residues have been found at any position.

Figure 5. Sequence determinations of blocked L and K peptides. +K proteins. ^{o}L proteins. All acetylated N-terminal K peptides were fragmented with subtilisin except HBJ 4 (see test). All blocked L peptides were sequenced in a manner analogous to HBJ 2 (see text). HBJ 11 was acetylated before TCPB digestion, hence the Arg can be assigned to the C-terminal position. Enzymatic fragments with established amino acid compositions are designated ⟨——⟩; sequences established by dansyl-Edman are indicated ⟨~~~⟩ ; sequences established by PITC procedure are indicated ——— ; sequences established by carboxypeptidase A and hydrazinolysis are indicated ⟨~~⟩ .

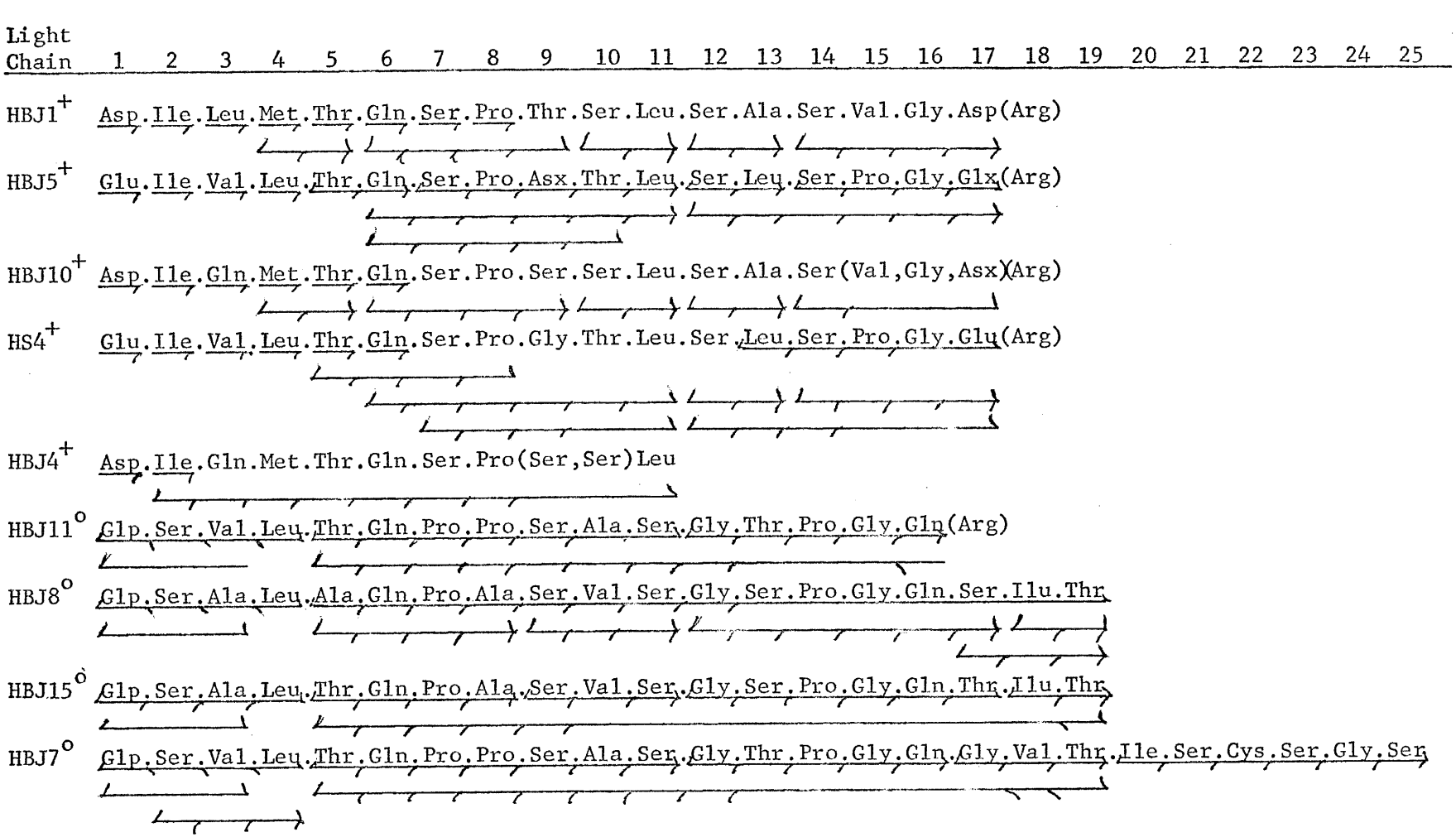

RESIDUE NUMBER
Light Chain
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
HBJ1+ Asp.Ile.Leu.Met.Thr.Gln.Ser.Pro.Thr.Ser.Leu.Ser.Ala.Ser.Val.Gly.Asp(Arg)
HBJ5+ Glu.Ile.Val.Leu.Thr.Gln.Ser.Pro.Asx.Thr.Leu.Ser.Leu.Ser.Pro.Gly.Glx(Arg)
HBJ10+ Asp.Ile.Gln.Met.Thr.Gln.Ser.Pro.Ser.Ser.Leu.Ser.Ala.Ser(Val,Gly,Asx)(Arg)
HS4+ Glu.Ile.Val.Leu.Thr.Gln.Ser.Pro.Gly.Thr.Leu.Ser.Leu.Ser.Pro.Gly.Glu(Arg)
HBJ4+ Asp.Ile.Gln.Met.Thr.Gln.Ser.Pro(Ser,Ser)Leu
HBJ11° Glp.Ser.Val.Leu.Thr.Gln.Pro.Pro.Ser.Ala.Ser.Gly.Thr.Pro.Gly.Gln(Arg)
HBJ8° Glp.Ser.Ala.Leu.Ala.Gln.Pro.Ala.Ser.Val.Ser.Gly.Ser.Pro.Gly.Gln.Ser.Ilu.Thr
HBJ15° Glp.Ser.Ala.Leu.Thr.Gln.Pro.Ala.Ser.Val.Ser.Gly.Ser.Pro.Gly.Gln.Thr.Ilu.Thr
HBJ7° Glp.Ser.Val.Leu.Thr.Gln.Pro.Pro.Ser.Ala.Ser.Gly.Thr.Pro.Gly.Gln.Gly.Val.Thr.Ile.Ser.Cys.Ser.Gly.Ser

Figure 6. Amino terminal sequences of myeloma light chains. Data taken from: *Hood et al. (91), Gray et al. (92); ≠ Wikler et al. (93), Putnam et al. (94); + Milstein (95); ** Hilschmann and Craig (67).

AMINO TERMINAL LIGHT CHAIN SEQUENCES

		5	10	15	20	25
SL	HBJ 11*	Glp-Ser-Val-Leu-Thr-Gln-Pro-Pro-Ser-( )-Ala-Ser-Gly-Thr-Pro-Gly-Gln(Arg)				
	HBJ 8*	Glp-Ser-Ala-Leu-Ala-Gln-Pro-Ala-Ser-( )-Val-Ser-Gly-Ser-Pro-Gly-Gln-Ser-Ile-Thr				
	HBJ 15*	Glp-Ser-Ala-Leu-Thr-Gln-Pro-Ala-Ser-( )-Val-Ser-Gly-Ser-Pro-Gly-Gln-Thr-Ile-Thr				
	HBJ 2*	Glp-Ser-Ala-Leu-Thr-Gln-Pro-Pro-Ser-( )-Ala-Ser-Gly-Ser-Pro-Gly-Gln-Ser-Val-Thr-Ile-Ser-Cys-Thr-Gly-Thr				
	HBJ 7*	Glp-Ser-Val-Leu-Thr-Gln-Pro-Pro-Ser-( )-Ala-Ser-Gly-Thr-Pro-Gly-Gln-Gly-Val-Thr-Ile-Ser-Cys-Ser-Gly-Ser				
	SH ‡	( )Ser-Glu-Leu-Thr-Gln-Asp-Pro-Ala-( )-Val-Ser-Val-Ala-Leu-Gly-Gln-Thr-Val-Arg-Ile-Thr-Cys-Gln-Gly-Asp				
		5	10	15	20	25
SK_I	MBJ 41*	Asp-Ile-Gln-Met-Thr-Gln-Ser-Pro-Ser- Ser -Leu-Ser-Ala-Ser-Leu-Gly-Glu-Arg-Val-Ser-Leu-Thr-Cys-Arg-Ala-Ser				
	Ag ‡	Asp-Ile-Gln-Met-Thr-Gln-Pro-Ser-Ser- Ser -Leu-Ser-Ala-Ser-Val-Gly-Asp-Arg-Val-Thr-Ile-Thr-Cys-Gln-Ala-Ser				
	Ker+	Asx(Ile,Glx)Met(Thr,Gln,Ser,Pro,Ser, Ser ,Leu)Ser-Ala-Ser-Val-Gly-Asp-Arg-Ile-Thr-Ile-Thr-Cys-Gln-Ala-Ser				
	BJ+**	Asx(Val,Glx)Met-Thr-Gln(Ser,Pro,Ser, Ser ,Leu)Ser,Ala,Ser,Val,Gly,Asp(Arg-Val-Thr-Ile-Thr-Cys-Gln-Ala-Ser				
	Roy	Asp-Ile-Gln-Met-Thr-Gln-Ser-Pro-Ser- Ser -Leu-Ser(Ala,Ser,Val,Gly)Asx-Arg(Ile,Thr,Ile,Thr,Cys,Glx,Ala,Ser				
	HBJ 10*	Asp-Ile-Gln-Met-Thr-Gln-Ser-Pro-Ser- Ser -Leu-Ser-Ala-Ser(Val,Gly,Asx)(Arg)				
	HBJ 1*	Asp-Ile-Leu-Met-Thr-Gln-Ser-Pro-Thr- Ser -Leu-Ser-Ala-Ser-Val-Gly-Asp(Arg)				
	HBJ 4*	Asp-Ile-Gln-Met-Thr-Gln-Ser-Pro(Ser, Ser)Leu				
		5	10	15	20	25
SK_{II}	MBJ 70*	Asp-Ile-Val-Leu-Thr-Gln-Ser-Pro-Ala- Ser -Leu-Ala-Val-Ser-Leu-Gly-Gln-Arg-Ala-Thr-Ile-Ser-Cys-Arg-Ala-Ser				
	HBJ 3*	Asp-Ile-Val-Leu-Thr-Gln-Ser-Pro-Leu- Ser -Leu-Pro-Val-Thr-Pro-Gly-Glu-Pro-Ala-Ser-Ile-Ser-Cys-Arg-Ser-Ser				
	HBJ 5*	Glu-Ile-Val-Leu-Thr-Gln-Ser-Pro-Asx- Thr -Leu-Ser-Leu-Ser-Pro-Sly-Glx(Arg)				
	HS 4*	Glu-Ile-Val-Leu-Thr-Gln-Ser-Pro-Gly- Thr -Leu-Ser-Leu-Ser-Pro-Gly-Glu(Arg)				
	MBJ 6*	Asp-Ile-Val-Val-Thr-Gln				
	HBJ 12*	Glu-Ile-Val-Val-Thr-Gln				

The human SK sequences can be divided into two subclasses, SK_I and SK_{II}, on the basis of "distinguishing" residues at certain positions (e.g. 3, 4, 9, 13, etc.) (Figure 6). The SK and the SL sequences can be distinguished by the presence of a deletion at SL (10) and "distinguishing" residues at certain positions (e.g. 1, 2, 10, etc.). Although each of these sets (i.e. SL, SK_I and SK_{II}) can be distinguished by certain features, they are also alike at many positions, suggesting a common evolutionary origin.

The mouse SK light chains can be divided into subclasses (SK_I and SK_{II}) analogous to the human proteins. Indeed, it is difficult to distinguish the mouse regions of these proteins from their human counterparts.

The observations on pooled human light chains are in complete accord with the findings among the myeloma proteins (Table 1). Thus, if all the homogeneous myeloma K light chains derived from the tumors were mixed together and a "pooled analysis" done, the results would be essentially identical to those shown for normal human light chains.

One important structural difference between K and most L chains should be stressed. Since the PITC reaction depends on the presence of a free α amino group, all blocked light chains presumably lack such a group. The PITC procedure is, therefore, capable of dividing all light chains into two groups, the blocked and the unblocked proteins. These groups correspond respectively to the L and the K classes as determined by immunologic analysis with a single exception (HBJ 9). It appears that the light chain class can be determined by chemical as well as immunologic procedures.

The blocked L N-terminal residue is a glutamic acid derivative, presumably pyrrolidone carboxylic acid by analogy with blocked heavy chain proteins (86).

DISCUSSION

Amino terminal sequences from 20 different light chains are given in Figure 6 and show that all light chains examined have unique S sequences. The mechanism responsible for antibody variability can obviously generate large numbers of different S sequences. A critical point in explaining the molecular and genetic basis for antibody diversity is whether the many different S genes could have arisen from a single gene by some special mutational process during somatic differentiation, or whether they are separate germ line genes which arose during the long course of evolution. If the S genes arose by a particular somatic mutational mechanism, there must be patterns reflecting that specific process imprinted on these proteins. On the other hand, if the S genes arose by normal chemical evolution, their phylogeny as determined by molecular level studies should be similar to that of evolutionarily related proteins such as cytochrome C or the hemoglobins.

The experimental evidence available at this time permits us to conclude with a high level of confidence that:

1. <u>There is no known single mutational or recombinational mechanism which can generate the observed sequences from a single germline gene.</u>

2. The pattern of amino acid variation is similar in all respects to that found in sets of evolutionarily related proteins.

3. At least two separate germ-line genes are responsible for encoding the S region of K light chains.

4. The common regions of human kappa chains are encloded by a single germ-line gene which segregates as expected for diploid organisms.

In our examination of the available date we find a completely straightforward phylogeny relating all of the light chains, which is analogous in all respects to phylogenetic trees for other proteins. There is certainly no evidence for any kind of hypermutational mechanism and no need to postulate any unusual amount of crossing-over between the genes involved. It is, of course, assumed that gene duplication and recombinational events do occur on a continuing basis as a major part of the organism's dynamic response to a changing environment.

Figure 7 shows five sets of protein sequences, two of hemoglobin chains (α and β) and three of light chains SK_I, SK_{II}, and SL; more extensive light chain data are shown in Figure 6. Each of these sets was aligned by maximizing amino acid sequence identity through the use of a minimum number of judiciously placed gaps.

The individual members of one hemoglobin set, α, can be distinguished from those of a related set, β, by the presence of different amino acid alternatives at given positions (e.g. 10, 12, 14 and 19) (Figure 7). The existence of these "distinguishing" features is seen even when much larger numbers of hemoglobin chains are compared; for

Figure 7. Characteristic sets of related proteins - amino terminal regions from hemoglobins and light chains. The one-letter code for amino acids is that used by Eck and Dayhoff in "Atlas of Protein Sequence and Structure", published by the National Biomedical Research Foundation, Silver Spring, Maryland. * Hemoglobin sequences are taken from this same source. Q* represents pyrrolidone carboxylic acid. Sources of light chain sequences are given in Figure 6.

SEQUENCE COMPARISONS OF EVOLUTIONARILY RELATED PROTEIN SETS

			1 5 10 15 20 25 30 35 40
α Hb*	Human		V - L S P A D K T N V K A A W G K V G A H A G E O G A E A L E R M F L S F P T T K T O
	Mouse		V - L S G E D K S N I K A A W G K I G G H A G E O G A E A L E R M F A S F P T T K(T.O.
	Horse		V - L S A A D K T N V K A A W S K V G G H A G E O G A E A L E R M F L G F P T T K T O
β Hb*	Human		V H L T P E E K S A V T A L W G K V N - - V D E V G G E A L G R L L V V O P W T Q F R
	Mouse		(V.H.L.T.A.E.B)K(S.A.V.G.A.L.W.G)K(V,N - -.V,D.E.A,G.G.E.A.L.G)R(L.L.V.V.O.P.W.T.Q)R(O
	Horse		V Q L S G E E K A A(V.L)A L W D K V N - - E E E V G G E A L G R L L V V O P W T Q R F
SL	Human	Sh	- S E L T Q D P A - V S V A L G Q T V R I T C Q G D S L R G - - - - O D A A W O Q Q K
	Human	2	Q* S A L T Q P P S - A S G S P G Q S V T I S C T G T
	Human	7	Q* S V L T Q P P S - A S G T P G Q G V T I S C S G S
SK_I	Mouse	41	D I Q M T Q S P S S L S A S L G E R V S L T C R A S Q D I G - - - - S L S N.W L Q.Q.G
	Human	Ag	D I Q M T Q P S S S L S A S V G D R V T I T C Q A S Q(D.I.B.-,-,-,-,S.F)L N W O Q Q G
	Human	Roy	D(I.Q.M.T.Q.S.P.S.S)L S(A.S.V.G)D.R(V.S.I.T.C.Q.A.S.Q.D.I.I.-,-,-,-,S.F)L(N.W.O.Q.Q.G
SK_{II}	Mouse	70	D I V L T Q S P A S L A V S L G Q R/A T I S C R/A S E S V B B S G I S F M D.W.F.Q)Q.K
	Human	Mil	D I V L T Q S P L S L P V T P G E P A S I S C R/S S Q N L L Z S B G(B)O L D W O L Q.K
	Human	HS4	E I V L T Q S P G T L S L S P G E(R)

example, the linked -Ala-Gly- sequence at positions 22-23 is found in all α chains sequenced to date and in no β chains. These two different sets of hemoglobin chains are clearly encoded by separate germ-line genes (α and β), and to argue that a single gene was responsible would require a mutation mechanism which could generate two discrete sets of proteins with their correlated gaps and "distinguishing" amino acid residues. The presence of the same amino acids at certain positions in both sets does not imply crossing-over between them but suggests that the amino acid was present in a common ancestor and performs a common function in the molecules.

Although less information is available on light chains, they can, by the above criteria, be divided into a minimum of three sets, SL, SK_I, and SK_{II}. For example, the subclasses of the SK region can be identified by a gap (SK (31-34)) and by the presence of "distinguishing" amino acid alternatives at given positions (e.g. compare SK_I and SK_{II} at residues 3, 4, 9, 13, etc.). Even from the limited data available we can identify at least twenty distinguishing positions throughout the SK region. Some of these show a tendency towards forming further subclasses (for instance positions 1, 10, 12, 13 within the SK_{II} regions). In evolutionary terms these classes and subclasses represent successive branches of the phylogenetic tree while individual proteins represent the terminal twigs. It is assumed that the number of known branches will grow as more and more precise molecular information becomes available. Note the SK_I and SK_{II} proteins in mouse, as well as man, can be distinguished by these criteria. These differences are illustrated

diagramatically in Figure 8. Similar criteria can be used in comparing the SL with the two SK sets of light chains. The SK regions observed in chicken and rabbit with an extra cystine bridge are assumed to comprise a third set of SK sequences which are designated SK_{III}. (These data will be presented in Chapter 4.) There are in each case significant structural features which distinguish members of one protein set from those of other related sets.

To argue that a single germ-line gene could generate both SK_I and SK_{II} sequences would necessitate postulating a mechanism which could, from a single nucleotide sequence, produce two discrete sets of proteins which differ in size. Correlated with the size difference, it would also have to produce coupled amino acid changes at many different positions. We find it impossible to imagine a single mutational mechanism which could generate two or more different "sets" of SK proteins. Rather, it is much simpler to assume that <u>each set of light chains must, as with the hemoglobin example, be encoded by a separate germ-line gene (or genes).</u> Since the sets of light chains share many common features with one another (e.g. compare SK_I and SL at positions 5, 6, 8, 12, etc.), the genes encoding each related protein must have diverged from a common ancestral gene.

The members of a given set are rather closely related to one another; for example, in comparing two members of SK_I over the entire specificity region, they differ in about 15% of their positions. The same is true of comparing two members of the SK_{II} subclass (Table 3). The differences between members of two sets, however, are much greater

Figure 8. Correlated changes in specificity regions of K chains. Squares and triangles indicate positions which are representative of one or other subclass. Many of these are found in both mice and humans, but it should be stressed that this figure is diagramatic, and should not be taken as a direct representation of four actual sequences. The open and closed rectangles indicate the absence or presence of four additional amino acids (96).

DIAGRAMATIC REPRESENTATION OF SK_I AND SK_{II} SEQUENCES

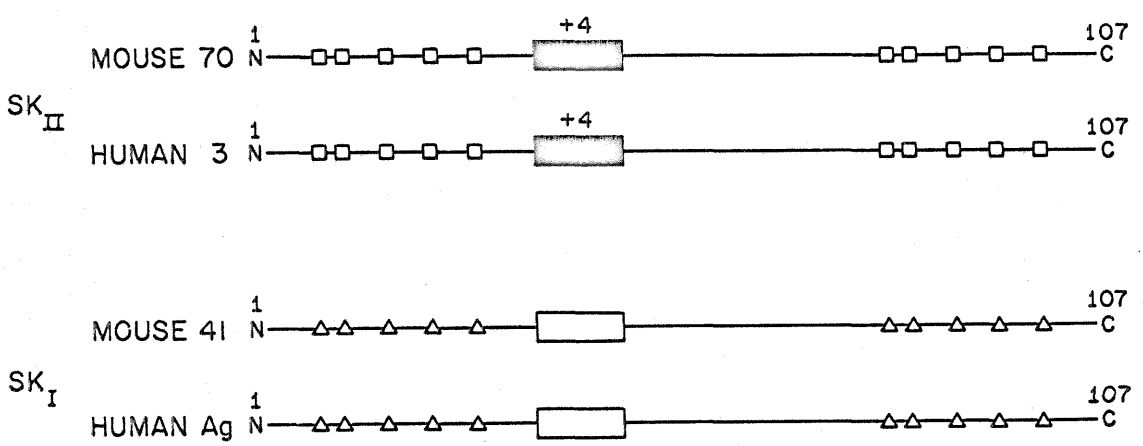


Table 3. Divergence of Various Light Chain Regions. Figures given represent approximate percentage changes in the amino acid sequences indicated. Sources of data are cited in Figure 6.

Table 3

Divergence of Light Chains

Between Classes	SK	vs	SL	60%	3 gaps
	CK	vs	CL	60%	3
Between Sub-classes	SK_I	vs	SK_{II}	40%	1
	CK_I	vs	CK_{II}	0	0
Within Sub-classes	SK_IAg	vs	SK_IRoy	15%	0
	SK_{II}HBJ 3	vs	SK_{II}Cum	12%	0
Between Species (Mouse - Human)	SK	vs	SK	40%	0
	CK	vs	CK	40%	0

and amount to a 40% difference in amino acid sequence throughout the S region. In comparing more closely the individual members of a set, one can see conservative substitutions, occasional radical amino acid interchanges, and regions of identity (Figure 6). The same is true in comparing similar numbers of hemoglobin chains (87). The amino acid alternatives at a single variable position can, for the most part, be generated by single base changes in the corresponding codons. There appears to be no uniform gradient of variability throughout the entire S region and no single mutational mechanism which can account for the types of variation noted in the S region (models tested include reading frame shifts, in-phase crossing over between two genes, inversions, and chromosomal rearrangements). Instead, these differences appear to have been generated by the gradual accumulation of highly selected point mutations similar to those found in evolutionarily related proteins.

In the case of the hemoglobins, different germ-line genes encode the various chains (e.g. α , β , γ , and δ chains). The simplest explanation for the origin of light chain specificity regions would be to propose that each light chain is encoded by a distinct germ-line gene.

Although most proteins which have been sequenced for at least 18 amino terminal residues are individually unique, the variation within a light chain set at a given residue is extremely limited (Figure 6). Generally, only 1, 2 or occasionally 3 residues have been found at any particular position. A few positions, however, appear exceptionally

variable, and show many changes. The same is true with the hemoglobins even when large numbers of sequences are compared. This would suggest that there are powerful selective forces severely limiting the amount of amino acid variability which can be expressed.

In summary, this comparison of light chain sequences and hemoglobin chains has established the following:

a) There is a minimum of three distinct sets of S sequences, SL, SK_I, and SK_{II}, which are encoded by separate germ-line genes; that is, at least two separate genes are required to code the SK subclasses.

b) The various members of a set appear to be evolutionarily related proteins, suggesting that each of them is encoded by a separate germ-line gene which arose by the process of chemical evolution in a manner similar to the hemoglobin genes.

A single, straightforward explanation for light chain diversity suggests itself at this point: namely, that each complete light chain is encoded by a separate germ-line gene. Unfortunately, this simple explanation runs into difficulties when the evidence relating to the common region is examined critically. This evidence is compelling and will be summarized here because it has very important implications regarding mechanisms of antibody diversity.

First, the human CK regions from different light chains have identical amino acid sequences except for the presence of an amino acid substitution at CK (191) (leucine for valine)(67,68). This subtle change can be detected by serological techniques, and accordingly it can be demonstrated that certain individuals have only the leucine variant,

others have only the valine variant, while others have both. Appropriate genetic crosses involving homozygous (i.e. only valine or only leucine) and heterozygous (i.e. valine and leucine) individuals demonstrate that these two variants behave as alleles of a single gene (88,89). This observation argues strongly for the singleness of the gene controlling the CK region. Furthermore, if it were multiple, one would not expect that the same point mutation would occur in all the members of the gene complex. If the mutational divergence occurred before successive gene duplications had taken place, recombination would have been expected to lead to randomization of the two alleles. This has not occurred, and the CK region in humans is clearly encoded by a single gene. The same situation probably occurs in mice though variation of this type has been eliminated by the use of the highly inbred BALB/c line of mice.

A second type of argument in support of this point appears abstruse at first but is, we believe, compelling. It is clear that the human and the mouse CK genes are evolutionarily related (60% of their amino acid residues are identical (90)) and that both have diverged from some common ancestor. If multiple CK genes existed in this ancestor, then upon divergence of the human and mouse lines each of the members within a set of CK genes must have evolved in identical fashion to produce in the mouse one set of identical CK genes and in the human a second set of identical genes. This is an impossible constraint to place on a large number of DNA sequences, each more than 300 nucleotides in length. Furthermore, the CK regions encoded by these two sets of genes

(CK human and CK mouse) had to diverge by 40% of their amino acid sequence thus revealing considerable possibility for variation with retention of function.

Quite clearly, then, the CK regions cannot be encoded by multiple genes; they must represent a single gene.

The importance of stressing this conclusion is, of course, related to the fact that we have previously concluded that the SK regions must be encoded by a minimum of two germ-line genes in both man and mouse. Either subclass SK_I or SK_{II} can be associated with a single CK region (e.g. in humans both SK_I and SK_{II} have been found combined with the genetic variant having valine at position 191; in mouse the two subclasses are attached to the unique CK sequence). An inescapable conclusion is, therefore, that each light chain is encoded by two genes (SK and CK) which are expressed as a single, continuous polypeptide chain. Thus the "one gene, one polypeptide chain" hypothesis must be broadened to include the special case of antibody light chains which need "two genes, one polypeptide chain ".

The S and C regions, being encoded by separate genes, must have had separate but interdependent evolutionary paths: separate since the S and C regions are encoded by distinct genes; interdependent because they must function together to make a light chain and at some level of protein synthesis have recognition or attachment sites. A joining mechanism is required to unite information from the S and C genes. The joining could occur at the polypeptide level or, as seems more likely, at the DNA (or RNA) level.

At what stage is variability generated within a given K subclass? A simple explanation is that many germ-line genes encode the SK sequences and that these arose through the normal process of chemical evolution involving gene duplication, mutation, recombination and selection for function. This genetic model for antibody variability will be discussed and contrasted with somatic models in Chapter 5.

CHAPTER IV

LIGHT CHAIN EVOLUTION

INTRODUCTION

A study of light chain evolution is a study of two discrete gene systems. On the one hand, the S region of light chains must be generated either by multiple germ-line genes or by an unusual genetic mechanism which generates light chain diversity from a few genes during somatic differentiation. On the other hand, each C region of a light chain class appears to be encoded by a single gene which behaves in an ordinary fashion.

The S regions of human myeloma proteins exhibit an amino acid sequence variability which is similar to that seen in other evolutionarily related protein systems (cytochrome C's and hemoglobins). The SK region variation seen in one species (and probably in one individual) exhibits a phylogeny normally seen in the sequences of a single type of protein (cytochrome C) from different points on the phylogenetic tree. Furthermore, the S regions of mouse are virtually indistinguishable from those of man. This is a very striking observation. Is this indistinguishability a consequence of the evolution of a multigene system or a result of intense selective pressures on a somatic mechanism? Will the S regions of other animals be indistinguishable from those of mouse and human? Can we, from the analysis of S regions in many different animals, generate clues as to the nature of the genetic mechanism responsible for anitbody diversity?

A study of C region evolution poses many of the more classic questions of protein evolution. Do all animals have CK and CL genes?

Are there other light chain classes? Will the expression of CK and CL in different animals give us any insights into the nature of control mechanisms in immunoglobulin synthesis or the nature of selective forces which are involved in the evolution of these genes?

Since myeloma proteins are not available from most species, light chain studies require the use of normally heterogeneous serum globulins. For example, in humans two classes of light chains are present (L and K), each having an unknown degree of heterogeneity in the S region. Light chains used in these studies were derived from the pooled immunoglobulins of several individuals. The heterogeneity of such materials quite naturally limits the applicability of many techniques commonly used in structural studies. We have taken two general approaches in studying pooled light chains from twenty species of vertebrates. General physical, chemical and immunological techniques have been used for rough characterization of the chains, and limited sequence studies have attempted to focus on small selected regions of light chains.

MATERIALS AND METHODS

Purification of IgG: The starting materials (IgG or serum), their source and methods of purification are given in Table I. Purification method 2, chromatography on DEAE Cellulose or Sephadex, has been described in Chapter 3. Purification method 3 was carried out as follows: 10cc of serum were mixed with an equal volumn of saturated ammonium sulfate, the precipitate spun down and dialyzed against borate buffer

TABLE 1

Sources of Materials. Purification procedures used were as follows: 1) used without further purification; 2) ammonium sulphate precipitation followed by chromatography on DEAE cellulose or DEAE sephadex; 3) starch block electrophoresis followed by gel-filtration on G-200 Sephadex. Antisera listed here were directed against whole serum from the appropriate species and were used for evaluating the purity of the IgG preparations.

TABLE 1

Animal	Starting Material	Sources of Material	Purification of IgG	Source of Antisera
Human	IgG	Pentex	1	Hyland Labs
Cow	IgG	"	1	"
Pig	IgG	"	1	"
Guinea Pig	IgG	"	1	-
Rat	IgG	"	1	Caltech
Dog	IgG	"	1	Hyland Labs
Rabbit	IgG	"	1	Hyland Labs
Horse	IgG	"	1	Hyland Labs
Sheep	IgG	"	1	Hyland Labs
Cat	IgG	"	1	Hyland Labs
Chicken	IgG	"	1	Caltech
Turkey	IgG	R. Templis	1	-
Monkey	serum	R. Owen	2	Caltech
Baboon	serum	Hyland Labs	2	Caltech
Mink	serum	F. Dixon	2	D. Porter
Duck	serum	H. Grey	2	H. Grey
Turtle	serum	H. Grey	2,3	H. Grey
Brown Trout	serum	J. Wright	2,3	Caltech
Tuna	serum	L. Barrett	2,3	Caltech

and then subjected to starch block electrophoresis in 0.2 M-borate buffer, pH 8.0 for 18 hours at 400 volts (115 v/cm) at 4°C (97). The IgG region was eluted from starch and submitted to gel filtration on G-200 Sephadex equilibrated with 1.0 M-NaCl and 0.1 M-tris buffer, pH 8.0 at 4°C. The peaks were characterized by the O.D. 280, pooled, dialyzed against distilled water and lyophilized.

Testing for Purity: Antiserum directed against the whole serum of individual species was used with immunoelectrophoresis to evaluate the purity of the respective IgG proteins. The sources for these antisera are indicated in Table 1. "Caltech" antisera were produced as described in Chapter 2.

Discontinuous Gel Electrophoresis: Totally reduced and alkylated chains (see Chapter 2) were examined at a slightly alkaline pH (approximately 8-9) to examine banding patterns. The method used was essentially that of Reisfeld and Small (98).

Aminoethylation of Proteins (99): This procedure was used to convert cystine residues into S-aminoethyl cysteine which is a trypsin substrate. Aminoethylated proteins can thereby be cleaved into smaller and hopefully more soluble fragments by trypsin. Large cystine peptides are notoriously insoluble in most chromatography and electrophoresis solvent systems.

200 mg of protein was dissolved slowly in 10 ml of a 0.2 M tris-10 M-urea solution titrated to pH 8.6 with concentrated HCl. 120 mg of beta-mercaptoethylamine-HCl, dissolved in 1 ml of water, was added to this solution and allowed to react for 1½ hours at room temperature

with constant stirring. 280 µl of ethylenimine were then added at 20 minute intervals three times with thorough mixing. Excess ethylenimine was reacted with 1.0 ml beta-mercaptoethanol, and the protein was dialyzed against repeated changes of an appropriate buffer (e.g. 0.1 N-ammonium bicarbonate for trypsin digestion or 1N-acetic acid for chain separation).

Light and heavy chains were separated from aminoethylated IgG by gel filtration on G-100 Sephadex or P-200 acrylamide equilibrated with 1N-acetic acid (see Chapter 2).

Peptide Maps: The light chains from 15 species were reduced and aminoethylated, digested with trypsin and fingerprinted in duplicate. One set of maps was dipped in collidine/ninhydrin solution and photographed in transmitted light using a 570 mµ interference filter to locate peptides, and the second was sprayed lightly with a dilute ninhydrin solution (.05% w/v in acetic acid/acetone, 1:20). This procedure is carried out in such a way that only the surface of the paper is moistened, thus assuring that only about 10% of the peptide amino groups are destroyed. Regions containing peptides were then cut out from the second map and eluted in water. Peptides were then hydrolyzed with 6N HCl at 105° and analyzed by high voltage electrophoresis on paper. Amino acid compositions of peptides at identical positions were then compared to confirm identity or difference.

Isolation of N-terminal Peptides: Amino terminal peptides were isolated from 3 animals with blocked light chains using the enzymatic digestion-Dowex 50 (H+) column procedure described in Chapter 3.

Subtilisin was used to obtain small blocked peptides and thereby minimize the problems of light chain heterogeneity.

After passage through the Dowex 50 (H+) column, the peptides (ninhydrin negative, peptide bond spray positive) were electrophoresed at pH 6.5 (10% pyridine acetate), eluted from paper, acid hydrolyzed, and amino acid compositions were determined by high voltage electrophoresis on paper.

Isolation of C-terminal Peptides: This procedure takes advantage of the fact that the C-terminal tryptic peptides from light chains of most species are, after oxidation, very acidic (they contain no basic amino acids and at least two acidic residues, i.e. glutamic acid and cysteic acid) (92,95). These acidic peptides can be separated from all others by electrophoresis at a neutral pH.

Reduced but not alkylated light chains were prepared from IgG. These proteins were performic acid oxidized, digested with trypsin, applied on paper as a 10 cm band (approx. 1 mg/cm), and run electrophoretically at 2000 v (35 v/cm) for 1½ hours at pH 6.5 (10% pyridine acetate titrated to pH 6.5 with concentrated acetic acid). After electrophoresis, ½ cm guide strips were cut from each side of the band and developed with ninhydrin to locate the appropriate acidic peptides. These peptides were then cut from the preparative paper and eluted with water.

Hydrolysates of all acidic peptides were quantitated on paper; amino acid compositions of larger acidic peptides were also determined on the amino acid analyzer.

Amide Determinations: Amide assignments were determined from peptide mobilities (relative to aspartic acid at pH 6.5) and molecular weights according to the method described by R. E. Offord (100).

RESULTS

Purification of IgG: The availability of the Pentex Inc. IgG proteins greatly simplified the task of purification. Most of these proteins, when subjected to immunoelectrophoresis, reacted against anti-whole serum to give a heavy IgG precipitin line and only minor serum contaminants. No further purification was attempted, however, as we were unable to demonstrate any contaminants in the light chain preparations after chain separation.

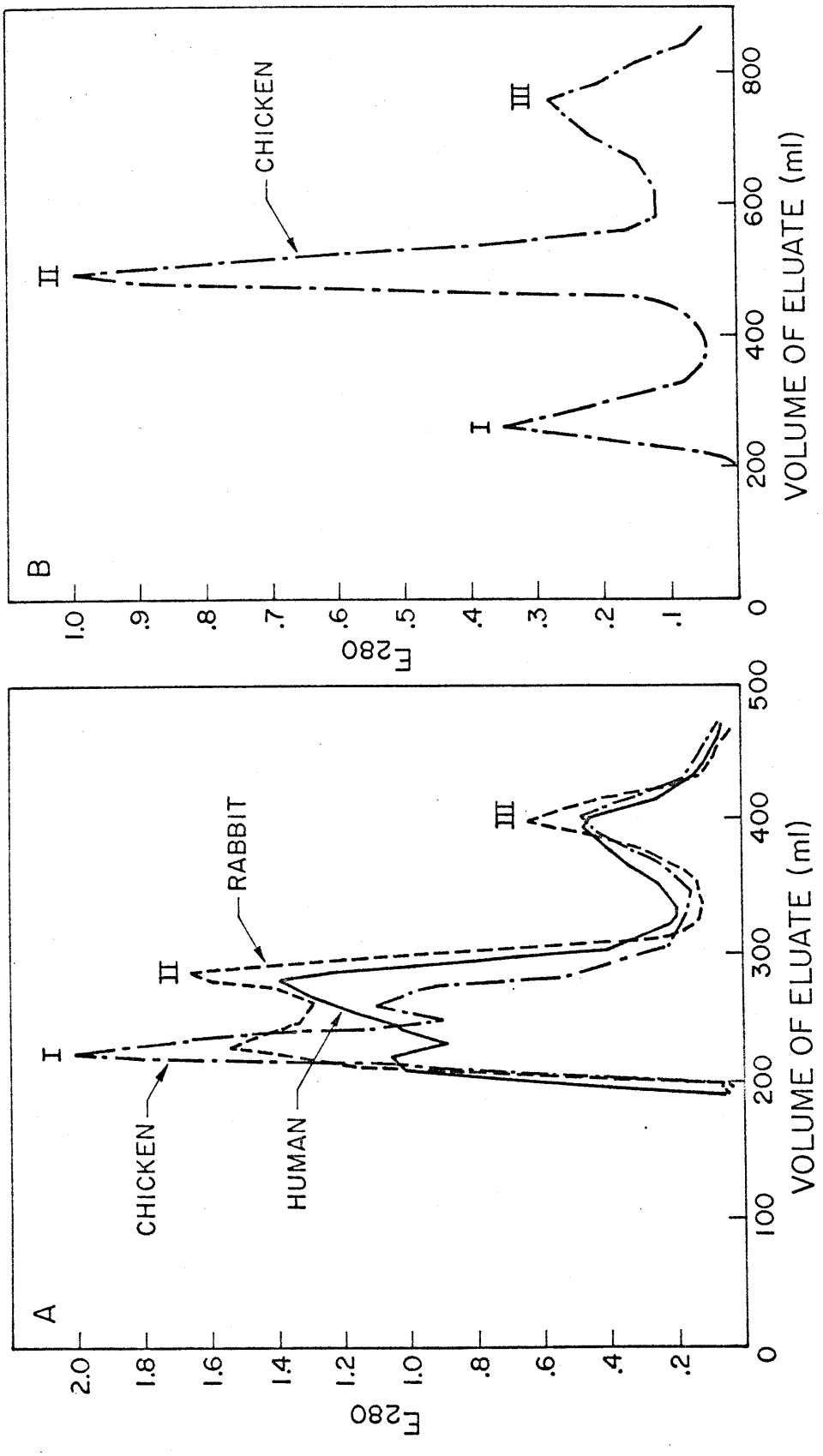
The isolation methods of serum IgG's are summarized in Table 1. Purification of the serum IgG on DEAE-Sephadex and Cellulose was relatively successful with the primate serums, giving 50-75% yields. These proteins gave a single IgG precipitin line against anti-whole serum. Purification of duck, mouse and mink IgG's, however, was much less successful as yields from both DEAE-Cellulose and Sephadex were approximately 10%. Again single IgG precipitin lines were obtained. IgG from turtle, brown trout and tuna could not be eluted from the DEAE columns, and an attempt was made to purify these proteins using a combination of starch block electrophoresis and G-200 Sephadex gel filtration. Significantly higher yields (10-20%) were obtained although it was impossible to remove all minor contaminants.

Chain Separation: The elution patterns of G-100 Sephadex and P-200 Bio-gel are presented in Figure 1. The P-200 column gave three well separated peaks; I is aggregated heavy chain, with some light chain contamination, II is heavy chain monomer (very slightly contaminated with light chain), and III is light chain uncontaminated with heavy chain. G-100 Sephadex did not separate the first two peaks but did yield clean light chains. Gel filtration yields were approximately quantitative.

The light chains of IgG from fish (tuna and brown trout) and reptile (turtle) sera could not be separated by this procedure; intact IgG emerged from the gel filtration column in each case. Presumably the noncovalent bonds joining light and heavy chains are not broken in acetic acid. This observation seems to suggest a stronger noncovalent association of light and heavy chains in cold blooded vertebrates. Immunodiffusion tests were carried out on intact IgG, but no attempt was made to perform chemical studies on the proteins from these species.

The light chains of all higher vertebrates were eluted from the gel filtration column at essentially identical elution volumes. This suggests that all these chains have similar molecular weights. The G-100 Sephadex column was calibrated with a mixture of trypsin (approximately 25,000 mol. wt.) and lysozyme (approximately 14,000 mol. wt.). Comparison of the elution positions of these proteins with those of the light chains indicates that the light chains have a molecular weight of 20,000 - 25,000. Clem and Small (101) used analytical ultracentri-

Figure 1. "A" represents the chain separation pattern of rabbit, human, and chicken Ig on Sephadex G 100 (2.5 cm x 125 cm). "B" shows the gel filtration pattern of rabbit IgG on Bio-gel P 200 (3.5 cm x 150 cm). Approximately 150 mg protein loads were used on these columns. Both fractionations were carried out at room temperature in 1N-acetic acid. I, II and III denote heavy chain aggregates, heavy chain monomers, and light chain monomers respectively.

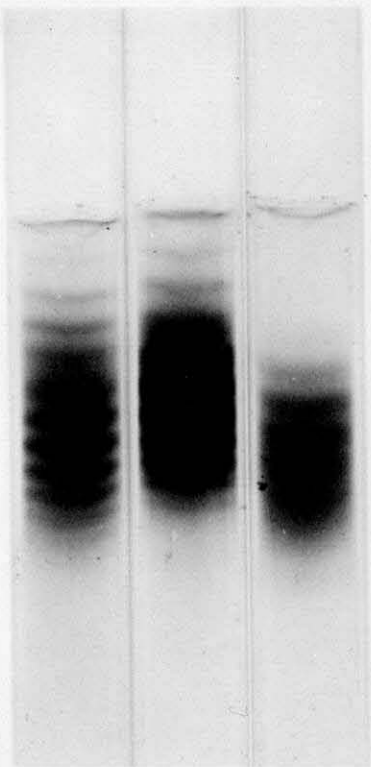

A
B
E_{280}
VOLUME OF ELUATE (ml)
I
II
III
CHICKEN
RABBIT
HUMAN

fugation to obtain a molecular weight of approximately 23,000 for lemon shark and rabbit light chains. Marchalonis and Edelman (102) obtained similar molecular weights for bullfrog light chains. These agree closely with those deduced from the sequence analysis of myeloma light chains of humans and mice (92,94). Thus it seems certain that all vertebrate light chains are similar in size.

Discontinuous Gel Electrophoresis: Cohen and Porter (103) demonstrated that normal human light chains are resolved into about ten regularly spaced components on starch gel electrophoresis at pH 7 to 8 and that a similar degree of complexity can be seen in other mammalian species (i.e. guinea pig, cow, horse, and baboon). Each electrophoretic band is itself heterogeneous and contains a large number of different chains. These light chain bands differ from one another by unit charges, as has been demonstrated by adding a single charge to a homogeneous myeloma light chain (104). Hence the pattern is made up of bands of light chains differing from one another by unit charges in the S regions (CL and CK are present in all bands).

Light chains from 12 species were examined by electrophoresis on acrylamide gels. Each species gave 6-8 bands (Figure 2) with comparable Rf values (Table 2). When light chains from two different species were mixed (1:1) and then examined, the band width and number were unchanged, confirming the pattern identity between species (Figure 2). Clem and Small (100) have noted similar banding patterns in the lemon shark. It seems quite clear that the complexity of this dimension of S heterogeneity is similar in primitive and higher vertebrate light chains.

Figure 2. Acrylamide electrophoretic patterns of light chains. A, rabbit; C, chicken and B, 1:1 mixture of rabbit and chicken. The missing bands in the rabbit sample were seen by examining higher concentrations of this light chain. The procedure is identical to that of Reisfeld and Small (98).



A B C

TABLE 2

Light Chain Electrophoretic Band RF's. $+R_f$ values are relative to a tracker dye, brom phenol blue.

TABLE 2

Source of Light Chains	"R_f" Values[+] of Light Chain Electrophoretic Bands							
	1	2	3	4	5	6	7	8
Human	.05	.09	.12	.15	.18	.21	-	-
Pig	.05	.08	.12	.14	.16	.19	-	-
Cat	.06	.09	.12	.14	.17	.19	-	-
Dog	.06	.09	.12	.15	.18	.20	-	-
Guinea pig	.04	.07	.09	.12	.14	.17	-	-
Rat	.05	.08	.11	.14	.17	.19	.22	
Duck	.04	.08	.11	.14	.17	.20	.22	-
Bovine	.05	.08	.11	.14	.17	.19	.23	-
Sheep	.05	.08	.11	.14	.17	.20	.23	-
Chicken	.04	.07	.11	.13	.17	.20	.22	.23
Rabbit	.04	.07	.10	.4	.17	.19	.22	.24
Turkey	.04	.08	.11	.14	.17	.19	.20	.22

Immunology: Antiserum to a heterogeneous population of polypeptide chains (e.g. human light chains) can sometimes recognize single amino acid changes (e.g. valine to leucine change at CK (191) in human light chains) (106) and also differences in structure unrelated to primary amino acid sequence (e.g. different carbohydrate moieties attached to light chains may be antigenic). In spite of these possibilities, it appears likely that most antibodies are directed against sequences of amino acids in the CL and CK regions since they are common to all chains in the pools. Particular antigenic configurations in the SK and SL regions are evidently present in very low concentrations due to extensive S region heterogeneity; hence, in general, antibody would probably not be directed against this region.

Antisera were produced in rabbits against the light chains from five species: three primates (human, monkey, and baboon), one carnivore (dog), and one artiodactyl (sheep). This group was selected to permit a study of immunologic similarities in closely and more distantly related mammals.

All antisera to light chains reacted strongly with their homologous antigens and with homologous IgG as measured by immunodiffusion. Results of immunodiffusion tests on light chains and IgG from eighteen species are shown in Table 3. All positive reactions were confirmed by immunoelectrophoresis to exclude the possibility of contaminating cross-reactivity (e.g. by albumin). As we expected, related species showed cross-reactivity to varying degrees (Figure 3) though pig light chains did not react with antiserum to sheep light chains.

TABLE 3

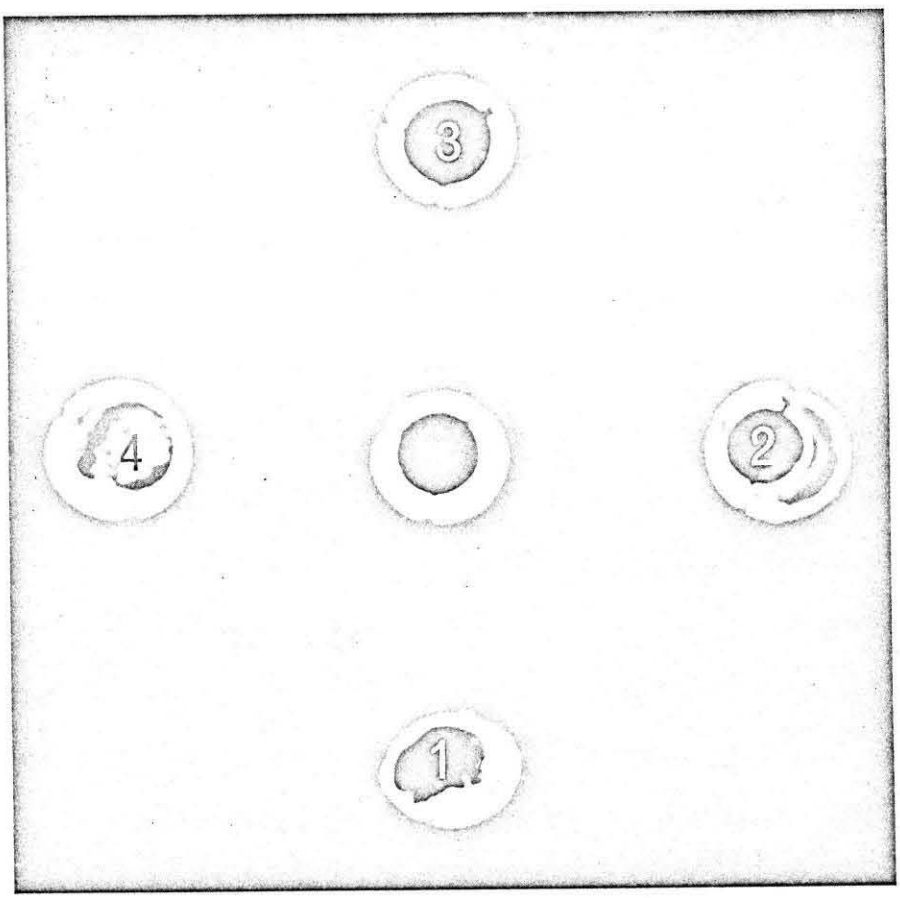
Summary of Immunodiffusion Analysis of Light Chains from Various Species.

Rabbit antisera were prepared against purified light chains from the five test species. Other light chains and/or IgG were tested against these by double-diffusion in agar gels. "C" indicates complete reaction between the light chain and the antiserum; that is, these proteins have most of the antigenic determinants that the antiserum is capable of recognizing. "P" denotes partial reaction between the light chain and the antiserum; that is, this protein does not have all the antigenic sites the serum can recognize and forms spurs with the "C" proteins.

TABLE 3

Antigen Source		Rabbit antisera against light chains from				
		Human	Baboon	Monkey	Dog	Sheep
Primates	Human	C	P	P	-	-
	Baboon	P	C	P	-	-
	Monkey	P	P	C	-	-
Rodents	Guinea Pig	P	P	P	-	-
	Rat	-	-	-	-	-
	Mouse	-	-	-	-	-
Carnivores	Dog	-	-	-	C	-
	Cat	-	-	-	P	-
Artiodactyls	Sheep	-	-	-	-	C
	Cow	-	-	-	-	P
	Pig	-	-	-	-	-
Perissodactyls	Horse	-	-	-	-	-
Avians	Chicken	-	-	-	-	P
	Duck	-	-	-	-	-
	Turkey	-	-	-	-	-
Reptiles	Turtle	-	-	-	-	-
Fish	Brown Trout	-	-	-	-	-
	Tuna	-	-	-	-	-

Figure 3. Immunodiffusion: monkey (1), baboon (2), human (3) and guinea pig (4) light chains reacted against rabbit anti-monkey light chain-serum.


3
4
2
1

Two unexpected reactions were observed: guinea pig light chains reacted with anti-primate sera (Figure 3), and chicken light chains reacted with antiserum to sheep light chains. In both cases similar cross reactivities were noted among the respective heavy chains and heavy chain antisera. The observation that none of the anti-heavy chain sera reacted against either of these light chains rendered unlikely the possibility that a common carbohydrate moiety shared by light and heavy chains was the basis of the unexpected cross reactivity.

The fact that guinea pig light chains reacted with antiserum against human chains indicated that they were reacting with antibodies to human CK or CL or both. When guinea pig light chains were reacted with antisera directed against human myeloma L and K chains, cross reactivity was demonstrated against both. The same was true of monkey and baboon light chains; hence on the basis of immunological cross reactivities it appears that guinea pig, baboon and monkey all have light chain regions which share some similarities with those of man (Figure 3). It should be mentioned that Nussenzweig, et.al. (106) have also demonstrated the existence of two distinct light chain classes in guinea pig by immunological techniques.

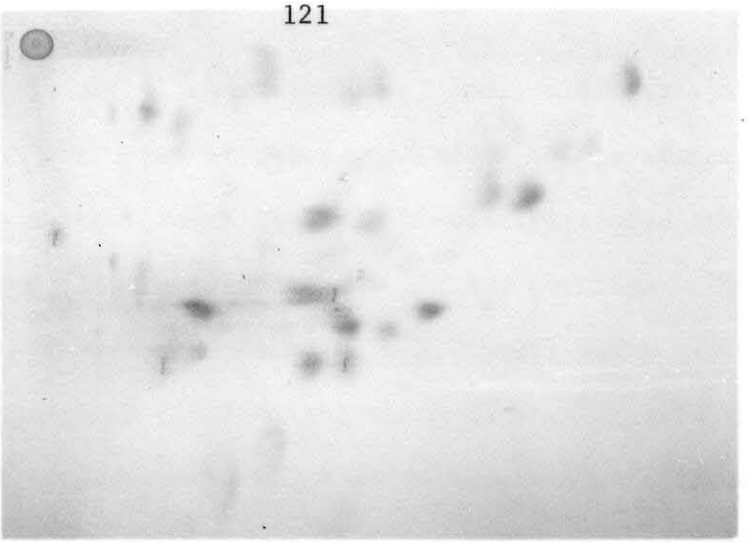
The observations are, for the most part, what would be expected from a set of evolutionarily related proteins; i.e. those animals which are most closely related have proteins with similar antigenic determinants in their primary amino acid sequences whereas light chains from distantly related animals are not sufficiently similar to reveal common antigenic sites. One can, even on the basis of the limited sets of

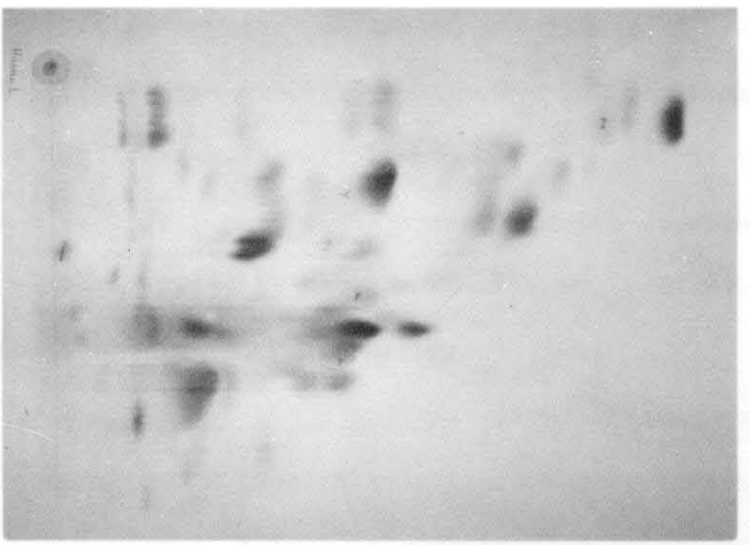
serologic reagents available, start to construct an "immunologic" phylogenetic tree based on common regions in CK and CL. The correspondence between this and the classical phylogenetic trees constructed on the basis of diverse morphologic features is hardly satisfactory, but the anomalies will probably be resolved as more reagents become available.

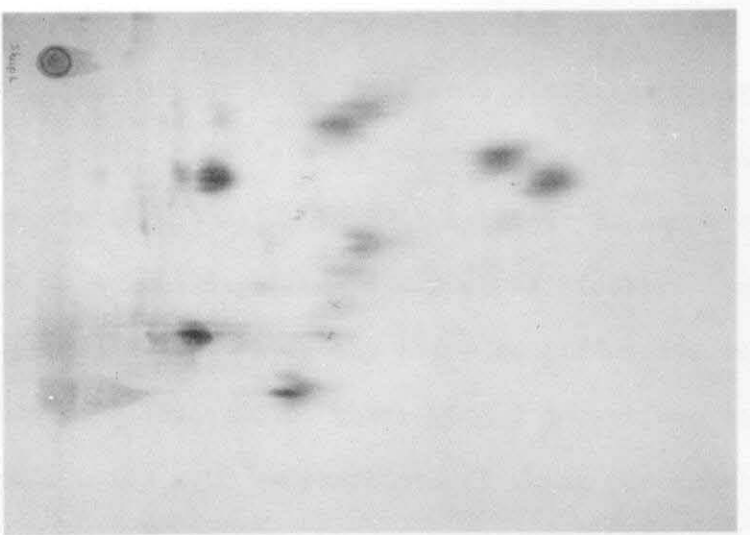
Peptide Mapping: Some limitations of this technique should be borne in mind. Peptide mapping reveals only those peptides which are present in sufficient concentrations for detection by stains such as ninhydrin; therefore many peptides in highly heterogeneous S regions are probably not seen by this procedure. Furthermore, this technique tends to emphasize differences rather than similarities in protein structure; for example, the change of a single charged amino acid in a decapeptide could change its map position even though 90% of the sequence is unchanged. Perhaps this is best illustrated by noting that although the mouse and human CK regions (107 residues) are alike at 64 of the 107 amino acid positions (90), they have but one tryptic peptide in common. Therefore, the presence of many common peptides implies a very significant degree of sequence identity.

As with the serological classification, the light chains from all animals can be divided into groups on the basis of fingerprint similarities. Members of the following five groups of animals share seven or more identical peptides with other members of the group (all peptides are not necessarily from a single class of light chain): 1) primates (human, baboon, monkey) (Figure 4); 2) artiodactyls (sheep and cow, but not pig); 3) carnivores (dog and cat); 4) rodents (rat

Figure 4. Peptide map comparison of pooled light chain from three mammalian species. Chromatographic dimension horizontal (left to right). Electrophoretic dimension vertical (pH 3.6, cathode at top). From left to right baboon, human and sheep.







and guinea pig); and 4) birds (chicken, duck, and turkey). The pig and the horse each share two or three peptides with the artiodactyls, indicating a more distant relationship. Distantly related animals (e.g. human and sheep, Figure 4) share few, if any peptides. Some of the peptide maps are strikingly simple (e.g. sheep) which probably indicates the predominance of a single light chain class. It should be noted here that the peptide maps, which contain much more information than immunodiffusion plates, did not indicate phylogenetic anomalies. (The serologic similarities between guinea pig and primate light chains and between chicken and sheep light chains are not reflected in the peptide maps.) Even fewer anomalies would be expected if complete amino sequences were available for the many species because these would enable us to distinguish between convergence and divergence of the actual structural genes of the organisms (109).

Amino Acid Analysis of Light Chains: Light chains were prepared from IgG using reduction without alkylation. When proteins are oxidized with performic acid, 95% of the cystine (cysteine) residues are converted to cysteic acid (62), a stable sulfur derivative which can be determined accurately by amino acid analysis. The analysis of these light chains is given in Table 4 as residues per mole of protein. Chicken and rabbit light chains have 7 ½-cystines as compared with 5 in most animals. This is a significant structural feature (perhaps comparable to a sequence gap) which may serve to differentiate light chains into two subclasses (i.e. those with 7 ½-cystines and those with 5). The overall amino acid compositions of these proteins are remark-

TABLE 4

Composition of Oxidized Light Chains. Samples of light chains were oxidized with performic acid (62) and then hydrolyzed in vacuo for 24 hours (105°C, constant boiling HCl). Analyses were carried out by automatic column chromatography on samples of 6-11 mμ moles. Results are expressed as residues per mole, normalized to a value of 210 residues per light chain, excluding tryptophan. (a) Arginine peak was lost, and value of 6 was assumed as typical of light chains.

TABLE 4

	Human	Rabbit	Chicken	Dog	Horse
$CySO_3H$	4.6	7.3	6.2	4.8	5.5
Asp	16.4	19.0	16.6	19.2	15.1
Met	-	-	1.3	-	-
Thr	16.6	27.7	22.1	16.5	20.7
Ser	27.9	21.5	30.7	25.9	32.4
Glu	24.2	20.5	18.7	20.7	19.7
Pro	19.3	11.9	14.1	15.9	13.5
Gly	14.2	18.4	19.2	17.5	20.0
Ala	14.3	16.1	16.4	13.9	15.7
Val	14.5	19.2	13.3	16.9	16.1
Ile	5.7	6.7	8.6	6.8	7.1
Leu	15.9	11.0	13.4	16.4	12.8
Tyr	7.6	9.9	7.8	6.4	7.2
Phe	7.3	6.3	5.7	7.1	5.2
Lys	9.2	8.5	8.5	10.6	11.2
His	2.9	1.2	2.8	3.4	2.3
Arg	7.0	3.0	7.5	(6)[a]	5.1

ably similar with characteristically high contents of hydroxy and carboxy amino acids (amides).

Light Chain C-Terminal Residues: Hydrazinolysis (Table 5) of oxidized light chains proved somewhat unsatisfactory in two respects: 1) the yields were low ranging between 20-50%, and 2) some of the light chains gave high background values of glycine, serine and alanine, presumably from the hydrolysis of their respective hydrazides. Hydrozinolysis of control proteins gave variable backgrounds of these amino acids. Apart from this variable background of certain amino acids, the hydrazinolysis results are consistent with those obtained from the sequencing of isolated C-terminal peptides.

Sequence Studies: Sequence analysis of L and K human myeloma proteins have shown that these classes can be readily distinguished by chemical techniques at both the N- and C-termini of the molecule (Table 6). The N-terminus of K proteins generally has a free α amino group whereas most L chains are blocked at this position. The C-terminal amino acids of the two classes are different, but the C-terminal tryptic peptides of both are extremely acidic due to the presence of cysteic acid in the oxidized proteins. As a working hypothesis, it was assumed that these features would be shared by light chains in other species. Chemical procedures were therefore designed on this basis to distinguish between the L and K classes of proteins in pooled light chains from 15 different species.

N-Terminal Studies: The Edman procedure (82), which removes one amino terminal residue per cycle, works only with those proteins having

TABLE 5

C-terminal Amino Acid Residues of Light Chains as Determined by Hydrazinolysis. +Approximately 1 mg of light chain (40 mμM) was subjected to hydrazinolysis. Residues are expressed as millimicromoles. *Yields were visually estimated from amino acid analysis on paper. Other results were determined by spectrophometric readings at 500 mμ after elution of cadmium stained amino acids from paper (see methods, Chapter 3). ∓Determined by column chromatography. ‡Significant residues are underlined.

TABLE 5

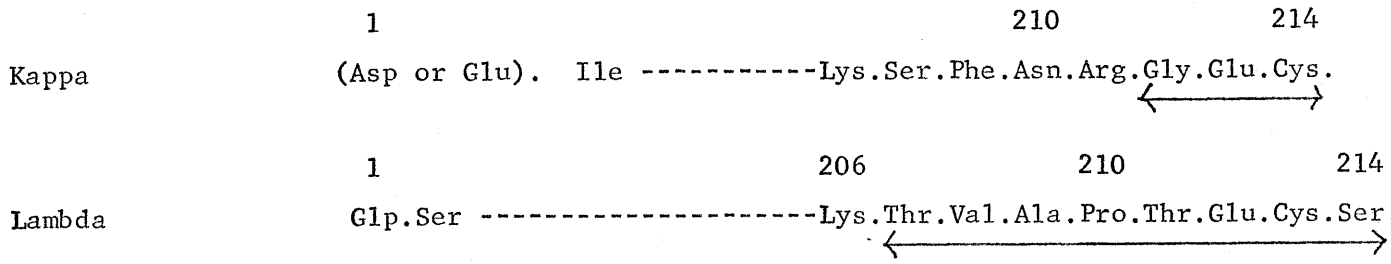
C-terminal Amino Acid Residues of Light Chains
as Determined by Hydrazinolysis

Protein	Amino Acid Residues					
	Glycine	Serine	Alanine	Cysteic Acid	Leucine	Proline
Lysozyme*	2	1	1	1	10‡	—
Turkey*	2	8	1	2	0	—
Duck*	2	3	1	5	0	—
Chicken	2	3	3	11	0	—
Cat*	3	12	1	2	0	—
Dog*	2	15	2	2	0	—
Mink	2	13	1	0	0	—
Sheep	5	14	3	5	0	—
Pig	3	3	11	2	0	—
Horse‡	2	1	1	0	0	10
Rat*	2	2	1	5	0	—
Guinea Pig	2	6	2	8	0	—
Rabbit	2	2	1	8	0	—
Human	4	9	3	8	0	—
HBJ 2 (L)	3	21	2	2	0	—
HBJ 4 (K)	5	4	4	13	0	—

TABLE 6

Terminal Sequences of Human Light Chains. Taken from Hood, Gray and Dreyer (113, 114); Milstein (95); Hilschmann and Craig (67); Titani et al. (68). Glp is pyrrolidone carboxylic acid. $\longleftrightarrow$, carboxyl terminal tryptic peptide.

TABLE 6


1
210
214
Kappa
(Asp or Glu). Ile ----------Lys.Ser.Phe.Asn.Arg.Gly.Glu.Cys.
1
206
210
214
Lambda
Glp.Ser ----------Lys.Thr.Val.Ala.Pro.Thr.Glu.Cys.Ser.

a free α amino group, i.e. K chains. When this procedure was carried out on myeloma K chains, yields were generally greater than 80-90%. Therefore, in pooled light chains the yield of end-group gives an approximate indication of the relative amounts of blocked and unblocked proteins (i.e. L and K chains). Light chains can thus be divided into three groups (Table 7): those whose amino terminal sequences are similar to those of human K light chains (human, baboon, rat and guinea pig); those whose amino terminal sequences are different from human K chains (chicken, rabbit, and pig) and those who have primarily blocked light chains (horse, dog, mink, sheep, and cow).

The following blocked N-terminal peptides from horse, dog, cow and two L myeloma light chains were isolated using the subtilisin-Dowex 50 (H+) column procedure:

		Yield
HBJ 2	Glp—Ser—Ala—Leu	30%
HBJ 7	Glp—Ser—Val—Leu	40%
Dog	(Glx,Ser,Val,Leu)	25%
Cow	(Glx,Ser,Ala,Leu)	30%
Horse	(Glx,Ser,Val,Leu,Ala,Thr,Pro?)	35%

These peptides were isolated in sufficient yields to suggest that they were the predominant amino terminal sequence of all light chains present. The compositions of dog and cow N-terminal peptides are identical to those of the myeloma peptides. On the other hand, the horse N-terminal peptide is quite distinct from those seen in myeloma proteins.

C-Terminal Studies: Performic acid oxidized light chains (L and K), after trypsin digestion, generally have very acidic C-terminal peptides which can be isolated by electrophoresis at pH 6.5. The mobility of

TABLE 7

Amino Terminal Regions of Light Chains. Results of Edman degradation procedure on light chains. All samples were pooled light chains unless otherwise noted. (a) Hood, Gray and Dreyer (113); (b) Doolittle (115); (c) Doolittle, personal communication. The yield refers to total PTH derivative obtained at the first step.

TABLE 7

Amino Terminal Regions of Light Chains

Light Chain Source	Residue Position 1	2	3	4	5	Yield
Human (pooled)[a]	Asp + Glu	Ile	Gln + Val	Met + Leu + Val	Thr	50%
Baboon	Asp + Glu	Ile				50%
Guinea Pig[c]	Asp + Glu	Ile				80%
Rat[c]	Asp + Glu	Ile				65%
Mouse (myeloma)[a]	Asp	Ile	Val	Leu		90%
	Asp	Ile	Gln	Met		90%
Rabbit[b]	Ala + Ile	Val	Val + Leu	Val	Gln	65%
Chicken	Ala	Leu				60%
Pig	Ala + Glu	Leu or Ile				22%
Sheep, Cow	Not identifiable					10%
Mink, Horse		-				0
Human (myeloma L)		-				0

kappa C-terminal peptides relative to aspartic acid is about 1.0 and that of lambda peptides is about 0.5 while most other peptides have mobilities of 0.2 or less, Table 8. Results are summarized in Table 9 which also includes the results of hydrazinolysis of oxidized light chains.

There is striking amino acid sequence homology among all of the C-terminal peptides. Although no C-terminal peptides could be isolated from cow, sheep and mink, serine was obtained as the C-terminal amino acid residue in the light chains in these animals, and the presence of blocked amino groups suggested the predominance of L proteins. Furthermore, a blocked N-terminal peptide was isolated in high yield from cow light chains, and this was identical in amino acid composition to an N-terminal human myeloma L peptide. The failure to isolate a lambda C-terminal peptide in these animals could be explained by either the final C-terminal basic residue (i.e. CL(206) in human) being converted to a nonbasic amino acid or the CL(207) residue being changed to a proline, which prevents trypsin from cleaving at the preceding basic residue. This would yield a large tryptic peptide whose mobility would probably not be sufficient to separate it from the smear of other tryptic peptides.

All peptides were sequenced completely by the dansyl-Edman method. Most of the K and L C-terminal peptides had a net charge of -2 (Table 8). This allowed us to deduce that the glutamic residue was in the acidic form since there were only two acidic groups in most of the peptides, namely glutamic and cysteic acid. The rat K peptide had three acidic,

TABLE 8

Electrophoretic Mobility and Charge of K and L C-terminal Tryptic Peptides+. +Mobility was calculated relative to aspartic acid at pH 6.5. Net charge was determined as described in text.

TABLE 8

Electrophoretic Mobility and Charge of K and L C-terminal Tryptic Peptides

		Relative Mobility	Net Neg Charge
Human	$K\text{-}C_{211\text{-}214}$	1.0	-2
HBJ 4	"	0.99	-2
Guinea Pig	"	0.98	-2
Chicken	"	0.96	-2
Turkey	"	0.97	-2
Duck	"	0.97	-2
Rat	"	1.0	-2
Pig	"	0.94	?
Rabbit	"	1.1	-2
Human	$L\text{-}C_{208\text{-}215}$	0.47	-2
HBJ 2	"	0.48	-2
Horse	"	0.55	-2
Pig	"	0.49	-2
Rabbit	"	0.64	?
Dog	"	0.66	-2

TABLE 9

Carboxyl Terminal Regions of Light Chains. Carboxyl terminal tryptic peptides obtained from performic acid oxidized light chains by electrophoresis at pH 6.5. Regions corresponding to positions of typical kappa and lambda peptides were eluted from all samples, even where no band was detected. Hydrazinolysis results are expressed as uncorrected yields (%). All samples contained small and variable amounts (2-10%) of Gly, Ser, and Ala.

TABLE 9

Species	K-Type (214)	L-Type (215)	Hydrazinolysis
Human (myeloma K)	Gly.Glu.Cys	–	$CySO_3H$ (59)
Human (myeloma L)	–	Thr.Val.Ala.Pro.Thr.Glu.Cys.Ser.	Ser (53)
Mouse (myeloma K)	Asn.Glu.Cys	–	
Human	Gly.Glu.Cys	Thr.Val.Ala.Pro.Thr.Glu.Cys.Ser.	Ser (23), $CySO_3H$ (20)
Rabbit	Gly.Asp.Cys	Gly.Asx(Gly,Asx,Ala,Pro,Thr,Glu,Cys)Ser	$CySO_3H$ (20)
Pig	Asx.Glx.Cys.Glx.Ala	Thr.Val.Thr.Pro.Ser.Glu.Cys.Ala	Ala (30)
Guinea Pig	Ser.Glu.Cys	Ser.Leu.Ala.Pro.Ser.Glu.Cys.Ser.	$CySO_3H$ (20) Ser (15)
Rat	Asn.Glu.Cys	–	$CySO_3H$ (13)
Chicken	Ser.Glu.Cys	–	$CySO_3H$ (30)
Turkey	Ser.Glu.Cys	–	Ser (20) $CySO_3H$
Duck	Ser.Gly.Cys	–	$CySO_3H$ (13)
Dog	–	Val.Ala.Pro.Ala.Glu.Cys.Ser	Ser (38), $CySO_3H$
Horse	–	Leu.Ser.Pro.Ser.Glu.Cys.Pro	Pro (25)
Cat	–	–	Ser (30), $CySO_3H$
Mink	–	–	Ser (33)
Sheep	–	–	Ser (35) $CySO_3H$ (13) Gly (13)

asx, glx, cys, and the amide was assigned by analogy with the published C-terminal sequence for mouse, Asn-Glu-Cys (90). We did not attempt amide assignment with the larger rabbit C-terminal peptide or the pig CK peptide — each with four potentially acidic residues.

The sequence data, although limited, are compatible with the hypothesis that all animals have K-like and/or L-like common regions. Henceforth, we shall designate these chains as K and L.

The free amino terminal residues of pooled human light chains are identical to their K myeloma counterparts (Table 7). Furthermore, other animals have similar SK amino terminal residues (e.g. mouse, baboon, guinea pig, and rat) (Table 6). Those proteins which have N-terminal alanine are correlated with kappa C-terminal peptides (rabbit, chicken, and pig) though pig has an additional two residues at the C-terminus. Blocked N-terminal peptides isolated from cow and dog are also analogous to their human L myeloma counterparts although the horse N-terminal peptide is somewhat different.

Yields can be estimated from all of the N- and C-terminal procedures athough some give only approximate results. When these yields are compared, the blocked N-terminal light chains generally correlate with lambda C-terminal peptides (Table 10). This is most easily seen in the extreme cases. For example, the horse has only blocked protein and only lambda C-terminal peptides whereas the rabbit has predominately K proteins. A rough approximation of the K to L ration for all species examined can be estimated from the yields obtained from the various procedures and is given in Figure 5.

TABLE 10

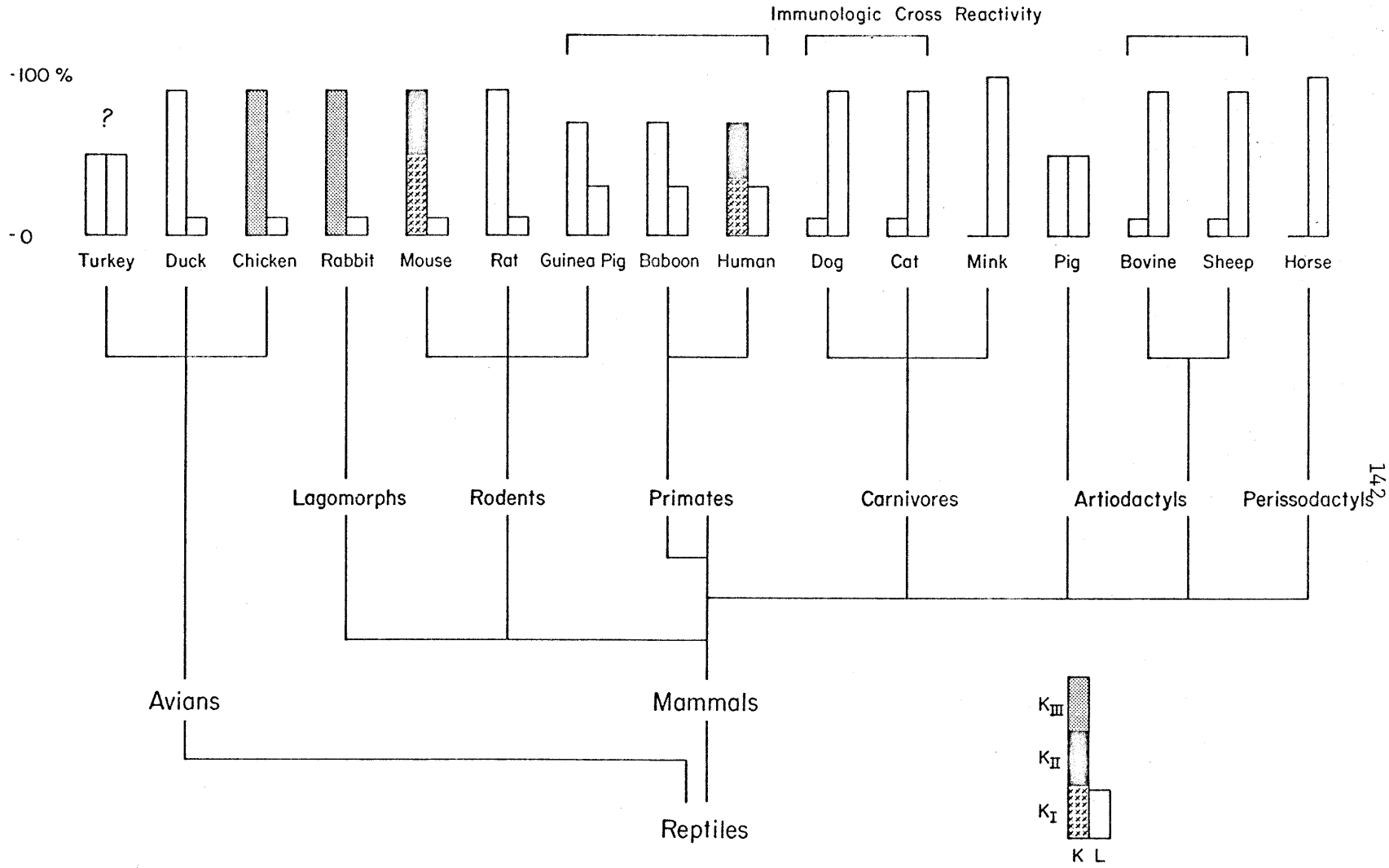
Summary of S and C Pooled Light Chain Sequence Data. $+K\text{-}S_{(1)}$ represents amino terminal residues which were determined by the PITC procedure. ‡Yield was calculated using spectra data as described in the text. ‡Blocked N terminal peptides were isolated using the enzyme digestion - Dowex 50 (H+) column procedure described in text. N.D. signifies not determined. °Percent of L & K C-terminal peptides was calculated relative to yields obtained from human myeloma proteins HBJ 2 (L) and HBJ 4 (K). X bar means looked for but not found.

TABLE 10

Summary of S and C Pooled Light Chain Sequence Data

Species	S Region			C Region	
	$K\text{-}S_{(1)}$+	Yield‡	Blocked N Terminal Peptides‡	%K°	%L°
Human	Asp Glu	50%	+	70	30
Baboon	Asp Glu	50%	N.D.	N.D.	N.D.
Rat	Asp Glu	65%	N.D.	100	——X
Guinea Pig	Asp Glu	80%	N.D.	70	30
Rabbit	Ala	65%	+	90	10
Pig	Ala Glu	20%	N.D.	50	50
Chicken	Ala	60%	N.D.	100	—
Duck	N.D.	N.D.	N.D.	N.D.	—
Turkey	N.D.	N.D.	N.D.	N.D.	—
Dog	N.D.	N.D.	+	—	100
Horse	Blocked	0%	+	—	160
Sheep	?	10%	+	—	—
Cow	?	10%	+	—	—
Mink	Blocked	0%	+	—	—

Figure 5. Distribution of Lambda and Kappa light chains among various species. Figure is based upon the data of Tables 5,7,9, and 10. Mouse ratios are estimated from frequency in myeloma proteins. Vertical bar on left above each species indicates proportion of light chains in K class (sub-divided where known according to K_I, K_{II} and K_{III}); right hand bar indicates lambda chains.


Immunologic Cross Reactivity
100 %
0
?
Turkey
Duck
Chicken
Rabbit
Mouse
Rat
Guinea Pig
Baboon
Human
Dog
Cat
Mink
Pig
Bovine
Sheep
Horse
Lagomorphs
Rodents
Primates
Carnivores
Artiodactyls
Perissodactyls
Avians
Mammals
Reptiles
K_{III}
K_{II}
K_I
K L

The high yield of a single blocked N-terminal peptide, the apparent absence of any free α amino groups, the high yield of a lambda C-terminal peptide, and the high yield of proline (plus some serine) upon hydrazinolysis, indicate that horse has only a single class of light chains (L). There is no evidence of any K class protein in the pooled horse chains examined. Our methods would not detect the presence of less than 2 or 3% K chains.

DISCUSSION

A number of general conclusions can be drawn from this and related studies.

1. The general structure of light chains appears to be very similar in all vertebrates from the most primitive to the most advanced.

2. The C regions of light chains behave as sets of evolutionarily related proteins. Indeed, all animals have light chains analogous to L and/or K classes though the amounts vary widely.

3. Distinct sets of S region sequences can be seen in different animals.

4. The complexity of light chains from even the most primitive vertebrate studied (shark) appears comparable to that of higher vertebrates.

The general structure of light chains appears to be very similar in all vertebrates from the most primitive to the most advanced: all are 20-25,000 in molecular weight; all have similar degrees of electro-

phoretic heterogeneity; and many can combine with the heavy chains of other species to produce normal IgG molecules (by chemical and physical criteria).

The C regions of light chains from 15 species of higher vertebrates are similar to sets of evolutionarily related proteins by immunologic and peptide map criteria, i.e. those animals which are most closely related, have proteins with similar antigenic determinants and peptides. Limited sequence evidence supports these evolutionary relationships and is compatible with the supposition that most higher vertebrates have two light chain classes which correspond to human L and K proteins. No evidence from the structural studies even suggests the existence of a third class. Furthermore, the corrected L and K yields generally account for 90% or more of the light chains present (Table 10).

A question must be raised at this point — namely, how does one distinguish a CK sequence from a CL sequence or either from a third type of C region sequence? Perhaps this distinction is analogous to that which can be made among the β, δ, and γ chains of hemoglobin in one type of animal. These chains are closely related, and over short stretches of amino acid sequence might be difficult to distinguish. Classification of closely related hemoglobin chain types can only come through extensive sequence data. The same is true of the C regions in light chains. Nevertheless, there seems to be sufficient structural homology in all cases (except perhaps the CK region of pig) to support the postulate that most animals have L and/or K chains.

Both light chain classes must have been present in reptiles because the two classes of higher vertebrates (mammals and birds) which independently originated from reptiles each have L and K chains (Figure 5). Recent evidence indicates that sharks also have both light chain classes. The smooth dogfish probably has mostly blocked light chains (L class) (108) whereas the leopard shark has at least 30% K light chains with an Asp-Ile-Leu amino terminal sequence closely homologous to the corresponding SK sequence in human K light chains (compare with Table 7) (109). It seems that L and K classes probably exist in all vertebrates from the most primitive to the most advanced.

The ratio of K to L chains varies markedly from one species to another (Figure 5). It is difficult to say what this means in terms of gene evolution. Some of this difficulty arises from the fact that we are examining highly selected products of the gene pool and not the genes themselves. What are the selective forces involved in producing the various L to K ratios? If we assume that an ancestor common to artiodactyls and rodents had a 1:1 ratio of L to K, then selective forces in artiodactyls had to shift this ratio to 9:1 and in rodents to 1:9. The particular gene configurations which are preserved during the course of evolution must have a positive survival value for the organisms. Since there are two functionally separate parts of the light chains, i.e. the S and C regions, selective forces could act on either component.

It would seem likely that one of the most important light chain selective forces is that of the antigenic environment acting on the

This is not a missing page. It is a result of page misnumbering.

S regions. Nessenzweig and Benacerraf (110) have demonstrated that particular antigens can stimulate the production of single light chain classes. If a particular pathogen (or set of pathogens) threatened the existence of a given species, and if the effective S regions in antibodies to this pathogen were located primarily in one light chain class, then this organism would probably, in time, shift the class ratio to favor the effective antibody. There are, however, some interesting questions which are difficult to answer in this regard. Why do the ratios of K and L in light chains taken from individuals of one species (e.g. humans) always seem to remain relatively constant? Indeed, there seems to be a tendency for closely related animals (e.g. primates) to have similar L to K rations (Figure 5). Why should very different kinds of animals (e.g. birds and rodents) have similar L to K ratios when one might expect them to have very different antigenic environments? Perhaps the answer, in part, to these questions lies with selection operating in the C region.

Common region selective forces would presumably be directed toward improvements in function. For example, a CK region which joins with the heavy chain in such a fashion that it is possible to have better binding sites or which enables the heavy chain to carry out some of its other functions more efficiently, has a definite selective advantage over its CL counterpart. Changes in one C region could also enable the corresponding polypeptide chains to be removed more easily from the polyribosomes, or in some other way could affect control mechanisms regulating immunoglobulin synthesis. Different internal

milieux present in different species could favor the effectiveness of one common region over the other; perhaps the higher body temperature characteristic of birds could favor the finding of one C region to heavy chains.

Admittedly all comments on the nature of selective forces must be highly speculative. Nevertheless, there are some general observations which can be made about the variation in the K to L ratio .

1) The K to L ratios are generally similar in animals which are phylogenetically related (Figure 5). It would be interesting to examine the light chain ratios of undomesticated species and determine whether or not a "domesticated" environment can change K to L ratios.

2) Apparently the loss of the ability to express a light chain class is compatible with organism survival. For example, the horse apparently has only a single class of light chains though some ancestors must have had both light chain classes. It is difficult to imagine why animals with a single light chain class were, in time, selected over those that had both classes. Presumably the loss of a light chain class was in some fashion correlated with an event which conferred a selective advantage.

3) Since horses lack (or repress) the CK gene while their ancestors had it, caution should be used in drawing conclusions about gene evolution from existent immunoglobulin populations. For example, it has been asserted that the M gene is the most primitive of the heavy chain genes because it is the only class found in the dogfish shark (111). It would seem just as plausible to suggest that the particular environ-

ment gave the M gene a selective advantage over its other heavy chain counterparts.

Information has also been obtained about the S regions of light chains. For example, distinct sets of S regions can be seen in different animals. The SK regions of rabbit and chicken probably differ in at least two significant structural features from human SK regions in that 1) they possess alanine as the N-terminal residue and 2) they have 7 rather than 5 ½-cystines (Table 4). Doolittle has indirect evidence in the case of rabbit chains that this extra disulfide bridge is in the S region (112). Hence, these two structural differences suggest that rabbit and chicken light chains can be divided into a discrete subclass of SK sequences distinct from those of mouse and human (see Discussion, Chapter 3).

The question of whether or not such species-specific subclasses can more easily be explained by germ-line or somatic theories must be deferred until more concrete sequence date are available. It will be interesting (and feasible) to obtain at least fragmentary N-terminal S region sequences on chains from selected animals and perhaps thereby reveal unsuspected sequence patterns. In any case, all theories of antibody formation will be required to explain the presence of distinct S regions in different animals.

One furtherpoint will be made about the S region of normal pooled human light chains. Amino acid alternatives expressed at the N-terminal five positions of these chains (Figure 7) are identical with those seen among the myeloma light chains (see Figure 6, Chapter 3). Oper-

ationally one could pool all of these K myeloma light chains and obtain from N-terminal analysis (the PITC procedure) the normal pooled light chain results. The observation strongly suggests that a random population of immunocytes has in fact been transformed by the myeloma process. This argues that myeloma proteins are individual members of the normal immunoglobulin population.

One final observation about light chain evolution is the striking structural similarity among the chains of primitive vertebrates (e.g. shark) and those of higher animals (e.g. man). The shark has light chains a) of similar size to those of higher vertebrates b) with a degree of electrophoretic heterogeneity equal to that of higher animals c) apparently with L and K light chain classes and d) with a degree of amino acid variability at the amino terminus of the SK region again comparable to that of higher vertebrates (109). Since the immune response has been observed only in vertebrates, and since the immune response is extremely complex even at the level of the lowest vertebrates, it may be that the evolutionary roots of this system extend into the invertebrate kingdom. A study of immunoglobulin evolution at the invertebrate level, of course, is dependent on finding a suitable assay system for primitive "immunoglobulin chains" which may have no classic immunoglobulin function. Immunologic detection (of C region) and assays for biologic functions other than antibody specificity (e.g. complement fixation) may provide the means to pursue this exciting area.

CHAPTER V.

ANTIBODIES: STRUCTURE, GENETICS AND EVOLUTION

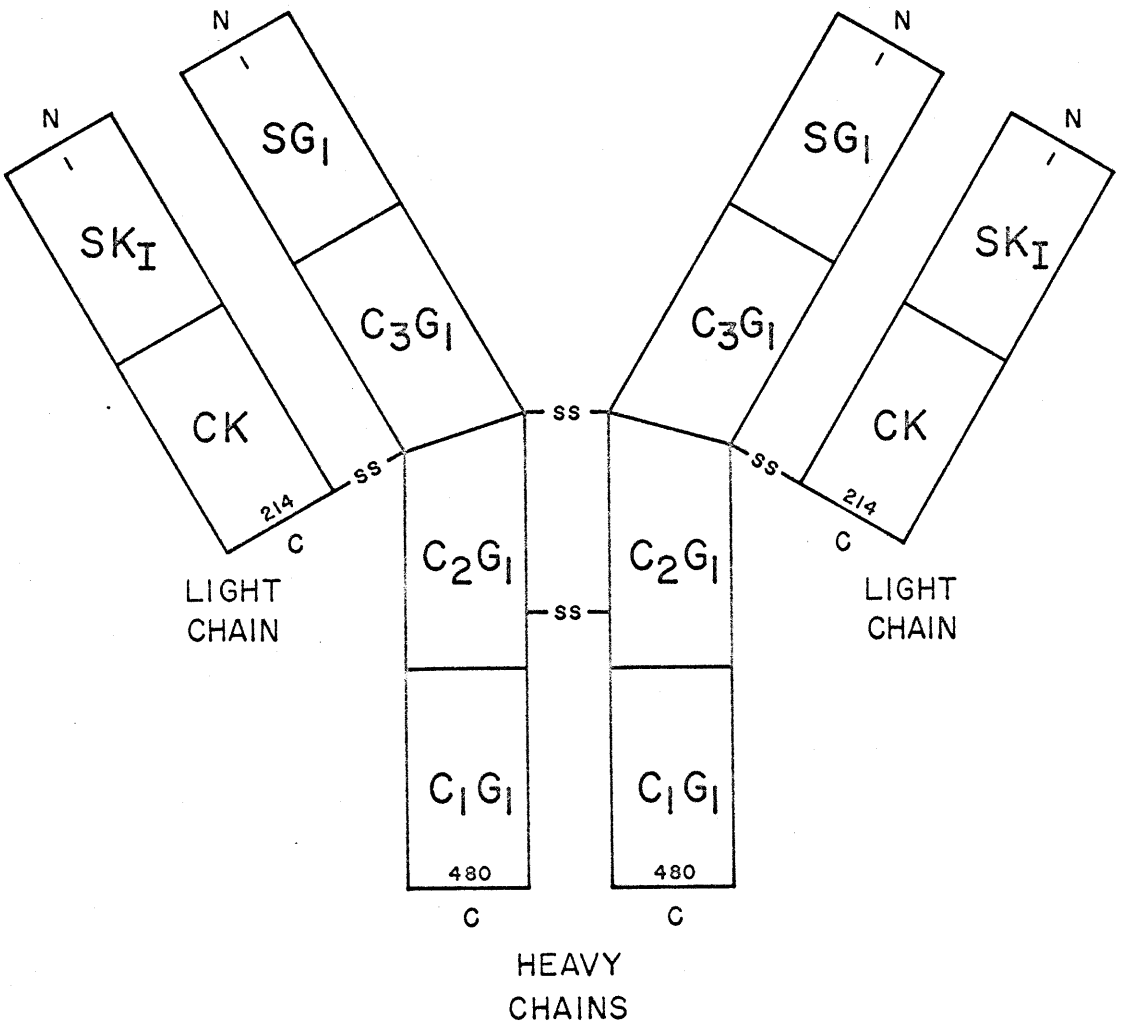
This chapter will be an attempt to integrate the observations made throughout this thesis with additional information on the structure of antibody molecules and to discuss the implications this picture has in regard to the genetic basis of antibody diversity and the evolution of immunoglobulin genes.

Antibody Specificity: The hallmark of the immune response is specificity, that is, the remarkable ability of antibodies to discriminate between closely related antigens. Since many different substances are antigenic, the immune system must be capable of synthesizing a large number of different antibodies, perhaps more than 10^5. Despite this large number of unique species, all antibody molecules have similar basic structures.

General Structure of Antibody Molecules (70)

Gross Structure: Three major types of antibodies (IgG, IgA, and IgM) occur in higher vertebrates. All immunoglobulins are multichain proteins containing light and heavy chains covalently joined through disulfide bonds (Figure 1). Each antibody type is defined by a particular class of heavy chain (G,A, or M). There are typically several members of each class; for example, G is represented by four subclasses- G_1, G_2, G_3 and G_4 which may all occur in the same individual (116). Hence the four G subclasses must be encoded by distinct cistrons and are not alternate alleles at a single locus. Common to all types of antibody are two classes of light chains, L and K. A light-heavy chain subunit seems to be the fundamental building block for all immunoglobulins from

Figure 1. Diagramatic model of IgG molecule. Light and heavy chains are joined through disulfide bonds. N and C indicate the N- and C-termini of these chains. The numbers adjacent to the carboxy terminus indicate approximate number of amino acid residues in each chain. The light chain is divided into SK and CK regions. The heavy chain may have three C regions of roughly equivalent size to the light C region (see text), C_1, C_2 and C_3 and N-terminal SG_1 region. This IgG molecule is of the SK_I - G_1 type.


N
SK$_I$
CK
214
C
SS
N
SG$_1$
C$_3$G$_1$
SS
C$_2$G$_1$
SS
C$_1$G$_1$
480
C
LIGHT CHAIN
N
SG$_1$
C$_3$G$_1$
SS
C$_2$G$_1$
SS
C$_1$G$_1$
480
C
N
SK$_I$
CK
214
C
LIGHT CHAIN
HEAVY CHAINS

those found in the most primitive of vertebrates (e.g. smooth dogfish shark) to those of higher forms (e.g. man)(117). Furthermore, each light-heavy chain pair usually forms a single antibody site to make the pair a functional as well as a structural subunit (118). The various types of antibody can be represented as a) IgG-$(KG)_2$ and $(LG)_2$; b) IgA-$(KA)_{2n}$ and $(LA)_{2n}$; and c) IgM $(KM)_{10}$ and $(LM)_{10}$. Hybrid molecules, combining different light or heavy chains (e.g. LKG_2), probably do not occur in vivo.

Biologic Function: The typical antibody molecule has a number of different biological functions which are in part associated with antibody type (70). Common to all major immunoglobulin types is, of course, antibody activity. The IgG molecule, for example, is also involved in transfer across the placental barrier and complement fixation (119, 120). Limited enzymatic cleavage (e.g. papain) of IgG disassociates the antibody activity from other functions and generates three enzymatic fragments (from the IgG molecule) — two identical Fab pieces and a single Fc piece (121). Each Fab piece (an intact light chain disulfide bonded to the N-terminal half of the heavy chain) is a univalent antibody, and under appropriate cleavage conditions can retain all of the combining activity of the intact IgG antibody (122). Other biologic functions (placental transfer and complement fixation) reside with the Fc piece (the C-terminal halves of the heavy chain pair joined through disulfide bonds) which is so homogeneous that it can be crystallized in some species (119, 120, 121).

Primary Sequence: The chemical heterogeneity of antibody (immunoglobulin) molecules has turned protein chemists to the study of the homogeneous products of multiple myeloma, a cancer of the antibody producing cells (plasma cells). Myeloma proteins appear to be normal immunoglobulins by physical, chemical, genetic and biologic properties (see Chapter 1). Amino acid sequence studies on myeloma light chains have demonstrated that the carboxy terminal half (common region, C) is essentially identical for all members of a given class and that the amino terminal half (specificity region, S) is unique to individual members of a class. Although little sequence data are yet available on heavy chains, peptide map studies indicate that they too can be divided into C and S regions (123). The C regions are certainly much larger than those of light chains (124). No definite information is available on the size of the heavy chain S region (see Figure 1).

Genetics of Common Regions

Separate genes must encode the common region of each type of immunoglobulin chain (i.e. CK, CL, CA_1, CA_2, CG_1, CG_2, CG_3, CG_4, CM) (116,125,126,127,128). The existence of C region genetic markers (see discussion, Chapter 3) has permitted linkage maps to be constructed for certain of these genes (89,129). For example, CG_1, CG_2 and CG_3 are closely linked in the human system (128) whereas two CG's and CA are closely linked in the mouse (130). Furthermore, there is no close linkage between the human CK gene (light) and the CG genes (heavy)(125). Further linkage data on the relationships among other heavy chain

C genes and on the relationship between CL and CK genes will await the discovery of suitable genetic markers in appropriate cistrons. Nevertheless, a cautious generalization can be drawn at this point: the C genes of immunoglobulins can probably be assigned to two independent linkage groups, the heavy chain set and the light chain set (this is similar to hemoglobin gene linkage, i.e. the "distinguishing" genes, β , δ , and γ , are closely linked to each other but not to the "common" gene, α). As we shall see in a moment, these linkage groups may play an important role in the control of immunoglobulin synthesis.

A cautionary note should be injected at this point. It has been tacitly assumed that each genetic marker is generated by a mutation in the gene encoding the primary structure of the polypeptide chains. While this is probably true for many of the immunoglobulin genetic variants (e.g. the Val and Leu found at CK (191)), other possibilities exist. For example, certain variants may be explained by modification of the chain after protein synthesis. In the rabbit system there is a set of common genetic markers (allotypic determinants) located on G,A, and M polypeptide chains (presumably encoded by non-allelic genes) (131). Such an unexpected result could most readily be explained by chain modification after protein synthesis (e.g. through the attachment of identical carbohydrate moieties).

Synthesis of Antibody Molecules

Even though each antibody producing cell potentially has the genetic information to produce many types of immunoglobulins (IgG, IgM,

etc), it appears that each immunocyte generally synthesizes a single molecular species of antibody. Although contrary results have been reported (132,133), the bulk of present evidence upholds the one cell: one antibody hypothesis:

1) In vitro experiments with antibody producing cells which have been stimulated in vivo with two or more antigens demonstrate that most cells make only a single type of antibody (134,135,136). Although occasional cells appear to be synthesizing two types of antibody, one must rule out nonspecific adsorption of antibody made by other cells and "engulfed" cytoplasm of other cells.

2) Myeloma proteins, presumed to be the products of single cell clones, generally have homogeneous light and heavy chains (90,137).

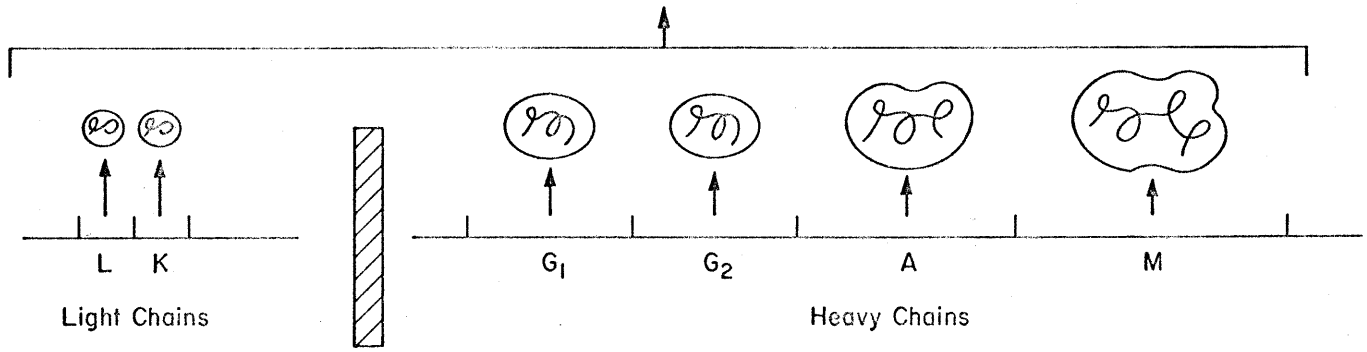
3) Immunofluorescent studies with type specific antisera show that individual lympoïd cells at a given point in time contain only one type of light and one type of heavy chain. These studies have been carried out in the human, rabbit, and mouse systems (138,139,140).

4) In rabbits heterozygous for allelic forms of a particular genetic variant, individual cells make only one of the two allelic forms (141).

These observations raise interesting questions concerning control mechanisms for immunoglobulin synthesis. Generally it seems that the genes of the differentiated antibody-synthesizing cell produce just a single type of light and a single type of heavy chain. That is, a single polypeptide chain is expressed from each of the two C region linkage groups (light and heavy chain genes) discussed in the previous section.

Figure 2. Diagramatic model of immunoglobulin common region gene system. Light and heavy chain genes are depicted in separate linkage groups. A number of heavy chain genes are closely linked; lacking suitable CL markers one can not be certain of light chain linkage; light and heavy chain genes are not closely linked. In most immunocytes only one type of light and one type of heavy chain can be synthesized.

IgG	IgA	IgM	Types
$(LG)_2$	$(LA)_{2n}$	$(LM)_{10}$	Molecules
$(KG)_2$	$(KA)_{2n}$	$(KM)_{10}$	



When the cell is heterozygous for genetic markers (e.g. CK (191) Val and Leu variants), only one of the two alleles is expressed. A similar phenomenon of "allelic exclusion" has been observed in genes located on the X chromosome of mammalian females; for example, in human cells heterozygous at the glucose-6-phosphate dehydrogenase allele, only one allele is active (142). Indeed, one of the X chromosomes is in a "contracted" and presumably repressed state. In contrast, however, both autosomal alleles controlling the synthesis of a hemoglobin chain (e.g. βS and βA) are expressed in each red blood cell (142). Immunoglobulin synthesis seems to be the first verified example of autosomal "allelic exclusion". Furthermore, these observations stress the point that the basis for the "heterogeneity of antibodies" is, in fact, a corresponding heterogeneity in antibody producing cells.

The introduction of a simple antigenic determinant into the immune system evokes an antibody response which is heterogeneous at three different levels:

1) Antibody <u>type</u> heterogeneity. There are at least 18 different types of human antibody (e.g. IgG_1, IgG_2, IgG_3, IgG_4, IgA_1, IgA_2, IgM, IgD, and IgE — each of which can have K or L light chains (70)).

2) Antibody <u>allele</u> heterogeneity. There are genetic variants in many of the antibody types (e.g. the Leu and Val alternatives at CK (191)).

3) Antibody <u>S region</u> heterogeneity. This heterogeneity is, of course, the essence of the immune response, and it implies that a) even a simple antigen may have many different antigenic sites and/or b) a single

antigenic site probably stimulates the production of many different antibody molecules with different binding affinities for this site. Obviously it is the S region heterogeneity which must be analyzed in order to obtain information as to the nature of the genetic mechanism responsible for the diversity of cells producing antibody.

Theories of Antibody Formation

In this section I will 1) consider in general terms the major theories of antibody formation 2) review the constraints which have been placed on these theories by light chain sequence data and 3) discuss specific theories for generating antibody diversity.

General Considerations: It will be helpful at this point to outline the various possibilities which exist for generating antibody variability:

I. Instructionistic Theories
   A. DNA level (replication)
   B. RNA level (transcription)
   C. Protein synthesis level (translation)
   D. Protein folding level

II. Selectionistic Theories
   A. Somatic Theories
      1. Replication level
      2. Transcription level
      3. Translation level
   B. Germ-line Theory

Instructionist theories postulate that the antigen must transfer information about the 3 dimensional nature of its antigenic sites to the antibody producing cell so that these cells can respond with the synthesis of complementary antibody. On the other hand, selectionistic theories hypothesize that the antigen need only select from a large population of immunocytes those cells which are already programmed to

synthesize complementary antibody. The "selected" cells are stimulated to divide and to synthesize their specific antibodies.

The instructionistic model can readily be visualized at the level of protein folding. The antigen permits the antibody molecule to fold around itself and thereby generate a complementary recognition site. This simple formulation is ruled out by at least two separate considerations:

1) Antibody fragments (Fab pieces) completely denatured in thiol-urea solutions can regain specific activity (i.e. refold into the native configuration) in the absence of antigen (143). The amino acid sequence, not the antigen, is responsible for specific antibody configurations.

2) The synthesis of a particular immunoglobulin is probably a stable and heritable trait for each differentiated immunocyte (45). This has been demonstrated in the author's experiments with the myeloma system where a single immunoglobulin species is produced throughout 100 generations of tumor passage. The stability of this system suggests that a permanent change must have occurred at the DNA (or RNA) level.

A mechanism whereby the instructionistic theory could operate at higher levels of protein synthesis (i.e. replication, transcription, or translation) is difficult to imagine. The three dimensional configuration of the antigenic site must be translated into a linear message (DNA or RNA); this information must be integrated in a stable fashion into the antibody synthesizing machinery, and finally an antibody molecule must be synthesized which is <u>complementary</u> to the original antigenic site. The arguments for one such theory have been discussed by Haurowitz (144) and will not be considered further here.

The greatest number of modern theories of antibody formation have been cast in a selectionistic framework (23,26,27,50,66,93,111,145,146, 147,148,149,150,151). These theories assume that each antibody cell is committed to the synthesis of a specific antibody even before the cell has seen the antigen. The critical point is to explain the molecular and genetic basis for antibody gene diversity which allows the organism to cope with an almost unbelievably large array of different antigens. Obviously any proposed theory must postulate many antibody genes — perhaps more than 10^5. These could arise from a single gene by some special process during somatic differentiation (somatic theory) or from separate germ-line genes which arose during the long course of chemical evolution (germ-line theory).

The germ-line theory is attractively simple:

1) Antibody genes are generated by the normal process of chemical evolution, i.e. gene duplication followed by mutation and selection. Those animals who meet the antigenic challenge of their environment survive to transmit their effective antibody genes to progeny; those who fail in this regard are eliminated (with their useless antibody genes).

2) Since each immunocyte is postulated to have the potential to generate all antibodies, differentiation requires the activation of particular antibody genes from among the many. This process is analogous to the differentiation of cells in other systems (e.g. red blood cells and lens fiber cells).

3) Since antibody genes can be highly selected by the normal process of chemical evolution, there is no need to depend on random somatic

mutational events in order to generate antibody specificities essential to organism survival.

Germ-line theories also raise a number of interesting questions:

1) How many different antibody molecules can the organism generate? How much DNA is required to encode these antibody genes? Can germ-line theories encode requisite antibody diversity without requiring a large percentage of the organism's DNA?

2) The transfer of large numbers of similar genes from an organism to its progeny raises questions concerning behavior of genetic material during meiosis. The Bar locus in Drosophila, the gene Hp^2 at the haptoglobin locus in man and the α and β hemoglobin genes in man are all tandem duplications. Each shows interchromosomal crossing over leading either to a return to the unduplicated form or to a triplication (152). The more extensive region of germ line genes postulated for antibody control might be expected to increase the probability of such events and lead to genetic instability.

Somatic theories postulate the generation of antibody variability at the level of gene replication, transcription or translation. These theories also have a number of attractive features:

1) Gene economy. A limited number of genes is responsible for antibody diversity. Less difficulty is encountered in the passage of genetic material from parent to progeny.

2) Extremely large numbers of antibody genes can be generated, starting from even a single gene.

This ability to produce antibody variability is, however, in turn a source of serious reservation about somatic theories. Are all of the mutant antibody genes (i.e. antibody cells) functional? This seems very unlikely. What percentage of the mutants produce functional antibodies? It appears that a somatic system could waste enormous numbers of antibody cells by generating nonfunctional mutants.

The amino acid sequence data have sharpened the focus of many of the points raised above and have allowed us to consider these numerous theories within the framework of certain molecular constraints.

Constraints of Immunoglobulin Structure

A study of amino acid variability in the S regions of light chains has provided a means for placing constraints on germ-line and somatic theories of antibody formation. Amino acid sequences of a number of light chains (and fragments thereof) have been determined and compared with one another to look for meaningful patterns of variation which might reflect the underlying genetic mechanism responsible for antibody diversity. As the genetic code dictionary is available, these light chain sequences can be translated into nucleotide sequences. Therefore, the search for meaningful patterns can be extended to the DNA level.

What is a meaningful pattern either at the amino acid sequence level or the nucleotide sequence level? If the S genes arose by a particular somatic mutational mechanism, there must be patterns reflecting that specific process imprinted on these proteins. For example, suppose we postulate that in-phase crossing over between two strands of DNA generates the variability seen in the S region of light chains. One

can then take the actual amino acid sequences, translate them to DNA sequences and determine whether or not all of the S region sequences can be generated by this simple model. Other simple models can be tested in a similar fashion.

On the other hand, if the S genes arose by normal chemical evolution, their phylogeny as determined by molecular level studies should be similar to that of evolutionarily related proteins such as cytochrome C or the hemoglobins.

What are the rules of phylogeny for evolutionarily related proteins (143)?

1) These proteins are generally of similar size and exhibit extensive areas of amino acid sequence homology when the sequences are properly aligned through the use of judiciously placed gaps (which represent sets of trinucleotide deletions at the DNA level). Sequence homologies are areas of amino acid sequence identity or conservative residue substitution (hydrophobic for hydrophobic: hydrophilic for hydrophilic; etc.).

2) The DNA is subjected to random evolutionary events, most of which are single base substitutions. These events are distributed randomly throughout the entire length of the DNA molecule (i.e. there is no gradient of variability). The base changes occur without respect to type; there are about twice as many transversions as transitions (the random chance for a transversional change is twice as great as for a transition because each purine can change to only one other purine but to either of two pyrimidines).

3) Evolutionary divergence has changed the duplicated genes of a single species (e.g. hemoglobin) from each other much as if they had been separated by speciation.

4) There is no single simple genetic mechanism which can account for the changes seen among evolutionarily related proteins.

There are now sufficient immunoglobulin sequence data (3,70,154) to search for the patterns mentioned above: five nearly complete K sequences (two mouse (92) and three human (67,94,96)), four nearly complete human L sequences (93,155,156), one Fc sequence from rabbit IgG (124) and many fragmentary sequences from light and heavy chains. These data support the following conclusions (see chapter 3):

1. There is no known single mutational or recombinational mechanism which can generate the observed sequences from a single germ-line gene. Models tested include reading frame shifts, in-phase crossing over between two genes, inversions and chromosomal rearrangments.

2. The pattern of amino acid variation is similar in all respects to that found in sets of evolutionarily related proteins. These sets obey all of the criteria for evolutionarily related proteins cited above. Indeed, an examination of the immunoglobulin sequences suggests that all immunoglobulin chains probably evolved from a common ancestor, half the size of the light chain (111,157,158). The few complete SK and SL comparisons which can be made show striking sequence homology which is fully substantiated by the extensive amino terminal data available (Figure 7, Chapter 3). Obviously, SK and SL genes have a common evolutionary origin. The S and C regions of light chains are of almost

identical size (which suggests a gene duplication), and furthermore the disulfide imposed ring structure in the two halves of the light chain is strikingly homologous (i.e. the ½-cystine residues are at comparable positions in the S and C regions) (111,158). The G heavy chain is about twice the size of the light chain and presumably is divided into four "evolutionary units," SG, C_3G, C_2G, and C_1G respectively from the N to C terminus (Figure 1). The human CL and CK sequences display striking sequence homology between themselves and with the C_1G sequence of rabbit heavy chain, suggesting again a common ancestral gene (124,155). The C_2G region has limited homology with other C regions but has probably incurred a large deletion (96). Peptide fragments from human C_2G and C_1G regions also exhibit similar homologies with light chain C regions (86,159,160). Very little data are available on other CG regions or other classes of heavy chains; nevertheless, current sequence information does suggest that all immunoglobulin chains may have originated from a gene encoding a polypeptide chain about 100 residues in length.

3. Each light chain is encoded by two genes (e.g. SK and CK) which are expressed as a single continuous polypeptide chain.

The human SK sequences can be aligned in at least two discrete sets termed SK_I and SK_{II} (Figure 7, Chapter 3). Hence, two separate germ-line genes or families of genes must be responsible for encoding the S regions of K chains. On the other hand, the CK region must be encoded by a single gene as alternative molecular forms of this gene (i.e. CK (191) Val or Leu) behave as alleles(88). If the SK regions are encoded by a minimum of two germ line genes and if the CK region is

encoded by a single gene (which can be associated with either SK gene), then each light chain must be encoded by two genes (SK and CK) which are expressed as a single continuous polypeptide chain. This will, presumably, be true of all immunoglobulin chains (e.g. G,A,M,L, etc.) by arguments of symmetry. Furthermore, this same general picture will probably be true in all higher vertebrates, for the general features of their light chains are similar to those of man and mouse.

If two distinct genes encode a single polypeptide chain, then some type of joining mechanism must exist for linking the separate bits of information. There are two general possibilities: either the S and C regions are linked at the informational level (i.e. DNA or RNA) or they are linked at the protein level.

Linkage at the informational level seems most likely because 1) polysomes involved in the synthesis of immunoglobulin are of appropriate size for heavy and light chains and not for smaller units (160) and 2) double-labeling experiments to determine the number of growing points on immunoglobulin polypeptide chains seem to indicate that at least the heavy chain is synthesized from a single point. Similar experiments with light chains are still equivocal (161,162).

The more remote possibility, namely that the S and C regions are linked at the protein level, has already been discussed by the author (113) and will not be considered further.

Specific Theories of Antibody Formation

Somatic Theories: In general terms, how do the constraints of the structural data affect these theories? 1) Some of the simplest

theories of this type can be unequivocally ruled out (e.g. simple hypermutation and in-phase crossing over between two strands of DNA). 2) Since discrete genes encode the S and C regions, every somatic theory must postulate two unusual genetic mechanisms - one to generate the sequence diversity of the S regions and a second to link the S and C regions. A consideration of specific somatic theories follows:

A. Specialized translational mechanisms (132,151): This type of theory suggests that an unusual codon exists at each variable position of the S gene. This codon is translated differently during protein synthesis, depending on which of a small number of activating enzymes is selected during immunocyte differentiation to charge the complementary transfer RNA. A number of objections arise to such a theory. 1) Are the special codons some of the redundant code words of the normal dictionary, or are they a supplementary set of words employing unusual nucleotide bases? There are at least 35 variable positions with unique amino acid alternatives. This seems like an excessive demand for unusual triplets. Presumably, this possibility could be tested experimentally. 2) The size difference in SK regions requires at least two genes with special codons for the K chain alone. 3) It is not obvious why, if a translation mechanism is the source of variability, the amino acid replacements are predominately single-base substitutions. There is no obvious reason to expect 80% of the variable amino acids to be related by single base changes.

This model could be approached experimentally. S-RNA could be isolated from several different myeloma tumors and characterized. Furthermore, the protein synthesizing machinery of one tumor could be

primed with the messenger from a second and the protein products characterized. Several workers are pursuing these and related experiments (163).

B. Transcriptional diversity: Variability could be generated by transcriptional errors of an RNA polymerase (3). The transcribed strand, like a RNA virus, would need to be replicated faithfully and passed on to progeny cells. Except for the need to postulate a "stable and heritable" RNA "virus," the results of somatic diversity generated at the transcriptal level are similar to those produced by somatic mechanisms at the DNA level, and both will be considered together in the next section.

C. Replication (DNA) diversity: A number of specific mechanisms have been proposed at this level. We shall consider three.

1. Errors in DNA repair (66,147). This model postulates that an enzyme recognizes a specific sequence of DNA in one strand and destroys a specific stretch of it (part or all of the S region gene). During normal repair, errors may arise with low frequency and new S genes are thereby generated.

A powerful argument can be leveled against this and, indeed, all forms of the simple hypermutational theory - the argument of cell wastage. In a random repair (or mutational) mechanism it would seem that the ratio of nonfunctional to functional antibody genes (i.e.antibody cells) produced would be immense.

This "error in DNA repair" mechanism would presumably make single base errors. On the average, five to seven different amino acids can

be generated from one codon by single base changes. If this somatic mechanism operated randomly then five to seven alternatives could be generated at each position in the region of variability. In point of fact, each position generally has three, two or even a single amino acid residue which is expressed in the light chain sequences. Hence at each position, "selection" must reduce the variability from five to seven possibilities down to three to two, or even one. It can be argued that somatic differentiation generates cells without any selection and that the antigen itself selects and activates those cells capable of producing complementary antibody. It seems, however, that cell wastage by such a process would be so great that the body would be essentially incapable of generating functional antibodies. For example, the region of variability is approximately 100 residues long and if the average codon can generate 6 different amino acids by single base changes, then the total number of different sequences which could be generated is 6^{100}. If one assumes that even three of the six alternatives at each site are functional, the ratio of nonfunctional to functional cells produced would be 2^{100}, or more than 10^{25} - more cells than are present in the entire organism (10^{14} for man).

Since the antibody-producing cells are probably committed to the synthesis of a particular immunoglobulin before they actually encounter the antigen, somatic theories require that variability be limited by a mechanism which is independent of the antibody molecule's ability to bind antigen. Somatic recombinational theories suggest a possible mechanism for limiting amino acid sequence variation.

2. Somatic recombination (148,149,150). Variability is postulated to be generated by repeated crossing-over among a small number of highly selected S genes. Since these S genes are highly selected, presumably the cell wastage argument is eliminated, for only those S genes capable of generating functional recombinant antibody sequences are passed on to progeny.

The simplest form of this theory (150), namely somatic crossing-over between two genes, can be ruled out with the sequence data at hand: 1) There are many residue positions in human light chains at which at least three different bases are required either in the first (SK (4,83,and 96)) or second (SK (91,96, and 100)) codon position. It is obviously impossible to generate three bases at one nucleotide position by crossing over between two strands of DNA. One can argue that human light chains are not a valid test of this proposition as the human population is probably polymorphic with respect to all genes (antibody genes included). This argument, however, is not valid in the BALB/c system, a highly inbred strain of mice. In this system there is at least one position (KS (4)) where three bases are required at the first position of the codon. 2) SK_I and SK_{II} sequences have characteristic features throughout the entire S region; - why are not these characteristic features randomized if all antibody diversity is generated by crossing over between two genes? 3) Many exceptions must be made at variable positions in the SK sequence in order to generate amino acid variability by recombination between two strands of DNA (e.g. 11 of 18 positions at the N-terminus of SK are variable. Eight of these

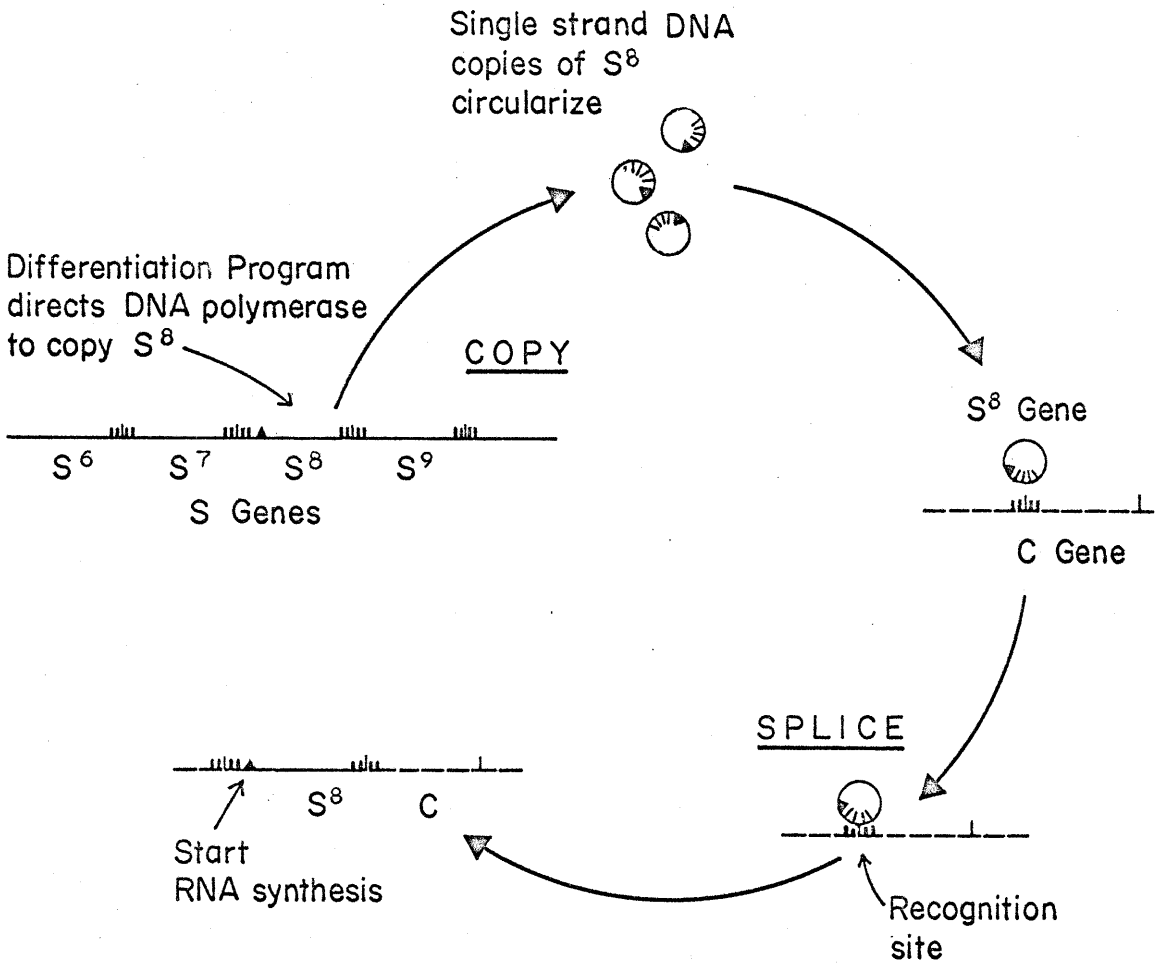
eleven variable positions have alternative amino acids which cannot be explained by a simple recombinational theory).

Clearly, more than two strands of DNA are necessary for a recombinational mechanism. Including additional genes in the recombinational mechanism can invalidate the above arguments. However, it leaves one with a model of antibody diversity which can explain almost any result and which does not suggest further experiments. The same is not true of an alternative hypothesis, the germ-line theory.

Germ-line theory (145,66,93,96): A striking constraint imposed by the sequence data is that immunoglobulin sequences behave as if they are evolutionarily related sets of proteins. An obvious solution to the antibody problem follows directly from this constraint, namely that for each species of antibody molecule there is a corresponding germ-line antibody gene. These genes arose by the normal process of chemical evolution; that is, gene duplication followed by mutation and selection.

The simplest formulation of this theory, namely that multiple genes exist for the entire antibody light chain, cannot be true because the C region is encoded by a single gene. Therefore, the germ-line theory must be modified to state that multiple germ-line genes encode each S region and a single germ-line gene encodes each C region. Many specific models can be proposed for such an hypothesis; one is demonstrated in Figure 3. The essential features of all such germ-line theories are as follows: 1) As it differentiates each immunocyte activates one S region, perhaps through a copy and splice mechanism which links the S gene to the C gene (Figure 3). 2) A few molecules

Figure 3. Copy-splice model of antibody differentiation. Multiple S genes are aligned in one or more chromosomes (perhaps in the heterochromatic regions which are generally repressed). The single C gene is probably not linked to the S genes. The "differentiation program" for each immunocyte directs a DNA polymerase to copy a particular specificity region (e.g. S^8). Multiple S^8 DNA copies are made and circularized. Each S^8 copy has a recognition site complementary to a C gene sequence (perhaps only 12 nucleotides in length - see text) which permits one S^8 copy to pair up and splice into the C gene to yield a continuous S-C nucleotide sequence. Antibody light chains are then transcribed and translated in the usual fashion. This model of differentiation suggests that information must be copied and spliced into a new region at the chromosome before it can be expressed (in contrast to the more classic derepression model which does not shuffle genes).


Single strand DNA
copies of S8
circularize
Differentiation Program
directs DNA polymerase
to copy S8
COPY
S6 S7 S8 S9
S Genes
S8 Gene
C Gene
SPLICE
S8 C
Start
RNA synthesis
Recognition
site

of specific antibody are synthesized and go to "trigger sites" (presumably on the immunocyte membrane) to serve as antigen receptors. Union of receptor antibody molecules and suitable complementary antigen will stimulate cell division and antibody synthesis.

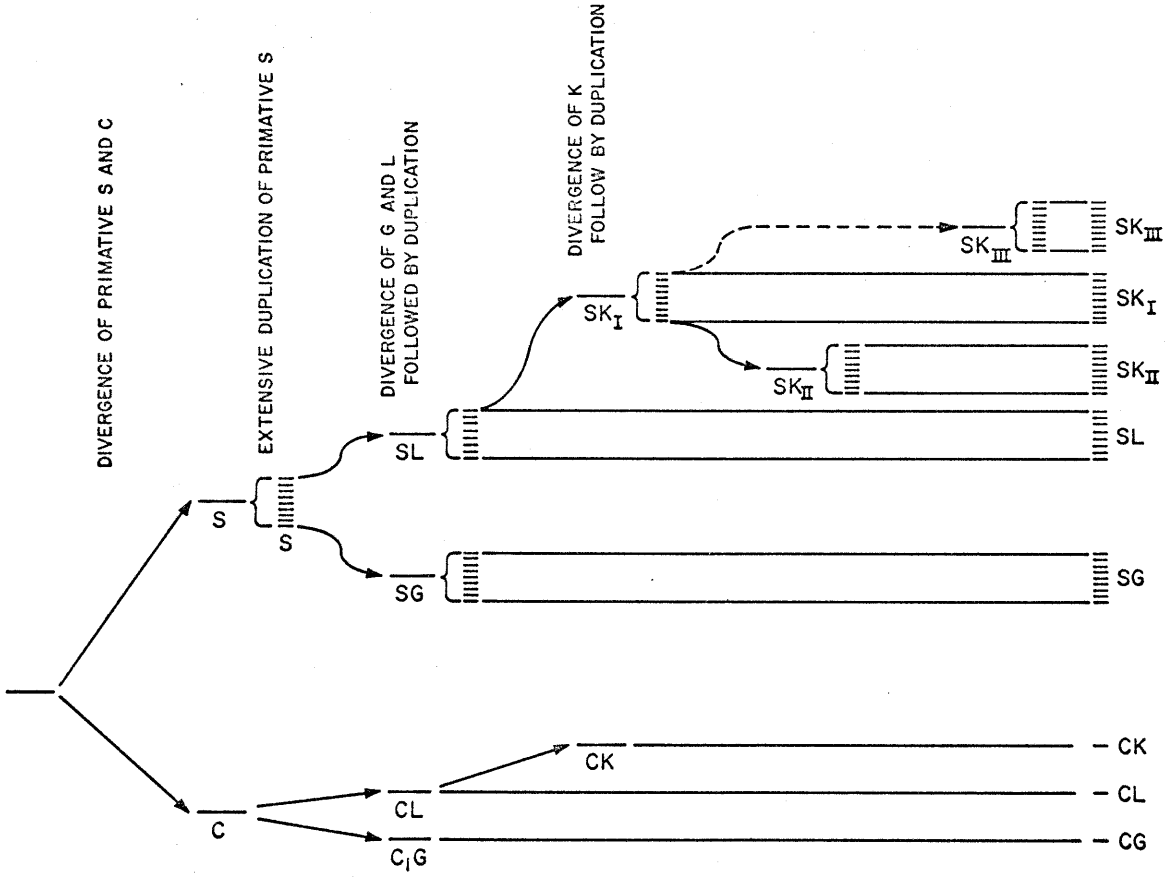
Since the C and S genes are presumably linked at the DNA level, they must have complementary recognition sites. Indeed, each SK gene must have the same recognition site to be able to combine with the single CK gene. Is there evidence for such a recognition site? It should be pointed out that such a recognition site need not be expressed in the light chain sequence itself. There is, however, a striking sequence invariance at SL (100-107). The same is true of the KS sequences in this region. These relatively invariant regions, different in the L and K proteins, could be the basis for discrete L and K recognition sites. Complementarity based on even 12 bases should be sufficient to promote the required pairing and splicing (164).

It is fair to ask at this point how such a multi-gene system might have evolved. A model is given in Figure 4.

The beginning. A primordial gene about 300 nucleotides in length (equivalent to the S and C genes of light chains) is duplicated to give a primordial S and a primordial C gene.

Formation of S gene pool. The S gene is duplicated repeatedly to form a large pool of specificity genes. This property of "hyperduplication" occurs only once in evolution as all other S genes will be selected from the original specificity pool and its descendents. The S pool is in constant state of flux; that is, it

Figure 4. Evolution of a multi-gene system. A primordial gene (encoding a polypeptide half the size of a light chain) duplicates to give primitive S and C genes. The S gene duplicates repeatedly generate an S gene pool. Complementary recognition sites are evolved (see text for details) and the divergence of present-day immunoglobulin types commences. The evolutionary tree as depicted in this figure is based on sequence comparisons which are discussed elsewhere (96,169).


DIVERGENCE OF PRIMATIVE S AND C
EXTENSIVE DUPLICATION OF PRIMATIVE S
DIVERGENCE OF G AND L FOLLOWED BY DUPLICATION
DIVERGENCE OF K FOLLOW BY DUPLICATION
S
S
C
SL
SG
CL
C I G
SK I
SK II
SK III
CK
SK III
SK I
SK II
SL
SG
CK
CL
CG

loses S genes and reduplicates new variants according to the laws of chemical evolution, i.e., mutation and natural selection of favorable genes.

Evolution of recognition sites. In time, one of the S genes (Sx) develops a recognition site which permits it to bind with and splice adjacent to the C gene. Presumably the product of this new gene pair confers a selective advantage; perhaps it is a pre-antibody which combines weakly with a foreign molecule(s) to render it less toxic, and natural selection leads to an increase of Sx frequency in the specificity pool. As Sx undergoes gene duplication and mutation (only the recognition site must remain invariant), a related set of genes is produced. Each can still combine with the original C region and each is capable of assuming a slightly different function (i.e., other pre-antibody specificities).

In time, a C gene duplication occurs which "frees" one of the two C genes and permits it to change its recognition site through normal mutational events. A process similar to that described above could engender a second set of S genes to match this new C gene recognition site. It is in this fashion that new types of immunoglobulin genes are evolved.

The germ-line theory described above predicts that there would be distinct sets of specificity regions for each major type of immunoglobulin (e.g. SL, SK, SG, SM etc.). The question arises as to whether the organism has sufficient DNA to handle this job. The following calculation can be made (165):

DNA content of human germ cell = 4×10^{-12} gm per cell
Therefore molecular weight = 24×10^{11}

Molecular weight of 1 base pair = 6.4×10^{2}
Therefore DNA content = 3.7×10^{9} base pairs per cell

1 specificity region = 107 amino acids
1 specificity region = 321 base pairs
Therefore 100,000 specificity gene = 3.2×10^{7} base pairs
Therefore 100,000 specificity gene = 1% of genetic material

Clearly, the organism has sufficient DNA to generate as many as 10^{4} S regions for each of 10 different classes of immunoglobulins.

Can one suggest experiments which might distinguish between the germ-line and the somatic hypotheses? These are at least three such experiments: 1) Many sequences. The demonstration of a thousand consecutive, unique light chain sequences would require an S region pool greater than the organism's DNA content. Perhaps this undertaking has become more feasible with the advent of the Edman machine (166,167). 2) DNA-RNA hybridization. One may be able to hybridize messenger from a myeloma tumor with the DNA from a second tumor cell (or perhaps with the DNA from normal immunocytes). If there are thousands of similar DNA copies, one should be able to detect them. This seems to be an experiment for the future because mammalian cell DNA-RNA hybridization is still an uncertain process (168). 3) Genetic marker in S region. If a genetic marker could be characterized in the S region, it would eliminate the germ-line theory from consideration. This appears to be the most substantial approach at this time.

In summary, the sequence data do place constraints on theories of antibody formation. The fact that light chain amino acid sequence variation is similar to that of evolutionarily related proteins does make the germ-line postulate attractive. On the other hand, the critical experiments which differentiate unambiguously between the germ-line and the somatic hypotheses have not yet been done.

ADDENDUM

The data in Chapter 2 have been published in a paper by J.C. Bennett, L. Hood, W.J. Dreyer & M. Potter (49).

The data in Chapters 3 and 4 have been published, in part, in three papers: L. Hood, W.R. Gray & W.J. Dreyer (113); L. Hood, W.R. Gray & W.J. Dreyer (114); and L. Hood, W.R. Gray, B.G. Sanders & W.J. Dreyer (91).

BIBLIOGRAPHY

1. W.C. Boyd. Fundamentals of Immunology. Interscience Publishers, Inc., New York. Chapter 3, 1956.

2. M. Sela. Advances in Immunology. 5, 30, 1966.

3. E.S. Lennox and M. Cohn. Annual Review of Biochemistry 36, 365, 1967.

4. R.K. Brown. Ann. N.Y. Acad. Sci. 103, 754, 1963

5. M. Fishman, J.J. van Rood, and F.L. Adler. In Molecular and Cellular Basis of Antibody Formation. J. Sterzl, editor. Publishing House of the Czechoslovak Academy of Science, Prague, 1966.

6. F.L. Adler, M. Fishman, and S. Dray. J. Immunology. 97, 554, 1966.

7. D.H. Campbell and J.S. Garvey. Advances in Immunology 3, 261, 1963.

8. R. Auerbach. In Organogenesis. R. DeHaan and H. Ursprung, editors. Holt, Rinehart and Winston, New York. Chapter 21, 1965.

9. J. Sterzl and A.M. Silverstein. Advances in Immunology 6, 161 1967.

10. T.N. Harris, K. Hummeler, and S. Harris. J. Exp. Med. 123, 161, 1966.

11. J.L. Gowans and D.D. McGregor. Progress in Allergy 9, 1, 1965.

12. R.W. Dutton and R.I. Mischell. J. Exp. Med. 126, 423, 1967.

13. I. MacKay and F.M. Burnet. In Autoimmune Disease. Charles C. Thomas Co., New York. Chapter 3, 1963.

14. P. Ehrlich. Proc. Roy. Soc. B. 66, 424, 1900.

15. F. Obermayer and E.P. Pick. Wein Klin. Wochschr. 17, 327, 1906.

16. K. Landsteiner. The Specificity of Serological Reactions. Rev. ed., Harvard University Press, Cambridge, 1947.

17. J. Alexander. Protoplasma. 14, 296, 1932.

18. S. Mudd. J. Immunology 23, 423, 1932.

19. R. Breinl and F. Haurowitz. Z. Phys. Chem. 192, 45, 1930.

20. L. Pauling. J. Am. Chem. Soc. 62, 2643, 1940.

21. F.M. Burnet and F. Fenner. The Production of Antibodies. MacMillan and Co., Ltd., Melbourne, 1949.

22. F.M. Burnet. Enzyme, Antigen and Virus. Cambridge University Press, Cambridge, 1956.

23. N. Jerne. Proc. Nat. Acad. Sci. U.S. 41, 849, 1955.

24. N. Jerne. Ann. Rev. Microbiol. 14, 341, 1960.

25. F.M. Burnet. Australian J. Sci. 20, 67, 1957.

26. F.M. Burnet. The Clonal Selection Theory of Acquired Immunity. University Press, Cambridge, 1959.

27. J. Lederberg. Science. 129, 1649, 1959.

28. E. Franklin. Progress in Allergy 8, 58, 1964.

29. S. Cohen and R.R. Porter. Advances in Immunology 4, 287, 1964.

30. A. Tiselius and E.A. Kabat. J. Exp. Med. 69, 119, 1939.

31. N.H. Eisen, E.S. Simms, J.R. Little, Jr., and L.A. Steiner. Fed. Proc. 23, 559, 1964.

32. J.F. Heremans and J.P. Vaerman. Nature 193, 1091, 1962.

33. H.A. Sober and E.A. Peterson. Fed. Proc. 17, 116, 1958.

34. O. Kratky, G. Porod, A. Sekora, and B. Paletta. J. Polymer Sci. 16, 163, 1955.

35. J. Almeida, B. Cinader, and A. Howatson. J. Exp. Med. 118, 327, 1963.

36. G.M.Edelman and J.A.Gally. J. Exp. Med. 116, 207, 1962.

37. P.A. Small, J.E. Kehn, and M.E. Lamm. Science 142, 393, 1963.

38. R.H. Pain. Biochem. J. 88, 234, 1963.

39. J.B. Fleischman, R.H. Pain, and R.R. Porter. Arch. Biophys. Suppl. 1, 174, 1962.

40. R.R. Porter and R.C. Weir. J. Cell Phys. 67, Suppl. 1, 51, 1966.

41. M. Potter, J.L. Fahey and H. Pilgrim. Proc. Soc. Exper. Biol. & Med. 94, 327, 1957.

42. F.W. Putnam, New England J. Med. 261, 902, 1959.

43. J.L. Fahey. Advances in Immunology 2, 42, 1962.

44. M. Potter and C.R. Boyce. Nature 193, 1086, 1962.

45. M. Potter, W.J. Dreyer, E. Kuff, and K.R. McIntire. J. Mol. Biol. 8, 814, 1964.

46. M. Mannik and H.G. Kunkel. J. Exp. Med. 117, 213, 1963.

47. J.L. Fahey and A. Solomon. J. Clin. Invest. 42, 811, 1963.

48. R.W. Putnam. Biochim. et Biophys. Acta. 63, 539, 1962.

49. J.C. Bennett, L. Hood, W.J. Dreyer, and M. Potter. J. Biol. Biol. 12, 81, 1965.

50. M. Cohn. Cold Spring Harb. Sym. Quant. Biol. 32, in press, 1967.

51. N.H. Eisen. Cold Spring Harb. Sym. Quant. Biol. 32, in press, 1967.

52. L. Hood and W.J. Dreyer. Unpublished observations.

53. M.E. Adams-Mayne and B. Jirgensons. Fed. Proc. 23, 474, 1964.

54. H.G. van Eijk, C.H. Monfoort, and H.G.K. Westenbrink. Koninlk. Nederl. Akad. von Wetenschappen. 66, 345, 1963.

55. M. Sogami and J.F. Foster. J. Biol. Chem. 237, 2514, 1962.

56. A.M. Katz, W.J. Dreyer, and C.B. Anfinsen. J. Biol. Chem. 234, 2897, 1959.

57. R.J. Block, E.L. Durrun, and G. Zweig. A Manual of Chromatography and Paper Electrophoresis. Academic Press, New York. p. 484, 1955.

58. I. Smith. Nature 171, 43, 1953.

59. O. Ouchterlony. Acta. Path. Microbiol. Scan. 25, 186, 1948.

60. J.J. Scheidegger. Intern. Arch. Allergy. 7, 103, 1955.

61. K.A. Piez and L. Morris. Anal. Biochem. 1, 187, 1960.

62. C.H.W. Hirs, S. Moore, and W.H. Stein. J. Biol. Chem. 219, 623, 1956.

63. G.L. Ellman. Arch. Biochem. Biophys. 82, 70, 1959.

64. B.J. Davis. Ann. N.Y. Acad. Sci. 121, 404, 1964.

65. J.W. Keyser. Anal. Biochem. 9, 249, 1964.

66. W.J. Dreyer and J.C. Bennett. Proc. Nat. Acad. Sci. U.S. 54, 864, 1965.

67. N. Hilschmann and L.C. Craig. Proc. Nat. Acad. Sci. U.S. 53, 1403, 1965.

68. K. Titani, E. Whitley, Jr., L. Avogardo, and F.W. Putnam. Science 149, 1090, 1965.

69. K. Titani, E. Whitley, Jr., and F.W. Putnam. Science 152, 1513, 1966.

70. S. Cohen and C. Milstein. Advances in Immunology. In press, 1967.

71. V. Kostka and F.H. Carpenter. J. Biol. Chem. 239, 1499, 1964.

72. R. Canfield and C.B. Anfinsen. J. Biol. Chem. 238, 2684, 1963.

73. C. Niu and H. Fraenkel-Conrat. J. Am. Chem. Soc. 77, 5882, 1955.

74. J.C. Bennett and W.J. Dreyer. Personal communication.

75. W.J. Dreyer and E. Bynum. Methods in Enzymology 11, in press, 1967.

76. H.N. Rydon and P.W.G. Smith. Nature 169, 922, 1952.

77. F. Sanger. Adv. Prot. Chem. 7, 1, 1952.

78. M.A. Naughton, F. Sanger, B.S. Hartley, and D.C. Shaw. Biochem. J. 77, 149, 1960.

79. J. Schultz. Discussion to F. Sanger. J. Polymer Sci. 49, 3, 1961.

80. W.R. Gray. Methods in Enzymology 11, in press, 1967.

81. W.R. Gray and J. Smith. In preparation.

82. P. Edman. Ann. N.Y. Acad. Sci. 88, 602, 1960.

83. R.F. Doolittle. Biochem. J. 94, 742, 1965.

84. J. Sjöquist. Biochim. Biophys. Acta. 41, 20, 1960.

85. J.L. Bailey. Techniques in Protein Chemistry. Elsevier Pub. Co. New York. p. 134, 1962.

86. E.M. Press, D. Givol, P.J. Piggot, R.R. Porter, and J.M. Wilkinson. Proc. Roy. Soc. B. 166, 150, 1966.

87. R.U. Eck and M.O. Dayhoff. Atlas of Protein Sequence and Structure 1966. National Biomedical Research Foundation. Silver Spring, Maryland.

88. G. Ropartz, J. Lenoir, and L. Rivat. Nature. 189, 586, 1961.

89. J. Oudin. J. Cell. Physiol. 67, Suppl. 1, 77, 1966.

90. W.R. Gray. Proc. Roy. Soc. B. 166, 146, 1966.

91. L. Hood, W.R. Gray, B.S. Sanders, and W.J. Dreyer. Cold Spring Harb. Sym. Quant. Biol. 32, in press, 1967.

92. W.R. Gray, W.J. Dreyer, and L. Hood. Science 155, 465, 1967.

93. M. Wikler, K. Titani, T. Shinoda, and F.W. Putnam. J. Biol. Chem. 242, 1668, 1967.

94. F.W. Putnam, K. Titani, and E. Whitley, Jr. Proc. Roy. Soc. B. 166, 124, 1966.

95. C. Milstein. Proc. Roy. Soc. B. 166, 138, 1966.

96. W.J. Dreyer, W.R. Gray, and L. Hood. Cold Spring. Harb. Sym. Quant. Biol. 32, in press, 1967.

97. H.G. Kunkel and R.J. Slater. J. Clin. Invest. 31, 677, 1952.

98. R.A. Reisfeld and P. Small. Science 152, 1253, 1966.

99. M.A. Raftery and R.D. Cole. J. Biol. Chem. 241, 3457, 1966.

100. R.E. Offord. Nature 211, 591, 1966.

101. L.W. Clem and P.A. Small. J. Exp. Med. 125, 893, 1967.

102. J. Marchalonis and G.M. Edelman. J. Exp. Med. 124, 901, 1966.

103. S. Cohen and R.R. Porter. Advances in Immunology 5, 287, 1964.

104. A. Feinstein. Nature 210, 135, 1966.

105. A. Wilson. Personal communication, 1966.

106. V. Nussenzweig, M.E. Lamm, and B. Benacerraf. J. Exp. Med. 124, 787, 1966.

107. E. Zuckerkandl and L. Pauling. In Evolving Genes and Proteins, V. Bryson and H.J. Vogel, editors. Academic Press, New York. p. 97, 1964.

108. L. Hood and G.M. Edelman. Unpublished results, 1967.

109. A. Suran and B. Papermaster. Cold Spring Harb. Sym. Quant. Biol. 32, in press, 1967.

110. V. Nussenzweig and B. Benacerraf. J. Exp. Med. 124, 805, 1966.

111. S.J. Singer and R.F. Doolittle. Pauling Festschr. In press, 1967.

112. R.F. Doolittle, personal communication, 1967.

113. L. Hood, W.R. Gray, and W.J. Dreyer. Proc. Nat. Acad. Sci. U.S. 55, 826, 1966.

114. L. Hood, W. R. Gray, and W.J. Dreyer. J. Mol. Biol. 22, 149, 1966.

115. R.F. Doolittle. Proc. Nat. Acad. Sci. U.S. 55, 1195, 1966.

116. W.D. Terry, J.L. Fahey, and A.G. Steinberg. J. Exp. Med. 122, 1087, 1965.

117. J. Marchalonis and G.M. Edelman. Science 154, 1567, 1966.

118. A. Nisonoff, M.H. Winkler and D. Pressman. J. Immunol. 82, 201, 1959.

119. F.W.R. Brambell, W.A. Hemmings, C.L. Oakley, and R.R. Porter. Proc. Roy. Soc. B. 151, 478, 1960.

120. A. Taranla and E.C. Franklin. Science 134, 1981, 1961

121. R.R. Porter. Biochem. J. 73, 119, 1959.

122. A. Nisonoff, F.C. Wissler, L.N. Lipman, and D.L. Woernley. Arch. Biochem. Biophy. 89, 230, 1960.

123. M. Potter, E. Appella, S. Geisser. J. Mol. Biol. 14, 361, 1965.

124. R.L. Hill, R. Delaney, H.E. Lebovitz, and R.E. Fellows, Jr. Proc. Roy. Soc. B. 166, 159, 1966.

125. A. Steinberg. Prog. Med. Genet. 2, 1, 1962.

126. W.D. Terry and M.S. Roberts. Science 153, 1997, 1966.

127. M. Harboe, J. Deverill, and H.C. Godal. Scand. J. Haemat. 2 137, 1965.

128. H.G. Kunkel and J.B. Natvig. Cold Spring Harb. Sym. Quant. Biol. 32, in press, 1967.

129. J. Oudin. Proc. Roy. Soc. B. 166, 207, 1966.

130. L.A. Hertzenberg. Cold Spr. Harb. Symp. Quant. Biol. 29, 455, 1964.

131. C. Todd. Biochem. Biophy. Res. Commun. 11, 170, 1963.

132. G. Attardi, M. Cohn, K. Horibata, and E.S. Lennox. J. Immunol. 92, 335, 1964.

133. R.N. Hiramoto and M. Hamlin. J. Immunol. 94, 214, 1965.

134. G.J.V. Nossal and O. Makela. Ann. Rev. Microbiol. 16, 53, 1962.

135. O. Makela and G.J.V. Nossal. J. Immunol. 87, 447, 1961.

136. G.J.V. Nossal. J. Exp. Path. 39, 544, 1958

137. M.J. Waxdal, W.H. Konisberg, and G.M. Edelman. Gold Spr. Harb. Sym. Quant. Biol. 32, in press, 1967.

138. B. Pernix and G. Chiappino. Immunology 7, 500, 1964.

139. E. Weiler. Proc. Nat. Acad. Sci. U.S. 54, 1765, 1965.

140. B. Pernis and G. Chiappino. Nature 211, 424, 1966.

141. B. Pernis, G. Chiappino, A.S. Kelus, and P.G.H. Gell. J. Exp. Med. 122, 853, 1965.

142. E. Beutler. Cold Spr. Harb. Sym. Quant. Biol. 29, 261, 1964.

143. E. Haber. Proc. Nat. Acad. Sci. U.S. 52, 1099, 1964.

144. F. Haurowitz. Nature 205, 847, 1965.

145. L. Szilard. Proc. Nat. Acad. Sci. U.S. 46, 293, 1960.

146. O. Smithies. Nature 199, 1231, 1963.

147. S. Brenner and C. Milstein. Nature 211, 242, 1966.

148. G.M. Edelman and J.A. Gally. Proc. Nat. Acad. Sci. U.S. 57, 353, 1967.

149. H.L.K. Whitehouse. Nature 215, 371, 1967.

150. O. Smithies. Science 157, 267, 1967.

151. J. Campbell. J. Theor. Biol. In press, 1967.

152. O. Smithies. Cold Spr. Harb. Sym. Quant. Biol. 29, 304, 1964.

153. T. Jukes. Molecules and Evolution. Columbia University Press, New York, 1966.

154. Most recent sequence data are summarized in Cold Spr. Harb. Sym. Quant. Biol. 32, in press, 1967.

155. F.W. Putnam, K. Titani, M. Wikler, and T. Shinoda. Cold Spr. Harb. Sym. Quant. Biol. 32, in press, 1967.

156. C. Milstein, J.B. Clegg, and J.M. Jarvis. Nature 214, 270, 1967.

157. S.J. Singer and R.F. Doolittle. Science 153, 13, 1966.

158. S. Cohen and C. Milstein. Nature 214, 449, 1967.

159. N.O. Thorpe and H.F. Deutsch. Immunochemistry. 3, 329, 1966.

160. A.L. Shapiro, M.D. Scharff, J.V. Maizel, and J.V. Uhr. Proc. Nat. Acad. Sci. U.S. 56, 216, 1966.

161. E.S. Lennox, P.M. Knopf, A.J. Munro, and R.M.E. Parkhouse. Cold Spr. Harb. Sum. Quant. Biol. 32, in press, 1967.

162. J.B. Fleischman. Cold Spr. Harb. Sym. Quant. Biol. 32, in press, 1967.

163. B. Mach. Cold Spr. Harb. Sym. Quant. Biol. 32, in press, 1967.

164. C.A. Thomas, Jr. Prog. Nucl. Acid Res. and Mole. Biol. 5, 315, 1966.

165. Calculation by W.R. Gray. DNA value from The Chemical Basis of Heredity, edited by W.D. McElroy and B. Glass, Johns Hopkins Press, Baltimore, 1957.

166. P. Edman and G. Begg. European J. Biochem. 1, 80, 1967.

167. H. Niall and P. Edman. Nature, in press, 1967

168. H.C. Birnboim, J.J. Pene, and J.E. Darnell. Proc. Nat. Acad. Sci. U.S. 58, 320, 1967.

169. W.R. Gray, W.J. Dreyer and L. Hood. In preparation.