INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

76-283

WASSERMAN, Richard Lawrence, 1948-
STRUCTURE AND EVOLUTION OF IMMUNOGLOBULIN
HEAVY CHAIN VARIABLE REGIONS.

The City University of New York, Ph.D., 1975
Health Sciences, immunology

Xerox University Microfilms, Ann Arbor, Michigan 48106

STRUCTURE AND EVOLUTION OF IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGIONS

by

Richard L. Wasserman

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements of the degree of Doctor of Philosophy, The City University of New York.

1975

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

19 August 1975
date

Chairman of Examining Committee

August 19, 1975
date

Executive Officer

Supervisory Committee

The City University of New York

ACKNOWLEDGEMENTS

The author gratefully acknowledges his debt to Drs. J. Michael Kehoe, Lawrence E. Mole, Pierre Querinjean and David G. Klapper for their teachings and to Dr. J. Donald Capra for being much more than a teacher.

DEDICATION

Dedicated to my wife, Tina Deborah Wasserman, whose support makes anything possible.

TABLE OF CONTENTS

I. INTRODUCTION — page

A. Immunoglobulin function.......... 1

B. Immunoglobulin structure.......... 2

C. Antibody diversity: Theoretical.......... 6

D. Antibody diversity: The Data.......... 9

1. Nucleic acid hybridization studies.......... 10

2. Inherited idiotypes.......... 10

3. Variable region allotypes.......... 12

4. Amino acid sequence studies.......... 13

II. MATERIALS AND METHODS

A. Immunoglobulins.......... 19

B. Purification of proteins and peptides.......... 19

1. Zone electrophoresis.......... 19

2. Salt fractionation.......... 21

3. Ion-exchange chromatography.......... 21

4. Molecular exclusion chromatography.......... 26

5. Separation of peptides by high voltage paper electrophoresis.......... 27

6. Separation of peptides by differential solubility.......... 27

C. Evaluation of protein purity.......... 27

1. Cellulose acetate electrophoresis.......... 27

2. Polyacrylamide gel electrophoresis.......... 28

3. Ouchterlony double immunodiffusion analysis.......... 28

4. Immunoelectrophoresis.......... 28

D. Preparation and quantitation of heavy chains.................... 28
1. Partial reduction and alkylation............................ 28
2. Complete reduction and alkylation........................... 29
E. Fragmentation of myeloma proteins.............................. 30
1. Cyanogen bromide cleavage of heavy chains................... 30
2. Cyanogen bromide cleavage of whole proteins................. 30
3. Tryptic digestion of cyanogen bromide fragments............. 30
F. Automated amino acid sequencing................................ 31
1. Liquid phase automated sequencing........................... 31
2. Identification of phenylthiohydantoin (PTH) derivatives of amino acids................................. 31
G. Amino acid analysis.. 32
III. RESULTS
A. Sequence studies of normal pooled sera......................... 34
B. The sequence of the variable region of a canine alpha chain..... 48
1. Digestion of canine alpha chain Gom with cyanogen bromide... 48
2. Digestion of IgAκ dimer with cyanogen bromide............... 51
3. The sequence of the amino terminal cyanogen bromide fragment of Gom alpha chain................................ 51
4. The sequence of the second cyanogen bromide fragment of the variable region of alpha chain Gom.................. 54
5. The sequence of the third variable region cyanogen bromide fragment of alpha chain Gom......................... 58
C. The sequence of the variable region of a canine mu chain........ 64
1. Digestion of IgMλ pentamer with cyanogen bromide............ 64

2. The sequence of the amino terminal cyanogen bromide fragment of the variable region of mu chain Moo.. 70

3. The sequence of the second cyanogen bromide fragment of the variable region of mu chain Moo........... 73

4. The sequence of the third cyanogen bromide fragment of the variable region of mu chain Moo........... 76

IV. DISCUSSION

A. Sequence studies of normal pooled IgG heavy chains............. 83

B. The amino acid sequence of the variable regions of two homogeneous canine immunoglobulins........................ 90

C. Conclusions.. 102

APPENDIX... 107

REFERENCES.. 108

LIST OF TABLES

page

II-1. Source and Nature of Serum Pools Used for the Preparation of IgG from Various Species........................ 20

III-1. Calculation of the Unblocked V_HIII Subgroup................... 35

III-2. Amino Acid Composition of Gom CB-1............................ 57

III-3. Amino Acid Composition of Gom CB-2 and Its Constituent Peptides... 61

III-4. Amino Acid Composition of Gom CB-3 and Its Variable Region Peptides.. 67

III-5. Amino Acid Composition of Moo CB-2 and Its Constituent Peptides... 79

LIST OF FIGURES

page

I-1. Structure of IgG.. 4

I-2. Prototypic sequences of the V_HIII subgroup of immunoglobulin heavy chains derived from the sequence of myeloma proteins.. 16

II-1. Chromatography of IgAκ Gom on DEAE Sephadex.................. 22

II-2. Chromatography of IgMλ Moo on DEAE Sephadex.................. 24

III-1. Gas chromatographic analysis of step 2 of the Edman degradation of pooled immunoglobulin heavy chains............. 37

III-2. Phylogenetic distribution of the V_HIII subgroup of immunoglobulin heavy chains among mammals..................... 40

III-3. Gas chromatographic analysis of the PTH derivative at position 23 of two myeloma heavy chains and of heavy chain pools from the same species............................ 43

III-4. Amino acid sequence analysis of the V_HIII subgroup of pooled mammalian and avian immunoglobulin heavy chains........ 45

III-5. Chromatography of cyanogen bromide peptides of alpha chain Gom.. 49

III-6. Chromatography of cyanogen bromide peptides of IgAκ Gom....... 52

III-7. Sequence analysis of fragment CB-1 of alpha chain Gom......... 55

III-8. Sequence analysis of fragment CB-2 of alpha chain Gom......... 59

III-9. Chromatography of tryptic peptides of Gom CB-3................ 62

III-10. Sequence analysis of the variable region portion of fragment CB-3 of alpha chain Gom............................. 65

III-11. Chromatography of cyanogen bromide peptides of IgMλ Moo....... 68

III-12. Chromatography of completely reduced and carboxymethylated cyanogen bromide fragments of IgMλ Moo............ 71

III-13. Sequence analysis of fragment CB-1 of mu chain Moo........... 74

III-14. Sequence analysis of fragment CB-2 of mu chain Moo........... 77

III-15. Sequence analysis of the variable region portion of fragment CB-3 of mu chain Moo................................ 81

IV-1. Complete amino acid sequence of two canine heavy chain variable regions....................................... 91

IV-2. Amino terminal sequences of canine myeloma and pooled heavy chains.. 93

IV-3a,b,c. Heavy chain variable region sequences of seven human and two canine myeloma proteins of the V_HIII subgroup...95,96,97

I. INTRODUCTION

Elucidation of the mechanisms of information accumulation, storage and utilization in living systems is one of the primary goals of modern biology. The vertebrate immune system is capable of specific recognition and response to a myriad of molecular species and is therefore an example of a biologic information processing apparatus. As such, immunology is a subject which has moved beyond gross empirical observation to rational, factually based explanations of the reaction of vertebrate organisms to alien molecules.

Exposure of an animal to antigen, a chemical structure which is not a normal body constituent of that animal, elicits a predictable reaction that is specific for the foreign substance. This reaction may be purely cellular, resulting in the activation or deactivation of lymphocytes, or it may result in the production of proteins, secreted by stimulated cells, which interact with the antigen, independent of cells. The antibody response to antigen is the most fully characterized feature of the immune response. Antibody proteins are physiologically active gene products; directly correlated with DNA, the chemical carrier of biologic information. Study of these immunoglobulin molecules provides data that helps define the information accumulation, storage, and utilization mechanisms of the immune system.

A. Immunoglobulin function

Immunoglobulins form a set of glycoproteins that have the ability

to bind other molecules with a high degree of specificity. Antibodies bind specifically to the antigens which induce them; other immunoglobulins (i.e. myeloma proteins) presumably have specificity for some antigen and differ from antibody only with respect to the mechanisms of induction and synthetic control (Kunkel et al. 1963; Grey and Kunkel, 1964; Kunkel and Prendergast, 1966).

In vivo, antibodies perform the coordinated functions of binding to foreign molecule and eliminating it as a threat to the host. The antigen binding function is specific; antibody will only combine with the inducing molecular structure or antigenic determinant. The specificity is a function of the conformation of part of the antibody molecule and is determined by the primary structure of the protein. Methods of eliminating the foreign substance as a threat to the host are called effector functions. These functions are accomplished by portions of the molecule distant from the antibody combining site (Porter, 1959; Kehoe et al., 1969; Yasmeen et al., 1973). The particular effector function is determined by the structure of the antibody and occurs after the reaction of antigen with antibody. Antigen binding and effector functions are defined by protein structure; the genetic information for these activities and their physiologic function are related by antibody structure.

B. Immunoglobulin structure

The functional immunoglobulin subunit is the four chain monomer, consisting of two identical 22.5 Kdal light chains and two identical 50 Kdal heavy chains (Edelman, 1959; Fleischman et al, 1962). The four chain subunit exists on its own or in higher order oligomers (Tomasi, 1972). In most cases one light chain is disulfide bonded to each heavy chain and

the heavy chains are disulfide bonded to each other (Fig. I-1).

X-ray crystallographic studies (Poljak et al., 1973; Padlan et al., 1973) have confirmed the hypothesis (Edelman and Gall, 1969) based on amino acid sequence studies that the immunoglobulin molecule consists of homologous, folded, globular domains linked by short stretches of non-folded chain. Each domain consists of a disulfide loop of approximately 100 residues contributed by two of the four chains. The fundamental pattern is a series of anti-parallel lengths of polypeptide chain arranged in beta-pleated sheet configuration. The characteristic feature of this structure is the immunoglobulin fold (Poljak, et al., 1974) which allows the formation of this tertiary structure. Particular attention has been focused on the domain which is formed by the amino terminal portions of the light and heavy chains; this domain has been isolated and shown to be the site of antigen binding (Inbar, et al., 1972; Hochman, et al., 1973). The X-ray crystallographic model of this domain shows a cleft or groove formed by the folding of several portions of each chain. This structural feature of the amino terminal domain was predicted by several other lines of investigation on the antibody combining site (Capra and Kehoe, 1975).

As the earliest sequence data on immunoglobulin chains accumulated, the unique characteristic of these polypeptides became clear. Bence-Jones proteins of a particular antigenic type were identical (except for minor genetic variants) in the carboxyterminal portion of the molecule, but each protein differed markedly from the others in the amino terminal half (Hilschmann and Craig, 1965; Titani et al., 1965; Milstein, 1966; Putnam et al., 1966). This observation has been extended to light and

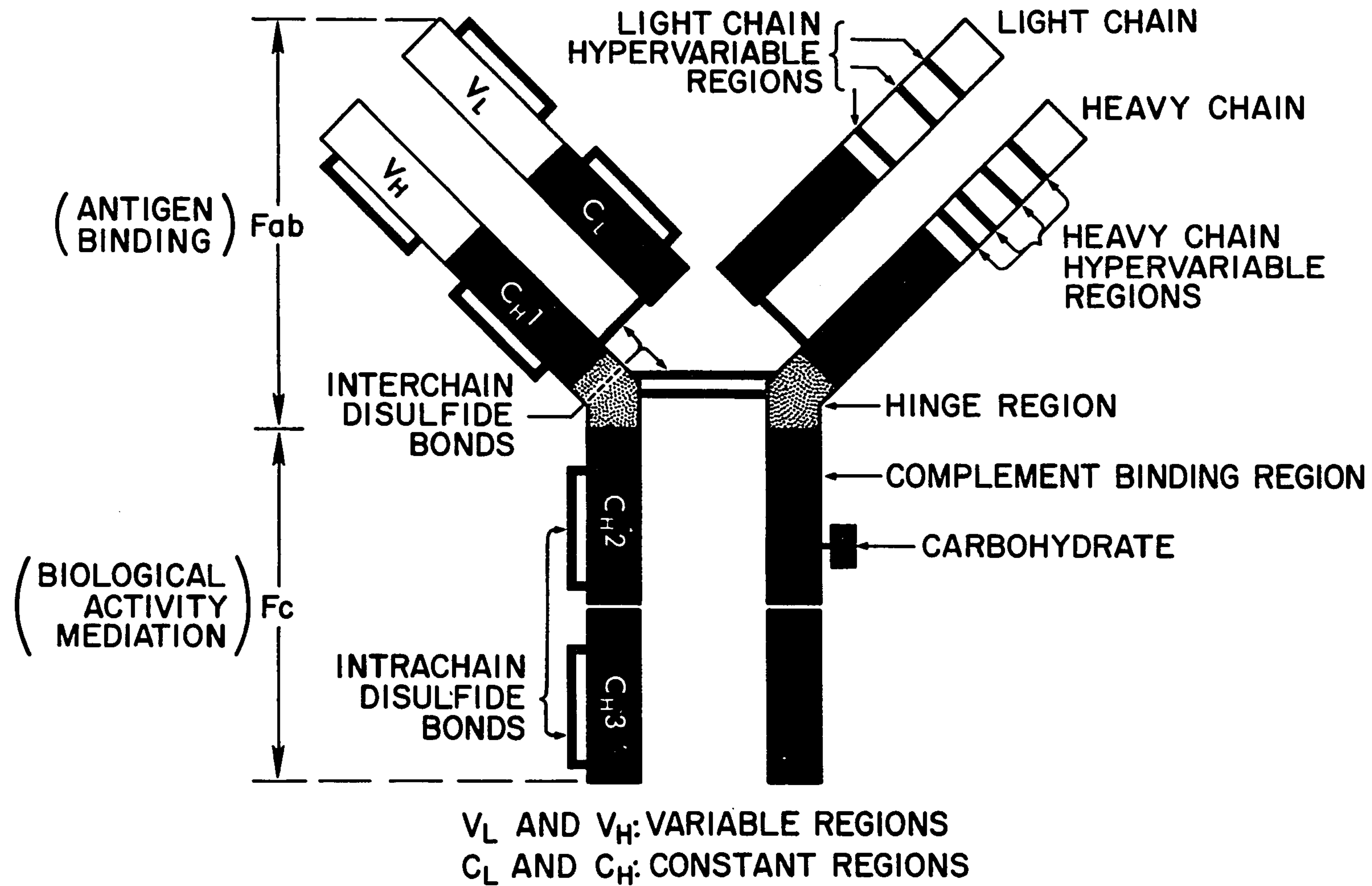

LIGHT CHAIN HYPERVARIABLE REGIONS
LIGHT CHAIN
HEAVY CHAIN
V_L
V_H
C_L
C_H1
(ANTIGEN BINDING) Fab
HEAVY CHAIN HYPERVARIABLE REGIONS
INTERCHAIN DISULFIDE BONDS
HINGE REGION
COMPLEMENT BINDING REGION
CARBOHYDRATE
C_H2
C_H3
(BIOLOGICAL ACTIVITY MEDIATION) Fc
INTRACHAIN DISULFIDE BONDS
V_L AND V_H: VARIABLE REGIONS
C_L AND C_H: CONSTANT REGIONS

FIGURE I-1

Structure of IgG

heavy chains derived from intact immunoglobulin molecules. The variability of the amino terminal domain of light and heavy chains was a powerful structural correlate for the functional variability of antibodies as antigen binding proteins. Extensive investigation of the primary structure of light and heavy chain variable regions led to the definition of areas of hypervariability (Milstein, 1967; Wu and Kabat, 1970; Franek, 1971; Capra and Kehoe, 1974a). Concurrent affinity labeling studies (Ray and Cebra, 1972) showed that amino acids, in or near hypervariable regions, could be affinity labeled by reactive haptens and therefore were in close proximity to the antigen binding site. The antibody combining site, part of the amino terminal domain of light and heavy chains, postulated to be related to hypervariable regions on the basis of sequence and affinity labeling studies, has been shown by X-ray crystallography to consist of hypervariable regions of both chains arranged in juxtaposition to form the interactional complement of antigen.

C. *Antibody diversity: Theoretical*

The finding that different immunoglobulin molecules had different sequences laid to rest the various instructional theories of antibody formation. Since each antibody had a different sequence and reverse transcriptase (Temin and Baltimore, 1972), let alone a "reverse translatease" had not yet been discovered, antigen could not act as a template for antibody formation. The Clonal Selection Theory of Burnet (Burnet, 1959), which predicted, in part, that specific antigen responsive cells preexist antigenic challenge, provided the most widely accepted mechanism of antibody formation. While Burnet's Theory postulated random generation of diversity during the early differentiation phase of embryonic life, the other components of the theory required only that specific anti-

gen reactive cells would be present before antigen. No molecular model for the generation of this diversity was provided.

Many different protein sequences require many different RNA messengers and many RNA messengers require many different DNA genes; but, are the many genes carried in the germ line of the animal or are they the products of differentiation of lymphoid cells during ontogeny? _Ab initio_, it is most productive to resolve this question into its purest form: Is there a germline gene for each different light and heavy chain variable region sequence, or is there a single gene for the light chain variable regions (and one for the heavy chain variable regions) which replicates inexactly to produce the many genes which code for the light and heavy chain variable regions of different sequence? The first explanation requires many genes and is therefore called the Multi-gene Theory and the second requires few genes and is called the Pauci-gene Theory. Each theory leads to predictions about the humoral immune response and each requires _ad hoc_ mechanisms to bridge the gap between the empirical findings and principles of genetics derived from other systems. Additionally, there are findings which are logical consequences of both theories.

At the level of the individual, the Multi-gene Theory is the simplest and most straightforward. Each member of a species, as a zygote, carries in its genome a full complement of variable region genes, perhaps 1000 V_H genes and 1000 V_L genes (Hood and Talmage, 1970; Smith, 1973). During ontogeny these genes gradually become expressed. All, or almost all, combinations of V_H and V_L are structurally possible and 10^6 specificities may be available. The ability to combine with antigen is accounted for by direct complementarity of the 10^6 specificities with all possible determi-

nants or by cross reaction. Multi-gene Theory suggests that the many genes evolved by gene duplication and subsequent mutation in a manner analogous to other proteins. Corollaries to this explanation are the expectations that: 1) Closely related individuals will have many identical variable region genes, these genes being inherited in a Mendelian fashion. 2) Identical myeloma proteins will be found from two different individuals. 3) There will be many genes in the germline which are maintained by positive selection pressure despite the high genetic load. 4) Within the constraints imposed by tertiary structural requirements, mutations resulting in new combining sites will carry a positive selection pressure; all other mutations will be random. 5) Diversity itself is an intrinsic positive selection pressure and new combining sites are maintained in the genome without contributing to the individual's survival. 6) Expression of a combining specificity occurs at early commitment to antigen; presentation of a new combining specificity by a clone involves repression of the original gene and derepression of another gene. On preliminary examination, therefore, the Multi-gene or Germline approach appears conservative; its predictions are compatible with the traditional findings and explanations of molecular biology.

The Pauci-gene Theory requires, at a fundamental level, a radical departure from accepted explanations of gene expression. Several mechanisms have been proposed by which antibody diversity may be generated from a few germline genes. Features common to these mechanisms are that they begin with a few genes which are altered at each mitotic cell division, thereby producing many clones of cells carrying different specificities. The multitude of combining sites are a product of cell differentiation.

Although each variation of Pauci-gene Theory is singular, they are all designed to account for antibody variability, differing in their processes, and some of their predictions. The features of the major models will be considered together.

Pauci-gene mechanisms augur the following: 1) Few genes are required for the immune system and consequently the genetic load on the organism is light. 2) All possible amino acid sequences have a finite probability of occurrence, limited only by the requirements of tertiary and quarternary structure. The sequences closest to the germline gene are most probable. 3) Two individuals may construct the same variable region. The likelihood of detecting such identical antibodies is greatly enhanced by pre-selection for specificity. 4) If gene duplication occurs there will be little or no selection pressure to maintain many germline genes. 5) Mutations of the germline genes, irrelevant to the antibody combining site, may occur and will be expressed on all variable regions in a method analogous to other protein polymorphisms. 6) Pooled antibodies and myeloma proteins, considered together, will, outside the hypervariable regions, behave as though they were encoded by one gene.

D. Antibody diversity: The Data

Theoretical constructs, developed to account for a body of data, must remain consistent with the empirical findings. Theories of antibody diversity are generated to explain the extensive functional and structural heterogeneity inherent in the humoral immune system, a characteristic for which there is no biologic precedent. Diversity is, therefore, the fundamental element. And the first consideration is the relationship of the data to diversity and then to the theory.

1. Nucleic acid hybridization studies

A priori, merely counting the number of genes would appear to be the simplest way of resolving the issue. Several investigators have approached this question by means of hybridization studies of mouse myeloma nucleic acids (Premkumar *et al*., 1974; Tonegawa *et al*., 1974; Leder *et al*., 1974). The results of these experiments are difficult to evaluate and, when taken together, are irreconcilable. Ignoring technical disagreements, the key question, which is basic to any interpretation of these studies, deals with the sequence similarity of nucleic acids which do hybridize and the sequence differences sufficient to prevent hybridization. Even variable regions of different subgroups and with different combining specificities are homologous. And it is clear that proteins of identical sequence may be encoded by DNA of different sequences (Murgola and Yanofsky, 1974). Recognizing these limitations the conclusions of the workers in this field differ widely. At present, nucleic acid hybridization studies have not resolved the issue.

2. Inherited idiotypes

Idiotypic determinants provide an experimental link between antibody combining site structure and function. Studies demonstrating the hapten inhibitable nature of the reaction of an idiotypic antiserum and the antihapten antibody (Brient and Nisonoff, 1970; Sher and Cohn, 1972) provide strong evidence that the idiotypic determinants are intimately related to the antibody combining site. Sequence studies of two anti γ-globulins with shared idiotypy have shown that, while these proteins vary as much as two unselected myeloma proteins in the framework portion of the variable regions, the hypervariable regions are almost identical (Capra

and Kehoe, 1974b). Thus, idiotypy has been shown to reflect both the binding function and the primary structure of the variable region.

Several investigators, working independently, have identified markers, based on idiotypic cross reactions, in the serum of mice immunized with specific haptens (Kuettner et al., 1972; Blomberg et al., 1972; Pawlak et al., 1973; Eichmann and Berek, 1973; Lieberman et al., 1974). These markers, which appear to be detectable in essentially all appropriately immunized animals of a particular strain, are completely absent from other strains. Mating experiments have shown that these markers are inherited in a simple Mendelian fashion. This is obviously most compatible with a germline model in which each combining site is encoded by a separate gene. These observations provide powerful support for Multi-gene Theory.

Inherited idiotypes have been attributed to critical germline genes by the proponents of somatic mutation models (Cohn et al., 1974). Such an explanation becomes increasingly less tenable as more inherited idiotypic markers are identified. It has been pointed out that many of the attractive features of the somatic mutation theory dissipate under this burden (Capra and Kindt, 1975). If, however, somatic mutation is considered as an ongoing process, not antigen induced but antigen driven (Cunningham and Pilarski, 1974), the appearance of inherited combining site markers may be rationalized. A few basic genes are carried in the germline. On interaction with antigen a clone derived from one of these genes is stimulated. As it proliferates it undergoes continuous mutation. The mutations leading to more favorable combining sites are selected by increased interaction with antigen. By the time that anti-

body secretion begins a single clone may predominate. Since all animals of the same strain start with the same basic genes and the diversity is generated and selected for by antigen, there is a high likelihood that the same combining site will be expressed. In another strain the same antigen may stimulate several clones, none of which originated from the same gene as in the situation described above. Therefore, no combining sites would share idiotypy between the two strains.

3. Variable region allotypes

Since it is not yet possible to count the number of variable region genes directly (e.g.: by hybridization) the best approximation may be obtained by examining the protein products of these genes. Polymorphism is a significant feature of many human enzymes and may be the rule rather than the exception (Harris, 1970). Such polymorphisms may be detected electrophoretically (Harris, 1970) or by means of antigenic analysis (Natvig and Kunkel, 1973). Examination of these polymorphisms has established the number of genes involved in several enzyme and immunoglogulin constant region systems. One of the most extensively studied protein groups is human immunoglobulins (Natvig and Kunkel, 1973; Milstein et, al., 1974). Antigenic analysis of human constant regions has defined the arrangement and number of genes coding for each family of proteins. The genetic markers, called allotypes, have been shown to be grouped into sets of alleles, each at a given locus. Linkage of most loci has been demonstrated and the amino acid sequence differences responsible for the antigenic differences have been identified.

A completely analogous system exists among rabbit immunoglobulins. The phenomenon of allotypy occurs more frequently in the rabbit than in the human and has been discovered in variable as well as

constant regions. Discovery of an allotypic marker, unrelated to light chain, but shared by IgM (Todd, 1963) and IgA (Feinstein, 1963) as well as IgG raised the possibility that a portion of heavy chain, common to all classes and therefore related to several loci, might be encoded by another locus. Early consideration of these findings focused on their implications for the one gene-one polypeptide theory but the significance of variable region allotypes to theories of antibody diversity is clear.

An axiom of classical genetics holds that only variants at a single locus can behave as alleles. In other protein systems polymorphisms which behave as alleles represent variants of one gene. Variants which are not allelic are the products of two or more loci which probably arose by gene duplication. During evolution such tandem loci may drift and become unrelated proteins; they may experience many crossovers and create a markedly polymorphic family of genes with widely varying gene frequencies or the polymorphism may be lost entirely as the interactions between adjacent genes obscure their differences. Three loci for rabbit V_H genes have been identified by allotypic analysis. While linkage studies have demonstrated crossovers between V_H and C_H genes, intragenic V_H crossover has never been described. From a genetic point of view, rabbit heavy chain variable regions behave as though they were encoded by one gene.

4. Amino acid sequence studies

Variations among proteins detectable antigenically or electrophoretically are a function of amino acid sequence differences. Despite the fact that the sequence of a protein is only an imperfect mirror of the nucleotide sequence from which it was decoded, the amino acid sequence is a precise and easily examined correlate of a gene. Many of the contributions sequence studies made to the understanding of antibody diversity

have been discussed. The first immunoglobulin sequences available initiated the controversy over the generation of diversity. Sequence studies of antibodies bearing an inheritable cross reacting idiotype have helped to structurally define that phenomenon (Capra _et al_., 1975). Allotypic markers in the rabbit have been shown to correlate with sequence differences in the V_H region (Mole _et al_., 1971). Thus, sequence studies of immunoglobulins bear directly on diversity, the number of gene products and the behavior of these gene products from a genetic point of view.

The earliest variable region sequence studies were performed on myeloma light chains and resulted in three significant observations. First, each protein had a different sequence in the variable region. Second, proteins could be grouped based on sequence similarities in the amino terminal twenty residues. Third, the degree of variation was variable and there were areas of extremely high variation and areas that were relatively invariant. The first and third observations have been discussed but the second observation is crucial to an understanding of the information generated by sequence studies. It is widely accepted that there is at least one germline gene for each variable region subgroup. Therefore, sequence studies of pooled immunoglobulin in which more than one subgroup is represented are not informative. There are many human, murine and rabbit light chain subgroups, only some of which have a cyclized amino terminus. (A protein with a cyclized amino terminus cannot be sequenced by routine automated techniques.) Therefore, the results of sequence studies of pooled light chains have been ambiguous because of multiple subgroups. In contrast, pooled porcine lambda light chains have a homogeneous amino terminal sequence and reflect a single subgroup

(Novotny _et al_., 1972). Comparison of the sequence of pooled porcine lambda chains with murine and human myeloma light chains showed that there were positions at which one species had a characteristic amino acid not present in the others.

Examination of several canine and feline myeloma heavy chains (Kehoe and Capra, 1972) showed them all to be homologous to the V_HIII subgroup of human heavy chains (Kohler _et al_., 1970). In addition, chains from each species possessed certain residue positions at which all proteins of one species had a particular amino acid while all proteins of another species had a different amino acid. These residue positions were termed phylogenetically associated since, as shown in Fig. I-2, a given substitution was not specific for a particular species but occurred within a phylogenetic group. These studies differed from previous attempts to find species specific residues in light chain variable regions in that only a single variable region subgroup was considered. Thus, phylogenetically associated residues are amino acid substitutions at a given position within a single subgroup that are peculiar to a Linnaen group. The fact that the V_HIII subgroup is only a minor constituent of human heavy chains and eight out of eight canine and feline myelomas were of this subgroup suggested that the distribution of variable region subgroups might be different in different species. Such a finding would be germane to the question of the role of subgroups in antibody diversity and to the question of the genetic mechanism of diversity itself. Since phylogenetically associated residues have been detected in other proteins and have been extensively used to draw conclusions concerning the evolution and genetics of those proteins, development of such a system in immunoglobulins could

	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	26	27	28	29	30
					5					10					15					20					25					
HUMAN	GLU	VAL	GLN	LEU	VAL	GLU	SER	GLY	GLY	GLY	LEU	VAL	GLN	PRO	GLY	GLY	SER	LEU	ARG	LEU	SER	CYS	ALA	ALA	SER	GLY	PHE	THR	PHE	SER
MURINE			LYS																LYS											
FELINE	ASP									ASP											THR		VAL							
CANINE										ASP													VAL							

FIGURE I-2

Prototypic sequences of the V_HIII subgroup of immunoglobulin heavy chains derived from the sequence of myeloma proteins (Kehoe and Capra, 1972). Only differences from the human are shown.

relate immunoglobulin variable regions to other proteins on a genetic level.

The studies reported herein are an examination of the V_HIII subgroup on two distinct levels. A survey of the subgroup distribution and a search for phylogenetically associated residues was made to determine if evolutionary patterns could be established for pooled immunoglobulin variable regions as they had been for other proteins. A more restricted but more complete examination of the V_HIII subgroup was made in the dog by determining the complete primary structure of the variable regions of two canine monoclonal proteins. Information derived from the survey provides the data necessary to compare immunoglobulin variable regions to other proteins. The sequences of two canine heavy chain variable regions, by comparison with sequences of heavy chain variable regions of other species, define the location of phylogenetically associated residues within the heavy chain variable region. Together, these studies allow conclusions to be drawn concerning the mechanisms for generating the sequence variability in heavy chain variable regions.

II. MATERIALS AND METHODS

A. Immunoglobulins

1. Normal pooled sera were obtained as indicated in Table I-1.

2. Sera containing canine paraproteins Gom and Moo were a gift from Dr. A. Hurvitz, Animal Medical Center, New York, New York. Gom was obtained as fresh frozen plasma from a plasmaphoresis of an Airedale terrier with multiple myeloma. Moo was fresh frozen plasma obtained by plasmaphoresis from a Scottish terrier with lymphosarcoma.

B. Purification of proteins and peptides

1. Zone electrophoresis

The monoclonal protein fractions of plasma from dogs Gom and Moo were initially isolated by zone electrophoresis using Pevikon (Pevikon C-870, Mercer Chemical Corp.), a copolymer of polyvinyl chloride and polyvinyl acetate, as the supporting medium (Kunkel, 1954). Pevikon was suspended in Barbital buffer, pH 8.6, ionic strength 0.1 and undiluted serum was applied to a 1 cm thick block by loading at a ratio of 0.3 ml/cm. Electrophoresis was carried out at 4°C by applying a constant current of 3 ma/cm to the block for 19 hours. After electrophoresis, 3 strips of filter paper 0.5 cm x 50 cm were laid across the block and then removed and stained with bromphenol blue. The portion of the block which contained the myeloma protein was removed and the protein eluted from the

Table II-1

Source and Nature of Serum Pools Used for the Preparation of IgG from Various Species

Animal	Immunoglobulin pool size and source		
	Number	Starting Material	Source
MOPC 173 (Mouse)	Pool of >10	Ascitic fluid	M. Potter
MOPC 21A (Mouse)	Pool of >10	Ascitic fluid	M. Potter
MOPC 40 (Mouse)	Pool of >10	Ascitic fluid	M. Potter
Tei (Human IgG1)	1	Plasma	Fresh frozen
Taf (Human IgG2)	1	Plasma	Fresh frozen
Lev (Human IgG3)	1	Plasma	Fresh frozen
Man	Pool of >10	Serum	Fresh frozen
Monkey(cynomolgus)	1	Serum	Fresh frozen
Monkey (rhesus)	Pool of >10	Serum	GIBCO*
Monkey (African Green)	Pool of >10	Serum	GIBCO
Dog	Pool of 3	Serum	Fresh bleeding
Cat	Pool of 3	Serum	Fresh bleeding
Cow	Pool of 4	Serum	Fresh bleeding
Goat	Pool of 4	Serum	Fresh bleeding
Pig	Pool of 10	Serum	GIBCO
Sheep	1	H chains	K. Dorrington
Horse	1	Plasma	Fresh bleeding
Opossum	Pool of >5	Serum	R. Genco
Guinea pig	Pool of 3	Serum	Rockefeller strain
Hamster	Pool of >10	Lyophilized serum	Rockland+
Rat-Sprague/Dawley	Pool of 5	Serum	T. Barka
Mink	Pool of 5	Serum	A. Hurvitz
Mouse	Pool of >10	Lyophilized serum	GIBCO
Rabbit	Allotypically balanced pool	Serum	T. Kindt
White whale	1	Serum	N.Y. Aquarium
Fin whale	1	Serum	N. Arnheim
Moose	1	Serum	N. Arnheim
Seal	1	Serum	N. Arnheim
Sea lion	1	Serum	N. Arnheim
Chicken	Pool of >10	Serum	GIBCO
Turkey	Pool of >10	Lyophilized serum	Rockland
Pigeon	Pool of >10	Lyophilized serum	Rockland
Duck	1	Serum	H. Grey

*Grand Island Biological Co., Grand Island, N. Y.

+Rockland, Gilbertsville, Pa.

Pevikon by washing with 0.05M sodium phosphate buffer, pH 7.85.

2. Salt Fractionation

The immunoglobulin fraction of serum from several species was initially isolated by successive precipitation with Na_2SO_4 (Stelos, 1967). The gamma globulin fraction was precipitated by adding Na_2SO_4 to a final concentration of 18% and redissolved in borate buffered saline, pH 8.0. The procedure was repeated with 14% and 12% Na_2SO_4.

3. Ion-exchange chromatography

a. Purification of immunoglobulins on Sephadex DEAE-A50

Zone electrophoretic isolates of canine IgMλ Moo and IgAκ Gom were applied to 2.5 x 100 cm columns of Sephadex DEAE A50 anion exchanger equilibrated in 0.05M sodium phosphate buffer, pH 7.85. After washing the column until the 280 OD returned to baseline, a gradient of 0 to 1.0M NaCl in starting buffer was applied to the column. Typical elution profiles are shown in Figures II-1 and II-2.

IgG was separated from pooled serum or plasma from several species by ion exchange chromatography on Sephadex DEAE A50. Generally, 30 ml of serum or its lyophylized equivalent was diluted 1:1 with 0.05M sodium phosphate buffer, pH 7.85, and applied to a 1.5 x 90 cm column of Sephadex DEAE A50 equilibrated with the same buffer. The column was then washed with 120 ml of the starting buffer, and 180 ml of eluate collected.

b. Separation of peptides on Sephadex DEAE A50 in dissociating conditions

Sephadex DEAE A50 equilibrated with 0.05M sodium phosphate buffer, pH 7.85, with 6M urea was used to separate certain peptide mixtures.

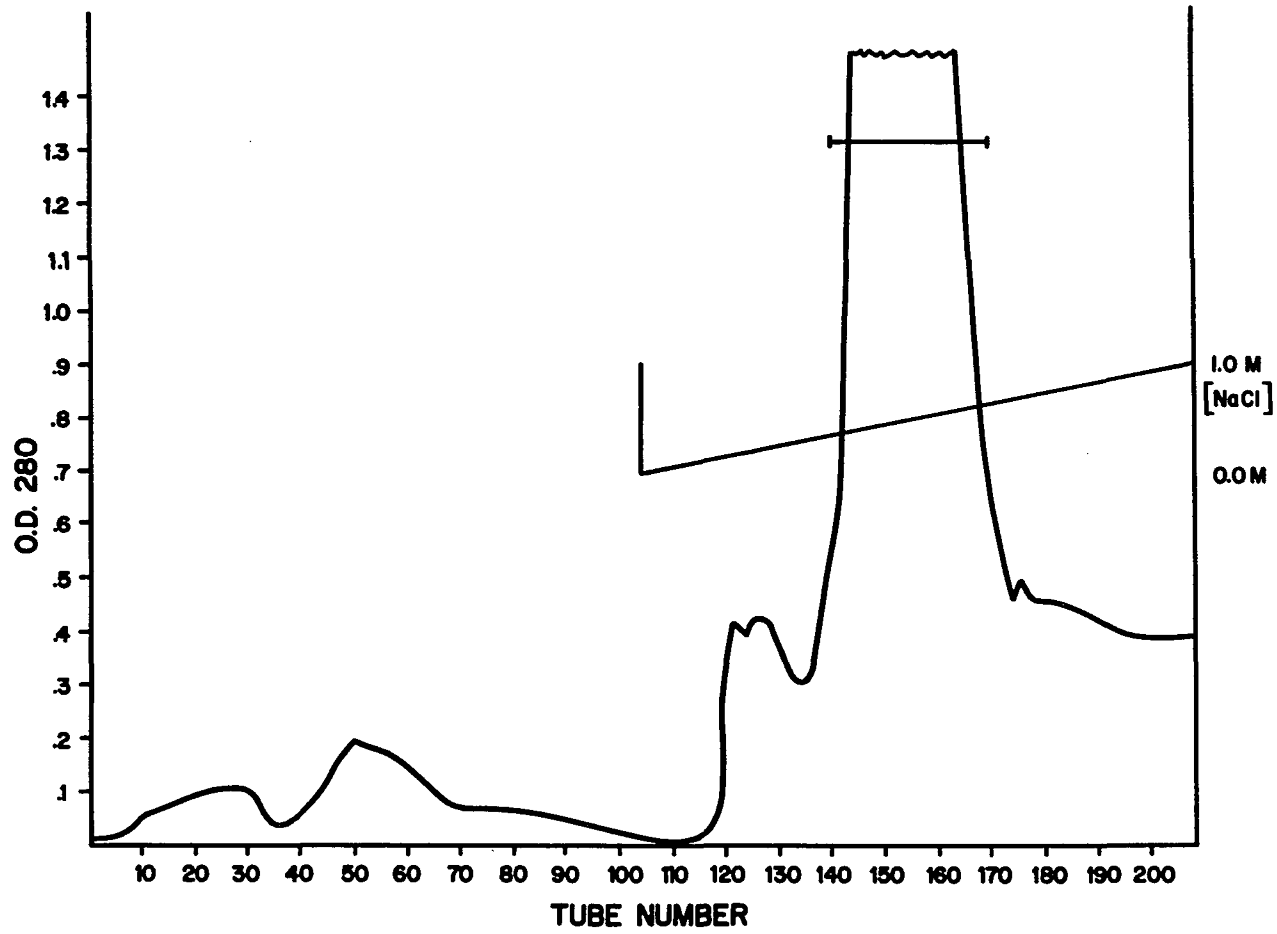
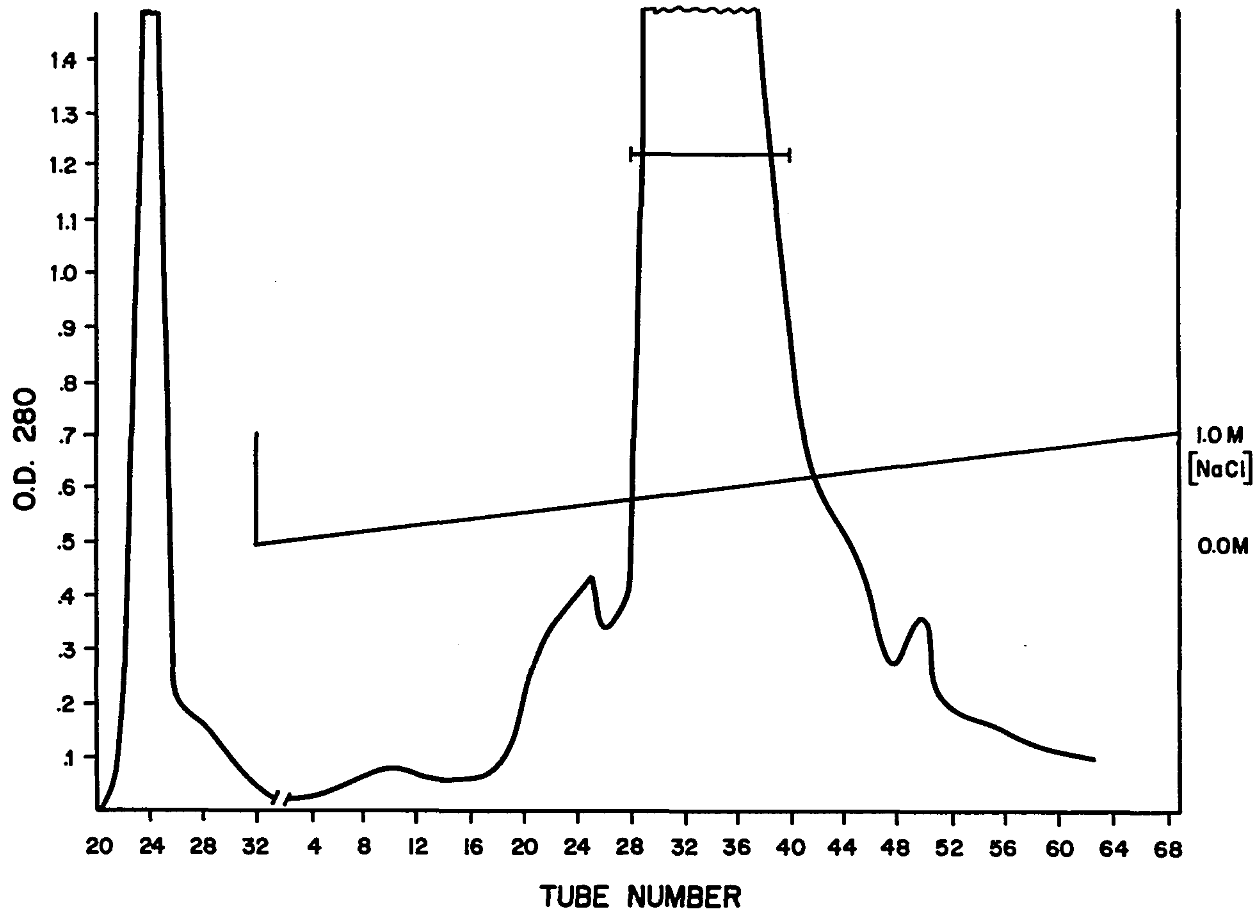

O.D. 280
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
TUBE NUMBER
1.0 M
[NaCl]
0.0 M

FIGURE II-1

Chromatography of IgAκ Gom on DEAE Sephadex. Starting buffer: 0.05M sodium phosphate, pH 7.85. Gradient: 0.0M to 1.0M NaCl in starting buffer, 500 ml in each reservoir. Column size: 2.5 x 100 cm.



O.D. 280
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
1.0 M
[NaCl]
0.0M
20 24 28 32 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68
TUBE NUMBER

FIGURE II-2

Chromatography of IgMλ Moo on DEAE Sephadex. Starting buffer: 0.05M sodium phosphate, pH 7.85. Gradient: 0.0M to 1.0M NaCl in starting buffer, 500 ml in each reservoir.

After applying peptide mixtures dissolved in starting buffer and washing with more than two column volumes of starting buffer, a gradient of 0.0M to 1.0M NaCl in the starting buffer was used to develop the column.

4. Molecular exclusion chromatography

a. Molecular exclusion chromatography in non-dissociating buffers

Sepharose 4B equilibrated in phosphate buffered saline and packed in a 2.5 x 100 cm column and Sephadex G-200 equilibrated in 0.5M Tris (hydroxymethyl) aminomethane (Sigma), pH 8.2, packed in a 2.5 x 100 cm column were used for purification of intact immunoglobulins.

b. Molecular exclusion chromatography in dissociating, volatile buffers

Sephadex G-100 developed with 1.0M propionic acid packed in a 2.5 x 100 cm column was used for the separation of partially reduced and alkylated heavy and light chains. Sephadex G50 fine equilibrated in 1.0M propionic acid or 0.5M ammonia in a 2.5 x 240 cm column was used to separate certain peptides. Sephadex G-25 fine equilibrated with 0.5M ammonia and packed in a 1.5 x 200 cm column was used to separate certain peptides. Desalting was performed on columns of G-25 coarse, 0.5M propionic acid (2.5 x 100 cm), G-25 coarse, 0.5M ammonia (1.5 x 90 cm) or G-25 fine, 0.5M ammonia (2.5 x 25 cm).

c. Molecular exclusion chromatography in dissociating, non-volatile buffers

Sephadex G-100 equilibrated with 5M guanidine HCl in columns 1.5 x 90 cm, 1.5 x 100 cm, 2.5 x 120 cm, 2.5 x 140 cm and 5.0 x 150 cm was used to separate heavy and light chains and to separate cyanogen bromide peptides. Guanidine HCl was "Absolute" grade (Research Plus,

Inc.) or reagent grade further purified by passage through a column of decolorizing charcoal. Sephadex G-75 and Sephadex G-50 fine equilibrated in 5M guanidine HCl in 2.5 x 120 cm columns were used for further separation of peptides.

5. Separation of peptides by high voltage paper electrophoresis

Certain peptide mixtures were separated by high voltage electrophoresis on paper. Peptide mixtures, dissolved in 0.5 M ammonia, were applied to Whatman No. 3 paper and electrophoresed for 20 to 75 minutes at a pH of 2.1 or 3.5 (Crumpton and Wilkinson, 1965). A strip of the electrophoretogram was cut and stained with cadmium ninhydrin to locate the peptides. Peptide containing strips were then cut from the electrophoretogram and the peptides eluted with 0.5 M ammonia.

6. Separation of peptides by differential solubility

The method of Rudikoff (Rudikoff and Potter, 1974) was employed to separate the CB-2 peptide of protein Gom. A mixture of 3 cyanogen bromide peptides was slurried in 0.01 M ammonium acetate buffer, pH 5.0, and allowed to incubate overnight at 4°C. The slurry was then centrifuged to pellet the precipitate. The precipitate was washed three times, dissolved in 50% acetic acid, and lyophylized.

C. Evaluation of protein purity

1. Cellulose acetate electrophoresis

Electrophoresis on cellulose acetate membranes was performed using a Beckman Microzone Electrophoresis Apparatus and barbital buffer, pH 8.6, ionic strength 0.1. Proteins in solution were applied to the membrane, which was subjected to a constant current of 6 ma for 20 minutes. After electrophoresis, the membranes were stained with Ponceau S in

trichloroacetic acid and destained in 5% acetic acid.

2. Polyacrylamide gel electrophoresis

Electrophoresis, in a 7.5% polyacrylamide gel matrix, contained in a 6 x 150 mm glass tube and formed in a 0.1M sodium phosphate running buffer, pH 7.1, with 0.5M urea and 0.1% sodium dodecyl sulfate, was performed using a constant current of 15 ma/gel for 2 to 4 hours. The gels were stained with Coumassie Brilliant Blue and destained in 7.5% acetic acid, 10% ethanol.

3. Ouchterlony double immunodiffusion analysis

Double immunodiffusion analysis was performed in a matrix of 1% ionagar #25 (Wilson Diagnostics, Inc.) and veronal buffer, pH 8.4. Antisera of multiple specificity as well as antisera rendered monospecific by suitable absorption were used.

4. Immunoelectrophoresis

Microimmunoelectrophoresis in 1% ionagar matrix with a veronal buffer, pH 8.4, was carried out by applying a voltage differential of 5V/cm through the gel matrix for 1 hour and then incubating with the appropriate antiserum overnight (Scheidegger, 1955).

D. Preparation and quantitation of heavy chains

1. Partial reduction and alkylation

Purified, pooled, IgG was partially reduced and alkylated according to the method of Fleischman _et al_. (1962). The IgG was dissolved in 1.0M tris (hydroxymethyl) aminomethane, (tris), pH 8.2, at a concentration of 10 to 30 mg/ml and reduced with 0.1M 2-mercaptoethanol under a nitrogen atmosphere for one hour at room temperature. Following a 15 minute preincubation in ice, the reduced chains were alkylated with 0.15M

iodoacetamide for 20 minutes. After overnight dialysis against 1.0M propionic acid, the heavy and light chains were separated on G-100 Sephadex equilibrated with 1.0M propionic acid. In certain cases the heavy chains were further purified on G-100 Sephadex in 6.0M urea--1.0M propionic acid or 5.0M guanidine HCl. After desalting, when appropriate, the heavy chains were lyophilized and quantitated by the Folin method. Several heavy chains were subjected to Kjeldahl nitrogen analysis and their dry weight and ash contents determined.

2. Complete reduction and alkylation

Canine IgAκ Gom was completely reduced in 1.0M tris, pH 8.2, 6.0M guanidine HCl, 0.005M EDTA using 0.1M 2-mercaptoethanol or 2.6 mM dithiothreitol for one hour. The protein concentration was 10-13 mg/ml. The reduced chains were alkylated with ^{3}H or ^{14}C labeled iodoacetamide or iodoacetic acid (0.3 μCi/mg of protein) for 5 minutes followed by unlabeled iodoacetamide or iodoacetic acid at 0.15M for 20 minutes at room temperature. The reduced and radiolabeled, carboxymethylated chains were separated on G-100 Sephadex equilibrated in 5.0M guanidine HCl.

Canine IgMλ Moo was completely reduced in 1.0M Tris, 6.0M guanidine HCl, 0.005M EDTA at a protein concentration of 20 mg/ml using 0.1M 2-mercaptoethanol. The reduced chains were carboxymethylated with ^{14}C iodoacetamide or iodoacetic acid for 5 minutes followed by unlabeled iodoacetamide or iodoacetic acid at 0.15M for 20 minutes at room temperature. The reduced and ^{14}C-carboxymethylated chains were separated on G-100 Sephadex equilibrated in 5.0M guanidine HCl.

E. Fragmentation of myeloma proteins

1. Cyanogen bromide cleavage of heavy chains

Completely reduced and carboxymethylated alpha chain was reacted in 70% formic acid (10-20 mg/ml) with cyanogen bromide (CNBr: protein 1.5-5.0:1) for 16 to 24 hours at room temperature. The reaction mixture was diluted with 10 to 20 volumes of distilled water and lyophilized. Fragments were initially separated on G-100 Sephadex equilibrated with 5.0M guanidine HCl.

2. Cyanogen bromide cleavage of whole proteins

Gom IgAκ dimer and Moo IgMλ pentamer were reacted in 70% formic acid (10-20 mg/ml) with cyanogen bromide (CNBr: protein 1.5-5.0:1) for 16 to 24 hours at room temperature. The reaction mixture was then diluted with 10 to 20 volumes of distilled water and lyophilized. Fragments were initially separated on G-100 Sephadex equilibrated with 5.0M guanidine.

After separation of the CNBr fragments of Moo IgMλ, the fragments were reduced and ^{14}C-carboxymethylated (vide II-D-2) and further purified on Sephadex G-75, or G-50 as appropriate.

3. Tryptic digestion of cyanogen bromide fragments

Cyanogen bromide fragments of Gom alpha chain and Moo mu chain were further fragmented by digestion with trypsin (enzyme:substrate 1:100) in 1% ammonium bicarbonate for 16 hours at 37°C. The reaction was stopped by lyophilization or application of the digest to a column of G-50 fine or G-25 fine Sephadex equilibrated in 1.0M propionic acid or 0.5M ammonia. Some peptides were pretreated with citraconic anhydride in 1.0M Tris - 6.0M guanidine HCl, pH 9.0.

F. Automated amino acid sequencing

1. Liquid phase automated sequencing

Fifty to four hundred nanomoles of peptide or protein in 50% acetic acid were applied to the cup of an updated Beckman 890A or a Beckman 890C automated sequencer. Some of the peptides, containing lysine, were treated with 4-sulfophenyl isothiocyanate (Pierce Chemical) prior to sequencing. Coupling of phenylisothiocyanate to amino terminal alpha amino groups was accomplished in dimethylallylamine buffer. Cleavage was performed with heptaflurobutyric acid and the phenylthiozolinone derivatives were extracted with butyl chloride. After evaporation to dryness under nitrogen stream, the phenylthiozolinone amino acids were converted to the phenylthiohydantoin derivatives by reaction with 1.0M HCl at 80°C under a nitrogen atmosphere for 10 minutes. Phenylthiohydantoin derivatives of all amino acids other than histidine and arginine were extracted from the HCl with ethyl acetate.

2. Identification of phenylthiohydantoin (PTH) derivatives of amino acids

a. Gas-liquid chromatography

PTH-amino acids were initially examined on a Beckman GC-45 or GC-65 gas liquid chromatograph equipped with glass columns packed with chromosorb W coated with SP-400 (Beckman Instruments). PTH derivatives of alanine, serine, carboxymethylcysteine, glycine, valine, proline, threonine, leucine, isoleucine, methionine, phenylalanine, tyrosine and tryptophan were detected in this manner (Pisano and Bronzert, 1969). Reaction of PTH amino acids with N,O-bis(trimethylsilyl)acetamide (Analabs, Inc.) allowed the differentiation of PTH-leucine from PTH-isoleucine and the detection of PTH-aspartic acid and PTH-glutamic acid.

b. Thin layer chromatography

Ascending thin layer chromatography in two dimensions was performed on 5 cm x 5 cm plates coated on both sides with a micropolyamide matrix (Gallard-Schlesinger Chemical Mfg. Co.). The first dimension solvent was methanol: toluene: acetic acid 30:15:11 with the fluor BBOT 2, 5-bis-2-(5-tert-Butylbenzoxazolyl)-Thiophene, Packard Instrument Company, Inc. The second dimension solvent was 25% acetic acid. This thin layer system allows identification of the PTH derivatives of proline, leucine/isoleucine, valine, phenylalanine, methionine, threonine, alanine, glycine, aspartic acid, asparagine, glutamic acid, glutamine, serine, carboxymethylcysteine, tyrosine, tryptophan and lysine (Summers *et al.*, 1973).

c. PTH hydrolysis

PTH amino acids were treated with HI for 20 hours at 150°C under a nitrogen atmosphere to liberate the free amino acid, which was subsequently identified on a Durrum D-500 Amino Acid Analyzer.

d. Scintillation counting

The presence of carboxymethylcysteine was confirmed by examining an aliquot of the appropriate position and its neighbors for the ^{3}H or ^{14}C radiolabel applied during reduction and carboxymethylation. Samples were dissolved in a dioxane based scintillation fluid containing 0.1g of 1,4-bis-2-(5-Phenyloxazolyl)-Benzene, 5.0g of 2,5-Diphenyloxazole (Packard Instrument Company, Inc.) and 120.0g of naphthalene per liter and counted in a Beckman LS-350 Liquid Scintillation System.

G. Amino acid analysis

Proteins and peptides were hydrolyzed in 6M HCl in vacuo for 20

hours at 110°C under a nitrogen atmosphere. The hydrolysates were analyzed on the Durrum D-500 Amino Acid Analyzer.

III. RESULTS

A. Sequence studies of normal pooled sera

Each heavy chain pool in which sequence data could be obtained was sequenced to at least position 11. These amino-terminal sequences confirmed the correspondence of all heavy chain pools with the prototype human V_HIII sequence. The percentage of the V_HIII subgroup for each mammalian species was quantitated by comparing the number of nanomoles of valine present in position 2 (except porcine chains, which had glutamine) of the pool with the number of nanomoles of heavy chain placed in the sequencer. Since very few immunoglobulin heavy chains of the V_HIII subgroup have been found to have a pyrrolidone-1-carboxylic acid as the NH_2 terminal residue, this determination of the percentage of unblocked V_HIII heavy chains will be an accurate minimal estimate of the percentage of this subgroup present in each pool.

Three homogeneous paraproteins known to belong to the V_HIII subgroup served as totally unblocked positive controls. Two blocked V_HII proteins (one human and one murine) and one blocked human V_HI protein served as negative controls. By repeating the entire procedure from the IgG isolation to the V_HIII quantitation for the dog, cat, mouse, and human pools, the total error of the procedure was found to be 5%.

The percentage of unblocked heavy chains for each species examined is listed in Table III-1. The first three proteins are the homogeneous

Table III-1

Calculation of the Unblocked V_HIII Subgroup*

	A Load	B Valine Step 2	C∞ Yield at Step 2		D§ Unblocked
	nmol	nmol	%		%
MOPC 173	152	128	84		
MOPC 21A	131	121	92	87	100
Tei IgG1	153	132	86		
Lev IgG3	148	7	5		
Taf IgG2	149	9	6	5	0
MOPC 40	166	7	4		
Opossum	155	131≡	85		97
Dog	152	122	80		91
Mink	80	68	85		97
Cat	161	130	81		92
Sea lion	95	80	84		96
Seal	85	74	87		100
Guinea pig	157	38	24		23
Mouse	154	36	23		22
Rat	102	22	22		20
Monkey (African Green)	103	30	29		29
Monkey (cynomolgus)	161	42	26		26
Monkey (Rhesus)	144	30	21		19
Man	156	34	22		20
Rabbit	107	8	7		2
Pig∿					40
Cow	137	7	5		0
Sheep	137	7	5		0
Goat	176	9	5		0
Moose	160	7	4		0
Horse	129	7	5		0
White whale	164	8	4		0
Fin whale	141	6	4		0

*Data used in determining percentage of unblocked residues. Column A indicates the molar quantity sequenced, column B the yield of valine measured at step 2 of the degradation, and column C the percentage yield of valine at step 2. The percentage of unblocked chains is given in column D, determined as described in the text.

∞Column B divided by column A x 100.

§Considering 87% as 100% unblocked and 5% as 100% blocked, thus, for the cat $(81-5)\left(\frac{100}{87-5}\right) = 92$

≡Measured as isoleucine in position 2.

∿Measured by dinitrophenylation, F. Franek.

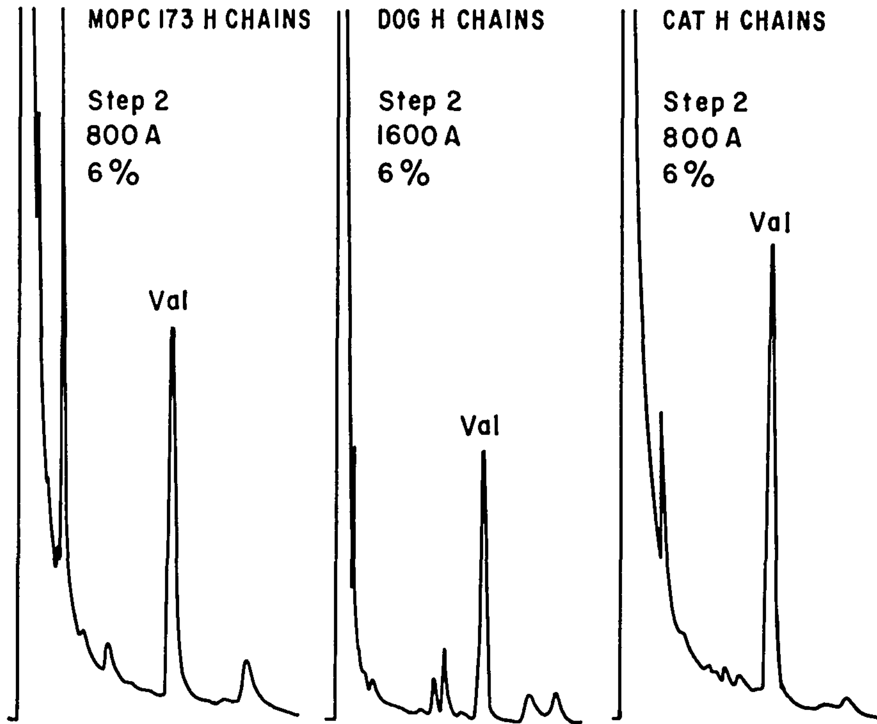
mouse (MOPC 173, MOPC 21A) and human (Tei, IgG1) unblocked V_HIII controls. Column C shows that these heavy chains, which were completely unblocked, gave PTH recoveries at step 2 varying from 84 to 92% (ave. 87%). This value corresponds closely to the yields that have been obtained previously with automated sequencers. Consequently, in all subsequent calculations, an 87% yield at step 2 was taken as an indication of a totally unblocked preparation of heavy chains. This parameter should be distinguished from the step-to-step repetitive yield, which is 95-96%.

The next three heavy chains (MOPC 40, Lev IgG3, and Taf IgG2) are homogeneous, blocked myeloma proteins of either V_HI or V_HII subgroup. As shown in column C of Table III-1, a small amount of valine (ave. 5%) was found in these preparations. This "background" level is the result of several factors inherent in automated sequencing (Smithies _et_ _al_., 1971). Thus a yield of valine of up to 5% at step 2 was assumed to indicate total blockage of a given preparation. All the test heavy chains were analyzed on this basis. The percentage of unblocked V_HIII heavy chains in each test pool is shown in column D of Table III-1.

Fig. III-1 shows actual gas chromatographic analyses of the PTH derivative at position 2 for some of these heavy chain preparations. The cat and dog pools show no significant differences from the unblocked mouse myeloma heavy chain. The figure also compares the horse and sheep heavy chain preparations with a V_HII protein that is known to be blocked.

As illustrated in Table III-1, the heavy chains of some species are completely blocked, those of others completely unblocked, and those of still others contain a mixture of both blocked and unblocked chains.

(a)



(b)

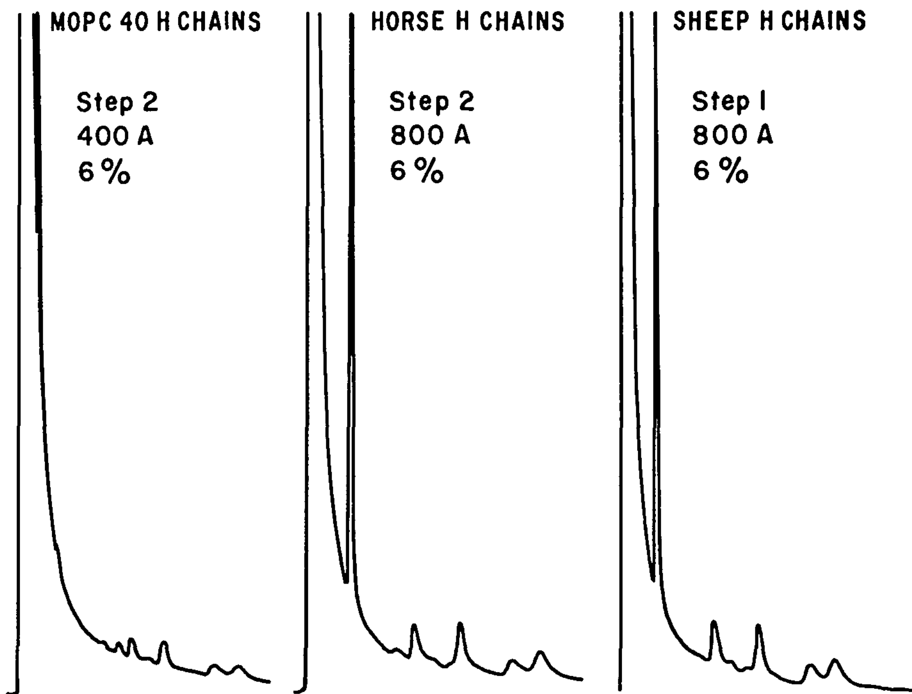


FIGURE III-1

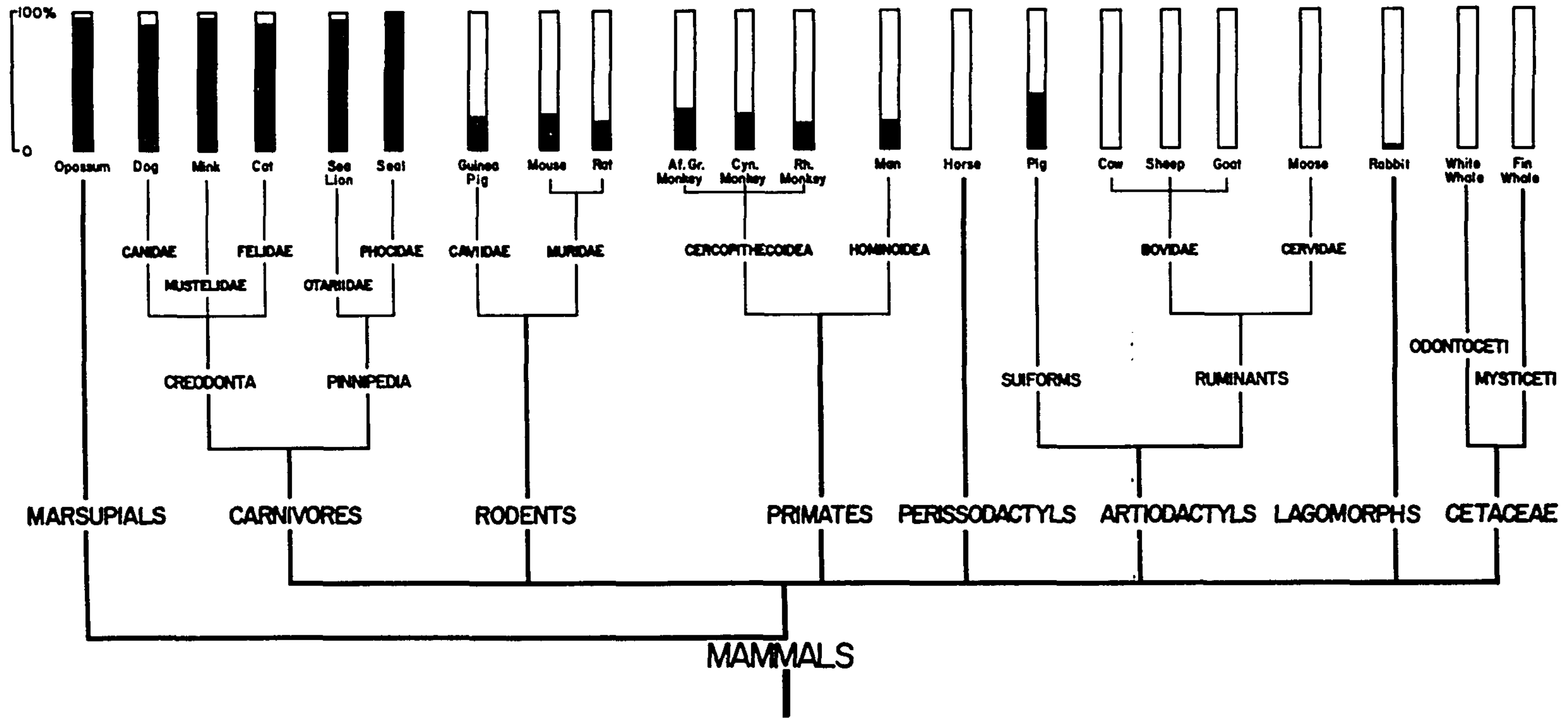
Gas chromatographic analysis of step 2 of the Edman degradation of pooled immunoglobulin heavy chains from (a) an unblocked myeloma protein and pools from the dog and cat and (b) a blocked myeloma protein and pools from the horse and sheep.

With rare exception, all the unblocked heavy chains were clearly assignable to the V_HIII subgroup.

As noted previously, this method of determining the percentage of unblocked chains leads to values that are minimal estimates of the actual amount of this subgroup present in a given preparation, since no V_HIII heavy chains with blocked amino termini (Ponstingl et al., 1970; Ray and Cebra, 1972) would be detected by the method. A composite of the results of the V_HIII subgroup distribution analyses for the various species is presented in graphic form in Fig. III-2.

The heavy chains from those animals in which the V_HIII subgroup was found to comprise a significant proportion of the immunoglobulin heavy chain V region pool were subjected to further analysis. These studies were therefore limited to three groups of animals: (a) marsupials and carnivores, which were found to have exclusively unblocked, V_HIII heavy chains; (b) rodents, primates, and suidae, which were found to have approximately 25% of their heavy chains as unblocked members of this subgroup; and (c) avians in which the V_HIII subgroup was not quantitated but accounted for more than one-half of each heavy chain pool. These data were correlated with previously available information on the primary structure of myeloma proteins from the dog and cat (Kehoe and Capra, 1972), mouse (Bourgois and Fougereau, 1970), and man (Capra, 1971).

The yield at each position was quantitated using the PTH derivatives by a combination of gas chromatography and acid hydrolysis followed by amino acid analysis. Background values were estimated by sequencing V_HIII proteins of known homogeneity and sequence, and by carrying out comparable analyses of horse heavy chains, which were shown to be totally blocked.



100%
0
Opossum
Dog
Mink
Cat
Sea Lion
Seal
Guinea Pig
Mouse
Rat
Af. Gr. Monkey
Cyn. Monkey
Rh. Monkey
Man
Horse
Pig
Cow
Sheep
Goat
Moose
Rabbit
White Whale
Fin Whale
CANIDAE
FELIDAE
MUSTELIDAE
OTARIIDAE
PHOCIDAE
CAVIIDAE
MURIDAE
CERCOPITHECOIDEA
HOMINOIDEA
BOVIDAE
CERVIDAE
ODONTOCETI
MYSTICETI
CREODONTA
PINNIPEDIA
SUIFORMS
RUMINANTS
MARSUPIALS
CARNIVORES
RODENTS
PRIMATES
PERISSODACTYLS
ARTIODACTYLS
LAGOMORPHS
CETACEAE
MAMMALS

FIGURE III-2

Phylogenetic distribution of the V_HIII subgroup of immunoglobulin heavy chains among mammals. The percent of the unblocked V_HIII subgroup is depicted in black.

This made possible the calculation of parameters such as the repetitive yield, nonspecific internal peptide bond cleavage, incomplete coupling, and incomplete thiozolinone cleavage (Smithies et al., 1971). In another control, a human and a canine myeloma protein were mixed in varying proportions and sequenced for 30 Edman degradation steps to highlight residues distinct for these two species. These various procedures showed that a 5% contamination by an alternative residue could be detected for virtually every position studied by the methods described.

The extraordinary structural homogeneity of the V_HIII subgroup of mammalian heavy chains allowed a significant amount of amino acid sequence data to be obtained on pooled IgG from these various species. Fig. III-3 compares the gas chromatographic analysis of position 23 of representative canine and murine myeloma proteins with pools from these same species. Previous studies (Kehoe and Capra, 1972) showed that all of a series of canine and feline myeloma proteins had valine in position 23 while human and murine myelomas had alanine at that position. A comparison of the data from the pooled heavy chains from these species shows that the canine pool contains exclusively valine at this position while the murine pool contains exclusively alanine (Fig. III-3). Quantitation of these residues by both gas chromatography and amino acid analysis indicated that, in each instance, greater than 95% yield of either valine or alanine had been obtained in position 23 in both the myeloma proteins and the pools. Thus, this position of the pools, as virtually every heavy chain pool position to residue 30 that has been analyzed to date, cannot be distinguished from myeloma heavy chains by the automated sequencer, with the few exceptions that are noted below. After position 30 this is no longer true, because

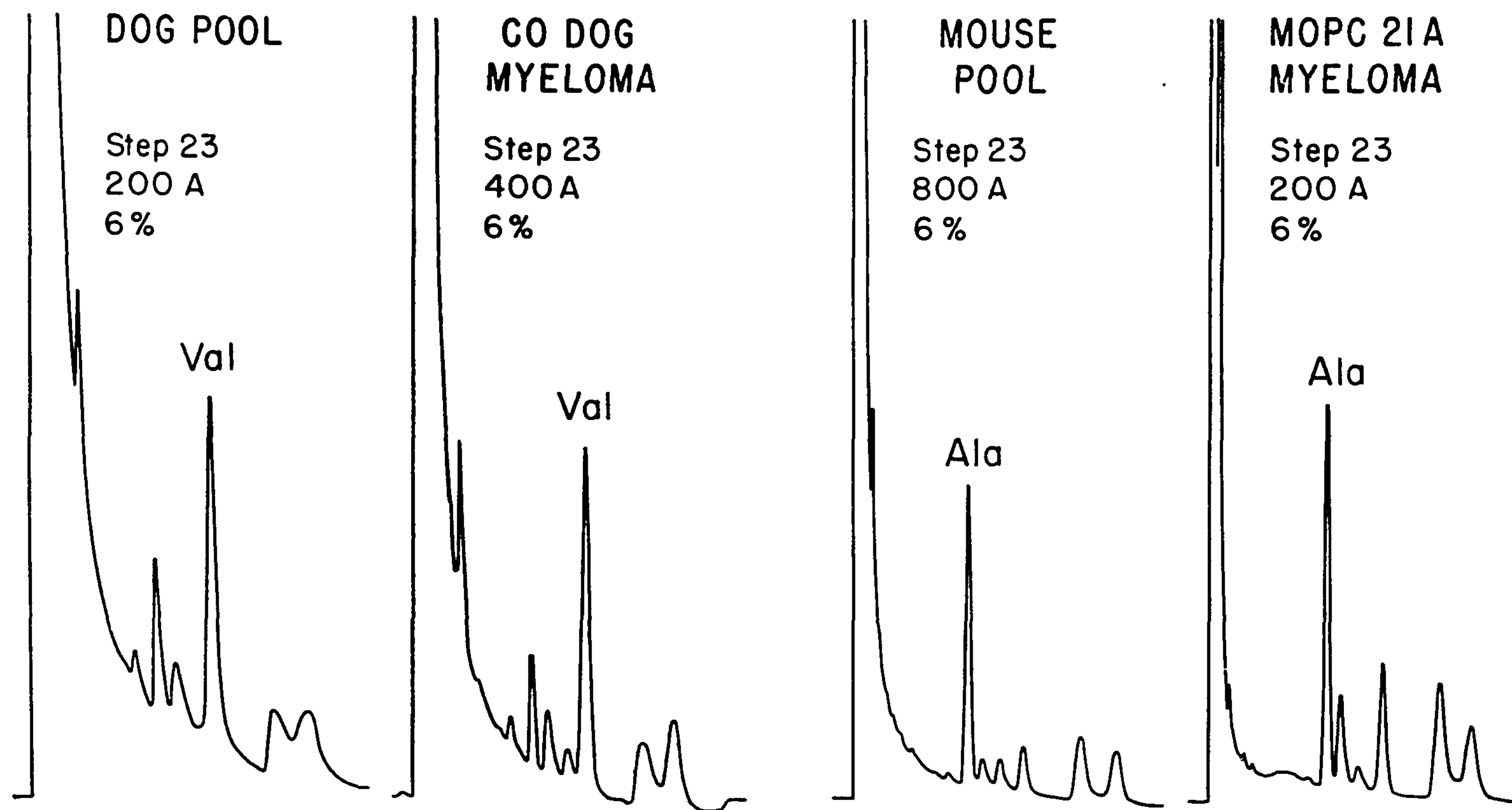

DOG POOL
Step 23
200 A
6%
Val
CO DOG
MYELOMA
Step 23
400 A
6%
Val
MOUSE
POOL
Step 23
800 A
6%
Ala
MOPC 21A
MYELOMA
Step 23
200 A
6%
Ala

FIGURE III-3

Gas chromatographic analysis of the PTH derivative at position 23 of two myeloma heavy chains and of heavy chain pools isolated from the same species. The different attenuation recorded reflects both the different number of nanomoles intially placed in the sequencer and the percentage of each population of molecules that is unblocked. In both the dog pool and the dog myeloma, the amount of valine relative to background is approximately the same. Similarly, the amounts of alanine in the mouse pool and the myeloma heavy chain are virtually identical. The uniqueness of both alanine and valine in the respective pools is clearly evident.

Figure III-4

PROTOTYPE	Glu	Val	Gln	Leu	Val	Glu	Ser	Gly	Gly	Gly	Leu	Val	Gln	Pro	Gly	Gly	Ser	Leu	Arg	Leu	Ser	Cys	Ala	Ala
Human	1				2					3														
Mouse			Lys																Lys					
Rat	4																							
Guinea Pig																							Val	
Dog										Asp			5										Val	
Cat	Asp									Asp											Thr		Val	
Mink										Asp														
Seal			Lys							Asp														
Sea Lion										Asp														
Opossum		Ile								Asp														
Pig		Glu																	( )				Val	
Pigeon	Ala	Ile	6														( )				Val		Gly	
Goose	Ala	Ile			Asp	7				Val			Gly			Val	Gln				Val		Gly	
Duck	Ala	Ala	Thr		Asp								Gly			Val	Gln				Val		Gly	
Turkey	Ala	8		9						10	Val	Gln	Gly			Val					Val		Gly	
Chicken	Ala		Thr		Asp							Gln	Thr					Ala			Val		Gly	

FIGURE III-4

Amino acid sequence analysis of the V_HIII subgroup of pooled mammalian and avian immunoglobulin heavy chains. A prototype sequence based on human proteins is given at the top. For the other proteins only those residues that differ from the prototype have been shown. Except where indicated by a number, all residues reported represent greater than 95% of the amino acids recovered at that position. Exceptions: (1) 10% aspartic acid; (2) 15% leucine; (3) 10% alanine; (4) 15% aspartic acid; (5) 10% alanine; (6) 50% glutamic acid; (7) 25% alanine; (8) 25% serine; (9) 25% glutamine; (10) 50% alanine.

of the influence of the first heavy chain hypervariable region (residues 31-35) (Capra, 1971). Since no given amino acid is present more than 25% of the time in these positions of pooled heavy chains, no predominant sequence is apparent. Myeloma proteins, in contrast, remain homogeneous through this region.

The predominant amino acid sequence of pooled immunoglobulin heavy chains from 11 mammalian and 5 avian species in which the V_HIII subgroup is represented, is shown in Fig. III-4. The degree of structural preservation within the V_HIII subgroup in these pooled heavy chains is clearly evident. A significant secondary sequence was detectable in only ten positions, as noted in the legend.

Among the mammals, positions 10, 21, and 23 were subjected to more careful study because of certain unique features that they showed. All the carnivores and the opossum yielded no detectable glycine at position 10, and the primates and the rodents gave no detectable aspartic acid at this position. In man, approximately 10% of the chains in the pool contained alanine in this position, consistent with the observation that about 15% of human myeloma proteins of the V_HIII subgroup have alanine at position 10. The most striking finding, however, was the total absence of aspartic acid in position 10 of the primates and rodents, and the absence of either glycine or alanine at this position in any of the carnivores or the marsupial. At position 21, the dog pool gave exclusively serine and the cat pool exclusively threonine. Finally, at position 23, those species that had alanine had no valine and those with valine showed no alanine. Thus, whenever a given species has a different amino acid at a particular position than do certain other species, the difference is generally apparent

in all of the heavy chains of that particular species. Significantly, the differences seen in the pool correspond to those seen in myeloma proteins (e.g., the threonine at position 21 of the cat, Kehoe and Capra, 1972).

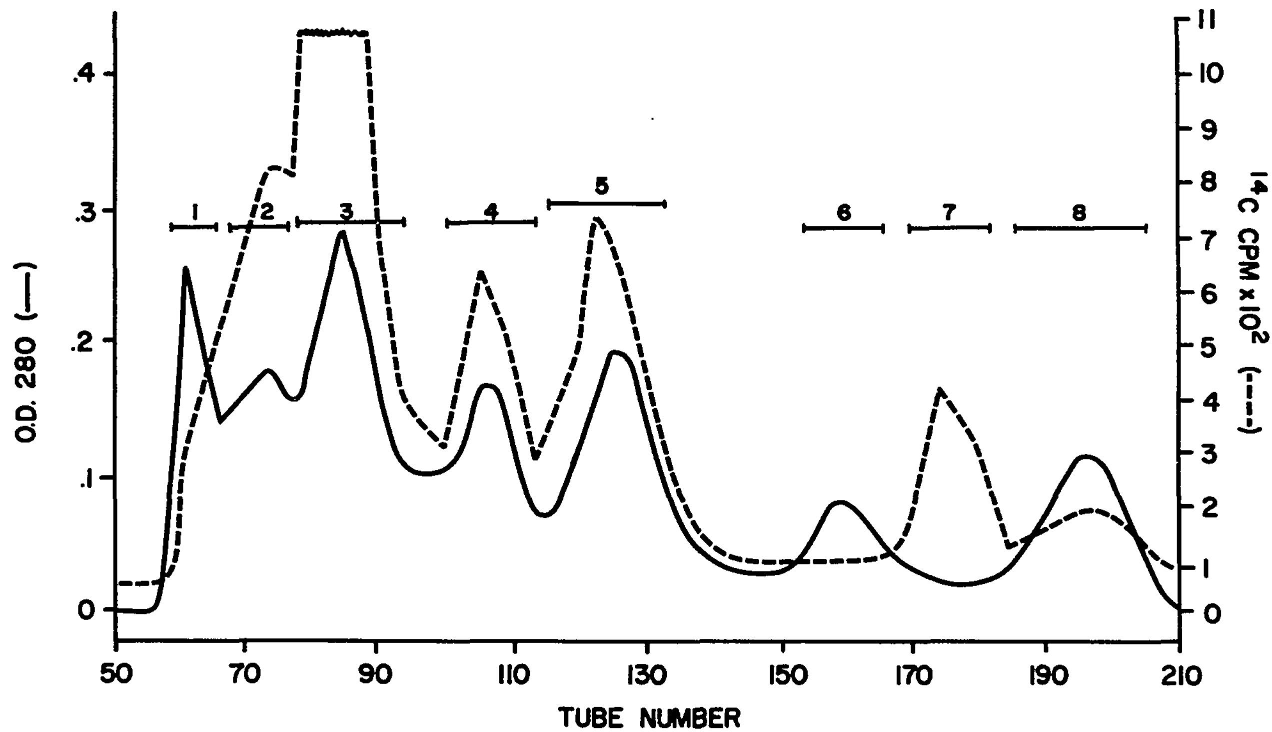
Despite the limited range of avian species examined, residues characteristic of this class (when compared to the mammals) could be identified. All members of the class Aves included in this study have alanine in position 1, valine in position 21, and glycine in position 23. These are distinctly different residues from any seen at these positions in the mammals studied and thus appear to be "avian-associated residues."

B. The sequence of the variable region of a canine alpha chain

Canine IgAκ myeloma Gom was purified as described in Materials and Methods. The purified protein migrated as a single band on cellulose acetate electrophoresis and formed a single arc in an immunoelectrophoretogram developed with rabbit anti-whole canine serum. Heavy and light chains were separated on Sephadex G-100 equilibrated with 5M guanidine HCl after complete reduction and carboxymethylation. The heavy chain preparation was evaluated for light chain contamination by comparing the yields, after sequential Edman degradation, of valine and isoleucine at position 2.

1. Digestion of canine alpha chain Gom with cyanogen bromide

Completely reduced and carboxymethylated heavy chains were dissolved in 70% formic acid and digested with cyanogen bromide overnight at room temperature (protein: CNBr ratio 1:5). The resulting peptides were separated by molecular exclusion chromatography on a 5 x 150 cm column



O.D. 280 (—)
.4
.3
.2
.1
0
^{14}C CPM x 10^2 (----)
11
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
50
70
90
110
130
150
170
190
210
TUBE NUMBER

FIGURE III-5

Chromatography of cyanogen bromide peptides of alpha chain Gom. The digest was applied to a 5.0 x 150 cm column packed with Sephadex G-100 equilibrated with 5.0M guanidine HCl.

of Sephadex G-100 equilibrated with 5M guanidine HCl. The elution profile is shown in Figure III-5. Each peak was pooled as indicated and tested in the automated sequencer. Peaks 5, 6, and 7 were found to contain variable region peptides.

Peaks 5 (CB-3) and 7 (CB-1) required no further purification. Peak 6 contained three peptides, one of which was identified as a peptide beginning at position 35; the other two peptides were not variable region peptides. The variable region peptide (CB-2) was isolated by precipitation in 0.01M ammonium acetate, pH 5.0, as described by Rudikoff (Rudikoff and Potter, 1974).

2. Digestion of IgAκ dimer with cyanogen bromide

Since peak 6 of the heavy chain cyanogen bromide digest was not radioactive, the peptides eluting at that position contained no carboxymethylcysteine. Therefore, intact IgAκ dimer was digested with cyanogen bromide in formic acid as described above (III-B-1) and the peptides separated by molecular exclusion chromatography on Sephadex G-100 in 5M guanidine HCl as shown in Figure III-6. Peak 4 of the digestion of the dimer was found to contain the same three peptides as peak 6 of the heavy chain digest and the variable region peptide, CB-2, was isolated as previously indicated (III-B-1).

3. The sequence of the amino terminal cyanogen bromide fragment of Gom alpha chain

Three hundred nanomoles of heavy chain were subjected to forty-one cycles of sequential Edman degradation. Unequivocal identification of all positions except 30, 31 and 38 was established by gas chromatography, with and without silylation, and by thin layer chromatography.

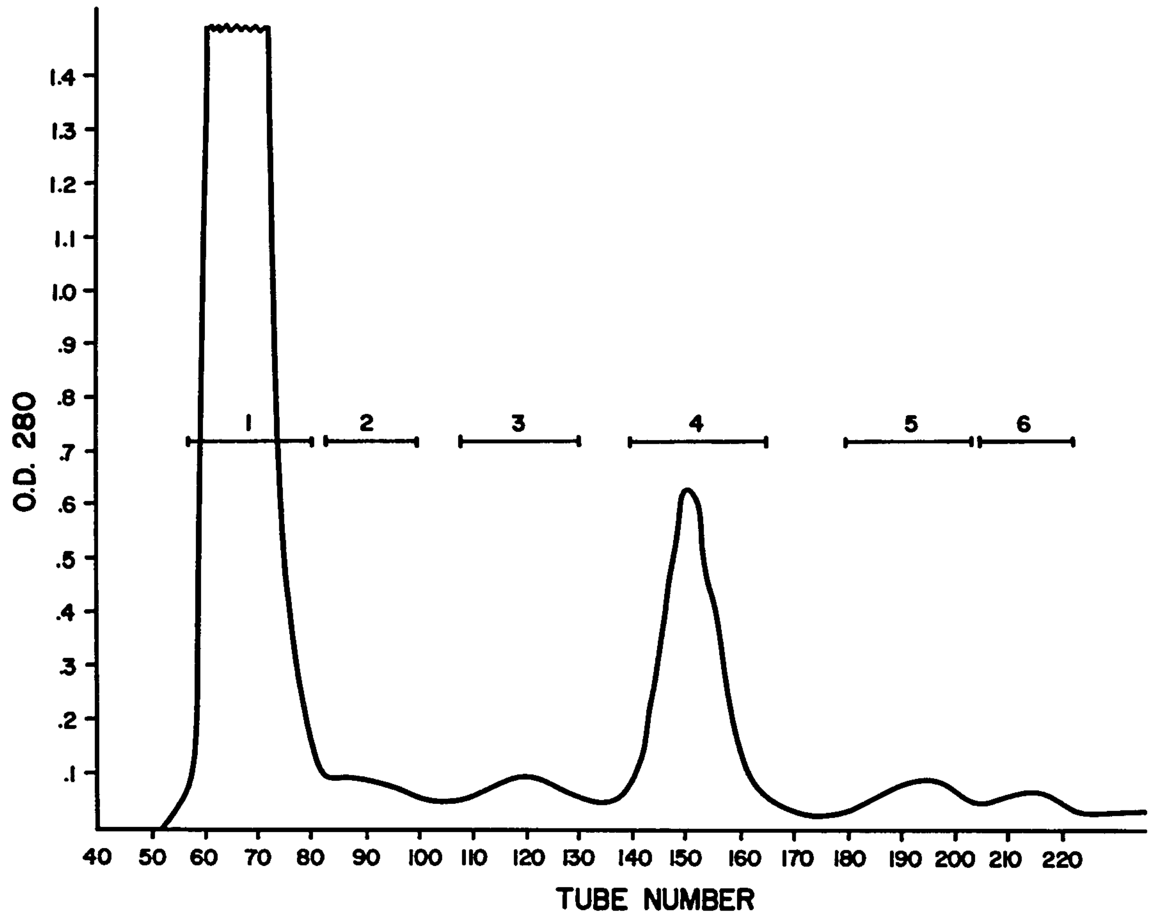

O.D. 280
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
1
2
3
4
5
6
40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220
TUBE NUMBER

FIGURE III-6

Chromatography of cyanogen bromide peptides of IgAκ Gom.

The digest was applied to a 5.0 x 150 cm column packed with Sephadex G-100 equilibrated with 5.0M guanidine HCl.

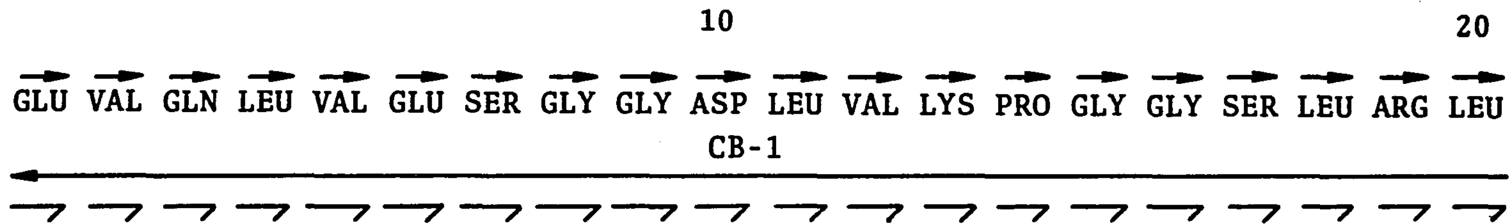
This experiment resulted in the identification of most of the amino-terminal CB-1 peptide. The two unidentified residues were identified by sequencing the isolated CB-1 peptide. Thirty-four cycles of Edman degradation were performed and all positions up to and including position 33 were identified. These two extended sequencer runs completed the sequence of the amino-terminal peptide and extended into the next peptide thus providing an overlap. The sequence and its derivation is shown in Figure III-7. The composition of this peptide as determined by the sequence and by amino acid analysis after acid hydrolysis is shown in Table III-2.

4. The sequence of the second cyanogen bromide fragment of the variable region of alpha chain Gom

CB-2, isolated as described in III-B-1 and III-B-2, was sequenced for 41 Edman degradation steps on the automated sequencer and identifications made of all positions up to and including step 30. In addition, positions 34, 36 and 41 were also identified.

One micromole of CB-2 was digested with trypsin (1% w:w) in 1% ammonium bicarbonate for 16 hours at 37°C. The tryptic peptides were initially separated by high voltage electrophoresis on Whatman No. 3 paper at pH 2.1 using 3000 volts for 30 minutes and eluted with 0.5M ammonia. Peptides which migrated together at pH 2.1 were separated by a second electrophoresis at pH 3.5 using 3000 volts for 20 minutes.

Except for determination of their amino termini, peptides T-1, T-2, and T-3 were not sequenced since their sequence had already been established by the amino-terminal run on the intact CB-2 fragment. In order to confirm certain positions, peptide T-4 was sequenced for 18


10
20
GLU VAL GLN LEU VAL GLU SER GLY GLY ASP LEU VAL LYS PRO GLY GLY SER LEU ARG LEU
CB-1

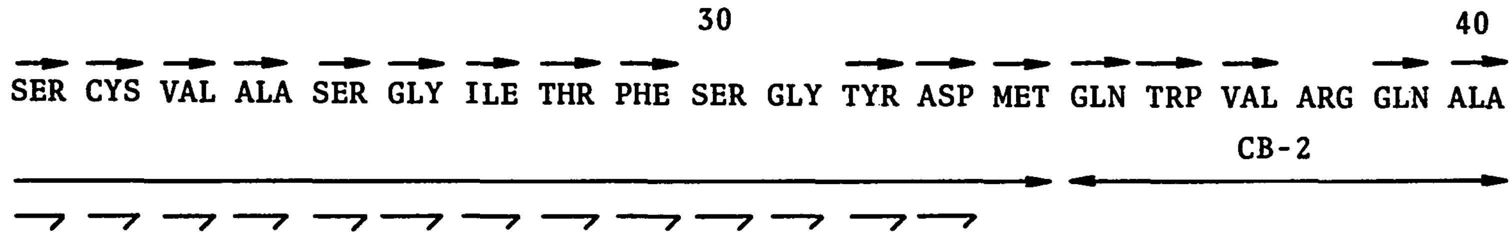

30
40
SER CYS VAL ALA SER GLY ILE THR PHE SER GLY TYR ASP MET GLN TRP VAL ARG GLN ALA
CB-2

FIGURE III-7

Sequence analysis of fragment CB-1 of alpha chain Gom. The sequence was established using automated Edman degradations on the intact heavy chain (⟶) and the isolated peptide (⟶).

TABLE III-2

Amino Acid Composition of Gom CB-1

Amino Acid	AAA*	SEQ+
CM-cysteine	0.9	1
Aspartic acid	2.8	2
Threonine	1.6	1
Serine	4.8	5
Glutamic acid	2.5	3
Proline	0.9	1
Glycine	5.4	6
Alanine	1.7	1
Valine	2.4	4
Isoleucine	1.3	1
Leucine	3.6	4
Tyrosine	1.3	1
Phenylalanine	1.4	1
Lysine	1.3	1
Arginine	1.1	1
Homoserine	+	1

* Amino Acid Analysis

\+ Composition derived from the sequence.
The sequence was determined as described in the text.

cycles after reaction with 4-sulfophenyl isothiocyanate to enhance the retention of the peptide in the sequencer cup (Inman et al., 1972). Peptides T-5, T-6, T-7, and T-8 were subjected to 2, 5, 4 and 7 cycles of Edman degradation, respectively. Peptide T-7 was also reacted with 4-sulfophenyl isothiocyanate before a second sequencer run. Peptides T-1, T-2, T-3, T-4, and T-6 were ordered by fitting their sequences to the sequence derived from the intact peptide. The two residue gap left between T-4 and T-6 was filled by T-5. T-8 was known to be the carboxy-terminal peptide because of the homoserine lactone which was found on amino acid analysis. Since all other peptides were accounted for, T-7 was therefore assigned the penultimate position. The ala obtained in step 41 of the sequencer run on the whole CB-2 fragment also places T-7 in this position. The complete sequence of CB-2 is shown in Figure III-8. The compositions of CB-2 and the tryptic peptides of CB-2 as determined by sequence and amino acid analysis are shown in Table III-3.

5. The sequence of the third variable region cyanogen bromide fragment of alpha chain Gom

CB-3, isolated as described in III-B-1, was subjected to 37 cycles of Edman degradation. Steps 1 thorugh 33 and step 37 were identified by gas chromatography, thin layer chromatography and amino acid analysis after hydrolysis of the phenylthiohydantoin amino acid with HI.

Approximately two micromoles of CB-3 were digested with trypsin (1% w:w) in 1% ammonium bicarbonate for 16 hours at 37°C. The resultant peptides were applied to a 2.5 x 240 cm column packed with Sephadex G-50 fine equilibrated with 0.5M ammonia. The elution profile is shown in Figure III-9. Peaks 12 and 13 contained peptide T-1 together with constant region peptides. Peptide T-1 was isolated by high voltage electro-

10 20

GLN TRP VAL ARG GLN ALA PRO GLY LYS GLY LEU GLN LYS VAL ALA TYR PHE ASN ASP ALA LEU SER ALA GLN GLY

CB-2

T-1 T-2 T-3 T-4

30 40

TYR ALA ASP ALA VAL LYS GLY ARG PHE THR ILE SER LYS ASP ASN ALA LYS ASP SER LEU TYR LEU GLN MET

T-5 T-6 T-7 T-8

FIGURE III-8

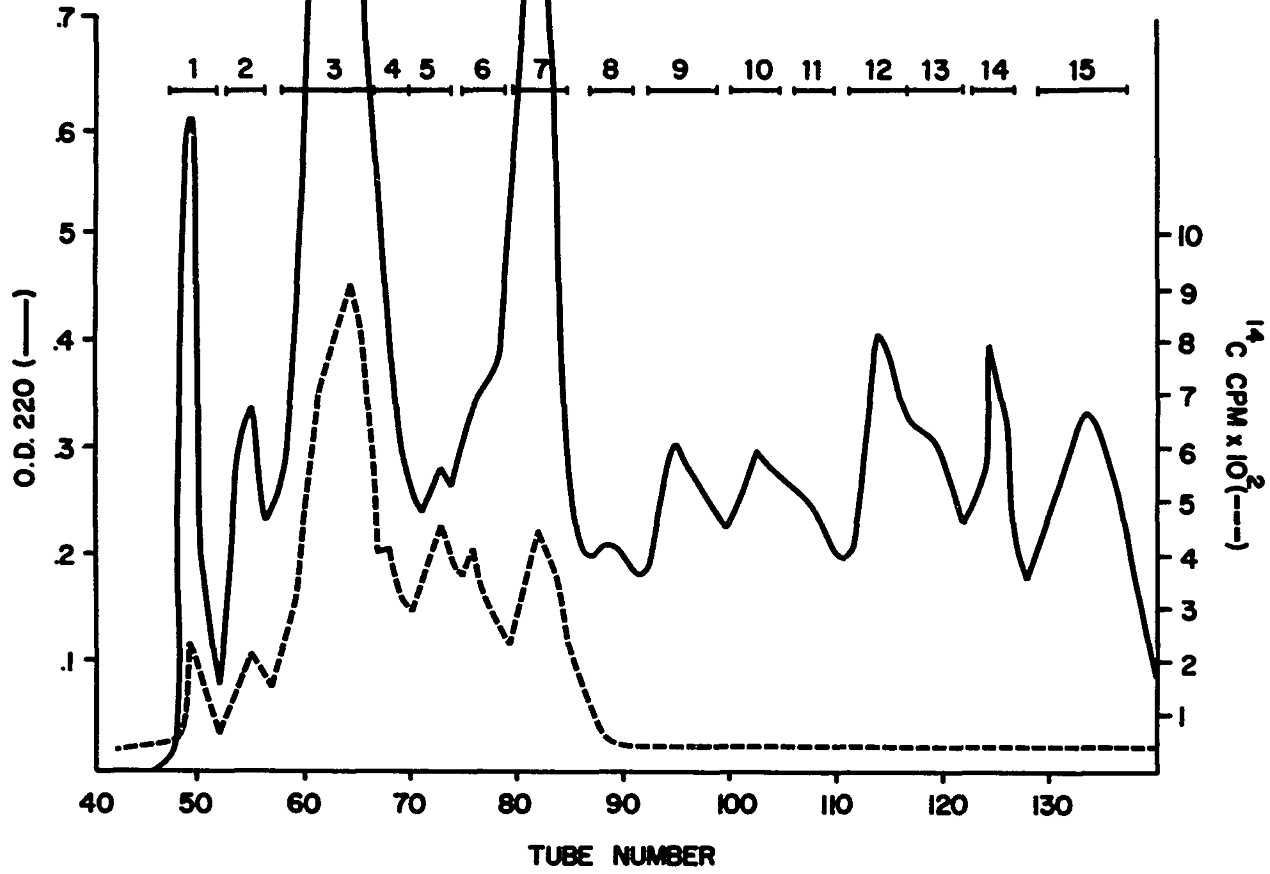
Sequence analysis of fragment CB-2 of alpha chain Gom. The sequence was established using automated Edman degradations on the intact CB-2 (⟶) and its peptides (⟶).

Table III-3

Amino Acid Composition of Gom CB-2 and Tryptic Peptides

Amino Acid	CB-2		T-1		T-2		T-3		T-4		T-5		T-6		T-7		T-8	
	AAA*	SEQ+	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ
Aspartic Acid	6.2	6			0.4				2.8	3	0.2		0.7		2.0	2	0.9	1
Threonine	2.5	1							1.2				0.8	1			0.8	1
Serine	3.0	3			0.7		0.7		2.0	1	0.4		2.0	1	1.4		0.4	
Glutamic Acid	4.7	5		1	1.1	1	1.0	1	1.5	1			0.3		0.6		1.0	1
Proline	1.2	1			1.0	1	0.3		0.6						0.5			
Glycine	3.7	4			1.6	1	1.7	1	1.8	1	1.9	1	0.6		2.3		0.5	
Alanine	6.3	7			1.2	1			3.2	5			0.5		1.8	1	0.4	
Valine	2.3	3		1	0.2				2.0	2							0.5	
Isoleucine	1.6	1			0.4				1.3				0.9	1			0.2	
Leucine	3.3	4			0.5		0.9	1	1.3	1			0.2		0.8		1.7	2
Tyrosine	3.8	3							1.6	2					0.6		0.8	1
Phenylalanine	2.1	2			0.3				1.1	1			1.4	1			0.5	
Lysine	3.4	5			0.7	1	1.0	1	0.8	1	0.4		1.0	1	1.5	1		
Arginine	2.2	2		1							1.0	1					0.3	
Homoserine	+	1															+	+
Tryptophan	ND	1		1	ND		ND		ND		ND		ND		ND		ND	
Mobility 2.1					-0.43		0.44		-0.75		1.02		0.25		0.44		-0.75	
Mobility 3.5							0.43		0.12						0.12		0.00	

* Amino Acid Analysis
+ Composition derived from the sequence. The sequence was determined as described in the text.
Mobility relative to lysine = +1, aspartic acid = -1, neutral = 0.



O.D. 220 (——)
^{14}C CPM x 10^{2} (----)
TUBE NUMBER
.7
.6
.5
.4
.3
.2
.1
10
9
8
7
6
5
4
3
2
1
40
50
60
70
80
90
100
110
120
130
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

FIGURE III-9

Chromatography of the tryptic peptides of fragment CB-3 of alpha chain Gom. The digest was applied to a 2.5 x 240 cm column of Sephadex G-50 fine equilibrated with 0.5M ammonia.

phoresis on Whatman No. 3 paper at pH 2.1 using 3000 volts for 30 minutes. The peptide was eluted with 0.5M NH_3. Peak 7 contained peptide T-2 in high yield and other peptides in very low yield. The material obtained from peak 7 was subjected to 29 cycles of Edman degredation and steps 1 through 26 were identified. The sequence of T-2 was identical to that obtained from the intact CB-3 peptide.

The sequence of CB-3 to the end of T-2 is shown in Figure III-10. The compositions of T-1 and T-2 as determined by sequence and amino acid analysis are shown in Table III-4. This peptide completes the sequence of the variable region and extends past the putative V/C bridge into the beginning of Cα1 domain.

C. The sequence of the variable region of a canine mu chain

Canine IgMλ myeloma Moo was purified as described in Materials and Methods. The purified material migrated as a single band on cellulose acetate electrophoresis and formed a single arc in an immunoelectrophoretogram developed with rabbit anti-whole canine serum. Heavy and light chains were separated on Sephadex G-100 equilibrated with 5M guanidine HCl after complete reduction and carboxymethylation. The heavy chain preparation migrated as a single band on polyacrylamide gel electrophoresis in sodium dodecyl sulfate.

1. Digestion of IgMλ pentamer with cyanogen bromide

Intact pentamer was digested with cyanogen bromide as described in Materials and Methods. The resulting peptides were initially separated on Sephadex G-100 equilibrated with 5M guanidine HCl. The elution profile is shown in Figure III-11. Peak 2 of the digest was resolved into two components by ion-exchange chromatography. Peak 2, dissolved

10 20

ASN SER LEU ARG ALA GLU ASP THR ALA VAL TYR TYR CYS ALA PRO TRP GLN PHE GLU TYR

CB-3

T-1 T-2

30

TRP GLY GLN GLY THR LEU VAL THR VAL SER SER GLU SER LYS

FIGURE III-10

Sequence analysis of the variable region portion of fragment CB-3 of alpha chain Gom. The sequence was established using automated Edman degradations on the intact peptide (⟶) and a tryptic fragment (⟶).

Table III-4

Amino Acid Composition of Variable Region Peptides of

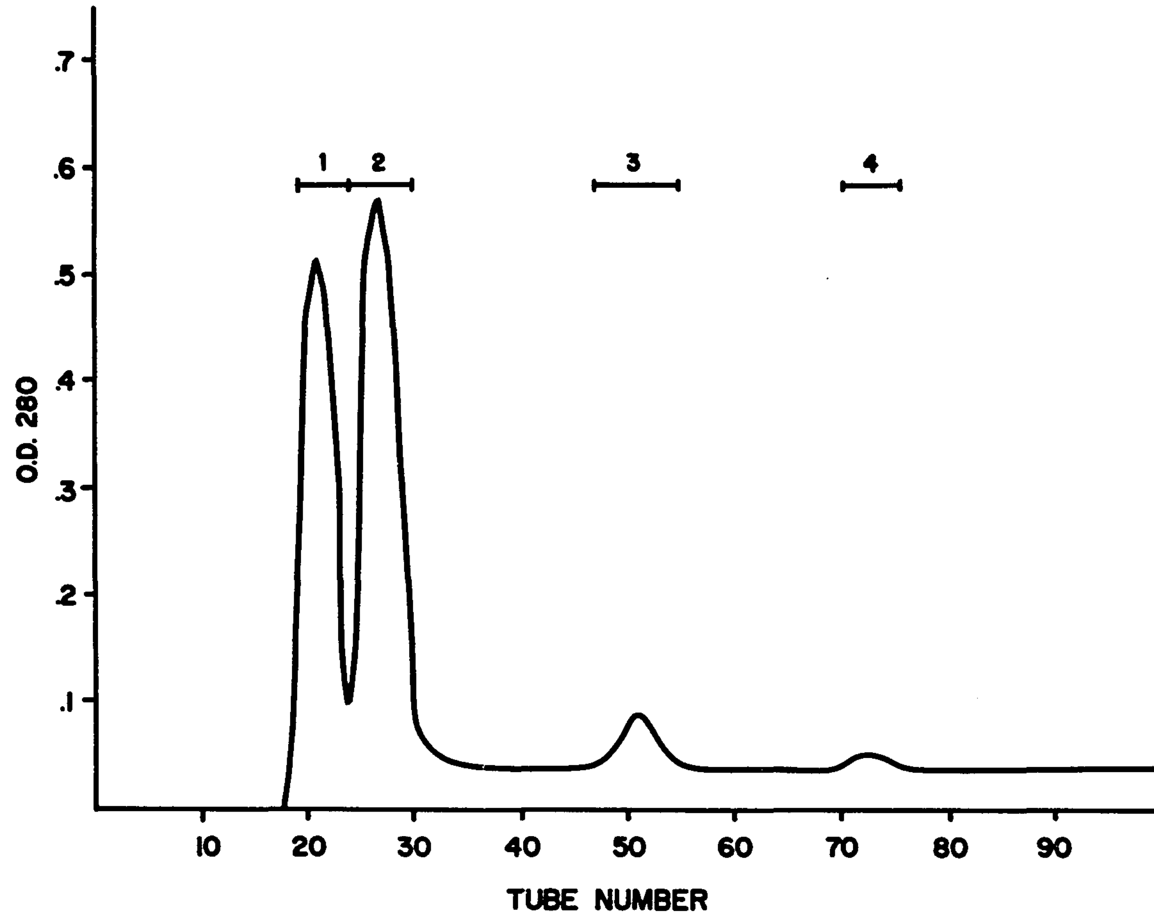
Gom CB-3

	CB-3 T-1		CB-3 T-2	
	AAA*	SEQ+	AAA	SEQ
CM-cysteine			0.7	1
Aspartic acid	0.9	1	2.3	1
Threonine	0.2		2.5	3
Serine	0.8	1	3.1	3
Glutamic acid	0.2		3.8	5
Proline			1.4	1
Glycine	0.2		1.8	2
Alanine	0.2		3.1	3
Valine			2.7	2
Isoleucine	0.2			
Leucine	0.8	1	1.4	1
Tyrosine			2.1	3
Phenylalanine	0.2		0.8	1
Lysine	0.2		1.0	1
Arginine	0.7	1		
Tryptophan	ND		ND	2
Mobility 2.1	0.32			

* Amino Acid Analysis

+ Composition derived from the sequence. The sequence was determined as described in the text.

Mobility relative to lysine = +1, aspartic acid = -1, neutral = 0.



O.D. 280
.7
.6
.5
.4
.3
.2
.1
1
2
3
4
10
20
30
40
50
60
70
80
90
TUBE NUMBER

FIGURE III-11

Chromatography of the cyanogen bromide peptides of IgMλ Moo. The digest was applied to a 5.0 x 150 cm column packed with Sephadex G-100 equilibrated with guanidine HCl.

in 6M urea, 0.05M sodium phosphate, pH 7.85, was applied to a column packed with Sephadex DEAE-A50 equilibrated with the same buffer. After washing and collecting the fall through, a gradient of 0.0M to 1.0M NaCl in the running buffer was applied to the column and a single peak eluted. Sequence analysis on the two separated groups of peptides showed them to be different; the material not retained by the ion exchanger contained two variable region peptides (CB-1 and CB-3); the peak eluted with sodium chloride contained only constant region peptides. Peak 3 of the digest was found to contain variable region peptide CB-2 uncontaminated by other peptides.

The variable region containing peptides derived from Peak 2 of the cyanogen bromide digest of the pentamer were reduced and carboxymethylated as described in Materials and Methods. The resulting peptides were separated on Sephadex G-75 equilibrated with 5M guanidine HCl. The elution profile is shown in Figure III-12. CB-3 was found in the fourth peak which also contained two constant region peptides. CB-3 was isolated from contaminants on Sephadex G-50 fine equilibrated with 0.5M ammonia.

2. The sequence of the amino terminal cyanogen bromide fragment of the variable region of mu chain Moo

Two hundred and fifteen nanomoles of heavy chain were subjected to 39 cycles of automated Edman degradation and definitive identifications were made for all positions except 31, 32, 35 and 38. A subsequent sequencer run on two hundred eighty nanomoles of heavy chain allowed the identification of the residues at positions 31 and 32. The complete sequence of this cyanogen bromide peptide and the overlap into

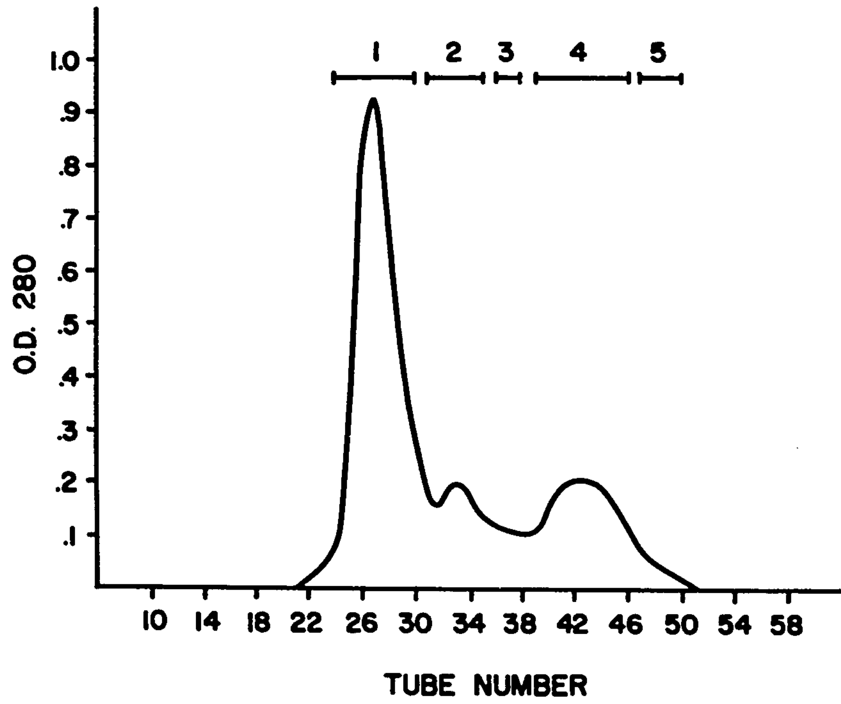

1
2
3
4
5
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
O.D. 280
10 14 18 22 26 30 34 38 42 46 50 54 58
TUBE NUMBER

FIGURE III-12

Chromatography of completely reduced and carboxymethylated cyanogen bromide fragments of the heavy chain of IgMλ Moo. These fragments were derived from peak 2 of the chromatography of the cyanogen bromide digest of the pentamer (see Fig. III-11) and was separated from other constant region peptides by ion-exchange chromatography as described in the text (vide supra III-C-1). The reduced and carboxymethylated peptides were applied to a 2.5 x 120 cm column packed with Sephadex G-75 equilibrated with 5.0M guanidine HCl.

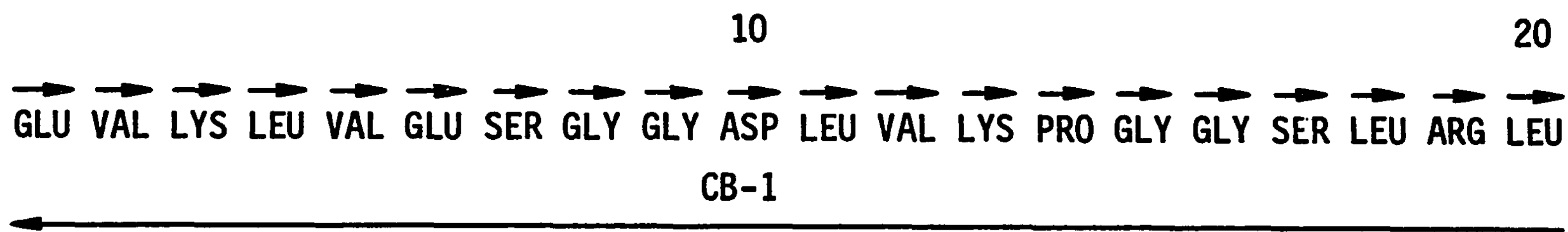
the next peptide are shown in Figure III-13. Since the sequence of this fragment was established by studies of the intact heavy chain this peptide was not isolated.

3. The sequence of the second cyanogen bromide fragment of the variable region of mu chain Moo

One micromole of the pure CB-2 peptide which was derived from the initial separation of the cyanogen bromide digest was subjected to forty cycles of automated Edman degradation and identifications made of every residue up to position 31. Positions 33 and 35 were also identified.

One micromole of CB-2 was digested with trypsin (1% w:w) in 1% ammonium bicarbonate for 16 hours at 37°C and the resulting peptides separated by high voltage electrophoresis on paper at pH 2.1 and pH 3.5. Peptide T-1 was not sequenced, since its sequence was determined on the sequencer run of the whole CB-2 peptide. T-2 could not be isolated from the paper. Peptides T-3, T-4, T-5, and T-6 were sequenced for 2, 4, 3, and 6 cycles respectively. The sequence of each of these peptides was derived from the sequencer runs and the compositions of the peptides. The assignment of the amide to the penultimate residue of peptide T-6 was based on the mobility of the peptide on paper electrophoresis.

In order to confirm the order of peptides T-3, T-4, T-5, and T-6, which was initially made by homology, one micromole of CB-2 was reacted with citraconic anhydride and then digested with trypsin (1% w:w) in 1% ammonium bicarbonate for 16 hours at 37°C. The peptide mixture was applied to a 1.5 x 200 cm column packed with Sephadex G-25 fine equilibrated with 0.5M ammonia. Two peaks were resolved by this


10
20
GLU VAL LYS LEU VAL GLU SER GLY GLY ASP LEU VAL LYS PRO GLY GLY SER LEU ARG LEU
CB-1

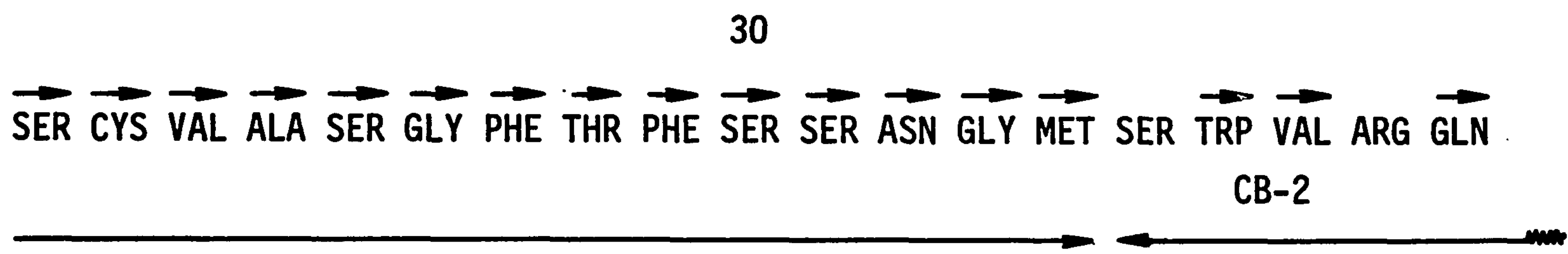

30
SER CYS VAL ALA SER GLY PHE THR PHE SER SER ASN GLY MET SER TRP VAL ARG GLN
CB-2

FIGURE III-13

Sequence analysis of fragment CB-1 of mu chain Moo. The sequence was established by sequential Edman degradation on the intact heavy chain (⟶).

procedure. The first peak was arbitrarily divided into an ascending and a descending portion. The ascending portion was found to contain CT-2 and the descending portion CT-4, both in high yield. CT-2 was sequenced for six cycles to establish its identification and purity. The placement of T-3 was confirmed by the composition of CT-2. CT-4 contained homoserine on amino acid analysis, placing it as the carboxy-terminal peptide. Six cycles of sequential Edman degradation showed that CT-4 was made up of T-5 and T-6, placing T-5 as the penultimate tryptic peptide of CB-2.

Although the sequence of T-2 was derived from the amino-terminal sequencer run, a peptide comprising the carboxyterminal portion of T-2 was sought to confirm unequivocally the sequence. CT-2, comprising T-2 and T-3, was digested with chymotrypsin for 2 hours in 1% ammonium bicarbonate and the resultant peptides were resolved by high voltage electrophoresis. The composition of CHY-3, the carboxyterminal peptide, agreed with the derived sequence.

The sequence of CB-2 is shown in Figure III-14. The composition of this peptide and its fragments as determined by the sequence and by amino acid analysis is shown in Table III-5.

4. The sequence of the third cyanogen bromide fragment of the variable region of mu chain Moo

The sequence of this portion of the variable region was determined by subjecting the peptide to 40 cycles of Edman degradation. All positions up to the V/C bridge were identified as shown in Figure III-15. Since this peptide contains substantial portions of the constant region its composition is not germane and is not included.

10 20

SER TRP VAL ARG GLN ASP PRO GLY GLU GLY LEU GLN TRP VAL ALA ASP ILE SER SER SER GLY GLN THR TYR TYR

CB-2

T-1 T-2

CT-1 CT-2

30 40

ALA ASP ALA VAL LYS GLY ARG PHE SER ILE SER ARG ASP ASN ALA LYS ASN THR LEU TYR LEU GLN MET

T-3 T-4 T-5 T-6

CT-3 CT-4

CHY-3

FIGURE III-14

Sequence analysis of fragment CB-2 of mu chain Moo. The sequence was established by sequential Edman degradation on the intact peptide (⟶) and its tryptic fragments (⟶).

TABLE III-5a

Amino Acid Composition of Moo CB-2 and Peptides

	CB-2		T-1		T-2		T-3		T-4		T-5		T-6	
	AAA*	SEQ+	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ
Aspartic Acid	5.5	6	0.2			3	1.2		0.3		2.0	2	0.9	1
Threonine	3.8	2				1			0.7		0.2		0.9	1
Serine	6.0	6	1.0	1		3	2.0		1.5	2	0.6		0.2	
Glutamic Acid	5.0	5	0.3			4			0.4		0.3		0.7	1
Proline	1.6	1				1								
Glycine	4.2	4	0.4			3	4.6	1	0.7		0.5		0.3	
Alanine	4.0	4	0.2			3	1.0		0.3		1.3	1	0.2	
Valine	2.1	3	0.7	1		2					0.2		0.3	
Isoleucine	1.9	2	0.2			1			1.0	1	0.3		0.2	
Leucine	3.5	3	0.3			1			0.5		0.4		2.0	2
Tyrosine	3.0	3				2					0.4		1.2	1
Phenylalanine	1.5	1							1.0	1	0.3		0.3	
Lysine	1.8	2				1					0.7	1		
Arginine	2.5	3	1.0	1			1.0	1	1.5	1				
Homoserine	+	1											+	
Tryptophan	ND	2	ND	1	ND	1	ND		ND		ND		ND	
Mobility 2.1					-0.37		0.43		-0.07		0.22		-0.37	
Mobility 3.5					0.02								-0.36	

* Amino Acid Analysis
+ Composition derived from the sequence. The sequence was determined as described in the text.
Mobility relative to lysine = +1, aspartic acid = -1, neutral = 0.

TABLE III-5b

Amino Acid Composition of Moo CB-2 and Peptides, continued

	CT-1		CT-2		CT-3		CT-4		CT-2 CHY-3	
	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ	AAA	SEQ
Aspartic Acid	0.2		2.8	3	0.3		2.0	3	1.1	1
Threonine			1.6	1	0.7		1.0	1		
Serine	1.0	1	2.6	3	1.5	2	1.0		0.2	
Glutamic Acid	0.3		3.0	4	0.4		1.3	1	0.2	
Proline			1.1	1			0.3			
Glycine	0.4		2.8	3	0.7		1.0		1.5	1
Alanine	0.2		2.0	3	0.3		1.1	1	1.9	2
Valine	0.7	1	1.7	2			0.6		1.0	1
Isoleucine	0.2		1.0	1	1.0	1	0.4			
Leucine	0.3		1.5	1	0.5		1.5			
Tyrosine			1.4	2			1.0	1	0.6	1
Phenylalanine			0.6		1.0	1	0.2			
Lysine			1.1	1			0.7	1	1.0	1
Arginine	1.0	1	0.7	1	1.5	1	0.2		1.0	1
Homoserine							+			
Tryptophan	ND	1	ND	1	ND		ND			
Mobility 2.1									0.48	
Mobility 3.5										

10 20

GLU ASP LEU ARG VAL GLU ASP THR ALA VAL TYR TYR CYS ALA THR GLU GLY ASP ILE GLU

CB-3

30

ILE PRO ARG TYR PHE GLY GLN GLY THR ILE

FIGURE III-15

Sequence analysis of the variable region portion of fragment CB-3 of mu chain Moo. The sequence was established by automated Edman degradation on the intact fragment. (⟶)

IV. DISCUSSION

A. Sequence studies of normal pooled IgG heavy chains

This study demonstrates the presence of the V_HIII subgroup in pooled IgG isolated from several mammalian species and shows that the inter-species distribution of this subgroup corresponds to established evolutionary patterns. In addition, the data confirm the presence of particular amino acids in certain positions of immunoglobulin V_HIII heavy chains from a number of mammalian species. The presence of these amino acids at a level of 95% or more at characteristic positions of pooled heavy chains of the various species is consistent with the earlier identification of phylogenetically associated residues in the heavy chain V_HIII subgroup on the basis of a study of myeloma proteins (Kehoe and Capra, 1972).

The class mammalia is currently divided into two subclasses, Prototheria (egg-laying mammals) and Theria (mammals that bear live young) (Johnson _et al._, 1969). No members of the Prototheria were available for study. The subclass Theria contains two infraclasses, Metatheria (pouched mammals) and Eutheria (placental mammals). The heavy chain result in the opossum (a pouched mammal) was of particular interest because, although the sequence of the pool showed unequivocally that this preparation was assignable to the V_HIII subgroup, the residue at position 2 was an isoleucine rather than the valine that is characteristic of V_HIII proteins

(Kohler et al., 1970; Capra, 1971). This undoubtedly represents a "phylogenetically associated" residue (Kehoe and Capra, 1972) and merits further analysis in other members of the Metatheria.

Among the Eutheria, three major groups were detectable on the basis of the relative predominance of the V_HIII subgroup (Table III-1 and Fig. III-2). The carnivores evidently utilize this subgroup to the virtual exclusion of the others, since members of this group possess 90% or more V_HIII proteins in their pools. In marked contrast, the artiodactyls (even-toed ungulates) except the pig, perissodactyls (odd-toed ungulates), cetaceae, and lagomorphs showed no detectable V_HIII proteins. The rodents, primates, and the pigs are intermediate and utilize the V_HIII subgroup to a level of approximately 25% (Table III-1).

The distribution of the V_HIII subgroup had a clear correlation with the Linnean grouping of the animal involved. For example, among the Artiodactyls, within the suborder Ruminantia, comprising two different families (the cow, sheep, and goat--suborder Ruminantia, family Bovidae; and the moose--suborder Ruminantia, family Cervidae) no V_HIII sequence was detectable in the IgG pool. In contrast, five different members of the order Carnivora (the cat--family Felidae; the dog--family Canidae; the mink--family Mustelidae; the sea lion--suborder Pinnipedia, family Otariidae; and the seal--suborder Pinnipedia, family Phocidae) had virtually 100% V_HIII proteins.

This pattern was equally apparent in the two orders, Rodentia and Primates, in which the V_HIII subgroup comprised approximately 25% of the total IgG pool. The narrowness of the V_HIII distribution (19-29%) among the various species examined in these orders was very striking (Fig. III-

2). Two different primate superfamilies (Cercopithecidea--old world monkeys; and Hominoidea--anthropoid apes and man) and two different rodent families (Caviidae--guinea pig; and Muridae--mouse and rat) have the same relative V_HIII subgroup distribution.

Any valid conception of the mechanisms responsible for the generation of antibody diversity must be reconcilable with these findings on the partitioning of the V_HIII subgroup among various mammals. There is currently general acceptance of the view that each variable region subgroup is encoded by at least one germ line gene. Assuming, for the moment, one such gene per subgroup, a satisfactory interpretation of the present data might be that four germ line genes (one for each of the four V_H subgroups) existed in a mammalian progenitor and that various phylogenetic groups have utilized each of these genes in varying amounts depending upon the particular selection pressures which have been operative on the respective groups at different times in evolutionary history. Certain groups (carnivores and marsupials) may have lost the V_HI, V_HII, and V_HIV genes, while others (ruminants and perissodactyls) may have deleted the V_HIII gene. Still others (primates and rodents) could have retained two, three, or all four, which may all be expressed in the present day variable region pool of these species.

It is also possible, of course, that a V_HIII gene was a very primitive V_H gene, which has given rise by gene duplication to the other subgroups. The data presented here could, from this point of view, be a reflection of the extent to which such duplication had occurred, varying from none (totally unblocked species) to extensive (totally blocked species), in which the putative V_HIII gene is presently represented to a

very limited extent, or not at all.

Multi-gene proponents might well look at these findings as consequences of large scale genetic expansion and contraction events (Milstein and Svasti, 1971; Hood and Prahl, 1971). We consider such an explanation unlikely because of the low probability that the V_HIII distribution among various species would follow the observed phylogenetic pattern. That is, it seems unlikely, on an a priori basis, that a pure germ line mechanism for diversity generation would lead to similarities in the V_HIII subgroup distribution within the following three groups of phylogenetically distant species: (a) marsupials, carnivores; (b) ruminants, perissodactyls, cetaceae, and lagomorphs; and (c) rodents, primates, and suiformes. This distribution could, however, be readily explained by assuming the existence of a relatively small number of V_H genes, some of which either never appeared or have been deleted during the course of evolutionary time in certain species. Such an interpretation could admittedly be complicated to a certain extent by the absolute confirmation of certain as yet disputed hypotheses in evolutionary zoology, such as, for example, the view that rodents and primates have an unexpectedly close phylogenetic relationship (Wood, 1972).

These findings on the V_HIII subgroup distribution should be compared and contrasted with those of Hood et al. (1971) concerning the distribution of kappa and lambda light chain types among various species. At the outset, the distinction between examining variable region subgroups and light chain types should be emphasized, since these are very different parameters. This is particularly true since, in contradistinction to the situation in heavy chains, where variable regions can associate with

different constant regions, the variable and constant regions of the two light chain types are not shared, as exemplified by the lack of association of kappa V regions with lambda C regions, and vice versa.

With these reservations, which are consistent with the known lack of genetic linkage between heavy and light chain genes (Dray et al., 1963), the light chain data can be compared with the present work. The distributional patterns of light chain type and the relative prevalence of the V_HIII subgroup among various species are not congruent. For example, although the light chain pool of both carnivores and perissodactyls is most, or all, of the lambda type, the carnivore heavy chains are overwhelmingly V_HIII, while the perissodactyl heavy chain pool has no detectable V_HIII proteins. Interestingly, animals overwhelmingly restricted to a single light chain type (cat, dog, horse, whale, sheep, goat, cow, rabbit) also showed either totally blocked or totally unblocked V_HIII heavy chains, while animals with significant amounts of both kappa and lambda chains (man, monkey, guinea pig, mouse, pig) possessed at least two heavy chain subgroups.

Additional information was obtainable from an analysis of the pool sequences themselves. Amino acid sequence data, obtained by others on human and mouse myeloma light chains (Hood and Talmage, 1970) and on light chains derived from rabbit homogeneous antibodies (Hood et al., 1970), have previously shown that most members of a single species have particular amino acid residues in certain locations in the chains. For example, position 11 in rabbit kappa chains from homogeneous antibodies has generally been valine, while mouse and human kappa chains contain leucine at this position. The identification and interpretation of such

residues have been complicated by the difficulty of specifying precisely what is implied by the terms "species specific" or "phylogenetically associated" (Kehoe and Capra, 1972). The residues in question are rarely, if ever, unique to a given species but are more accurately associated with broader level phylogenetic groupings. As an example, a residue that is shared by more than one species, such as valine in position 23 of the dog and cat V_HIII heavy chains (Fig. III-4), is very likely a residue that was present in the phylogenetic precursor of these two species, rather than being specific for either the dog or the cat [valine is, in fact, also found in position 23 of guinea pig heavy chains (Ray and Cebra, 1972)]. Thus the term phylogenetically associated would seem a more accurate designation for these residues than species specific.

The interpretation of the significance of phylogenetically associated residues in variable regions would obviously depend on whether the subgroups arose before or subsequent to speciation. Although this question has not yet been completely settled, Milstein and Svasti (1971) have proposed, on the basis of sequence studies of mouse kappa chains, that distinct subgroup genes existed <u>before</u> the divergence of the mouse-man evolutionary branches. This reinforces the necessity for considering phylogenetically associated residues on a subgroup-specific basis in both light chains and heavy chains (Kehoe and Capra, 1972), and implies, as a corollary, that such residues are not necessarily to be found in all members of either kappa or lambda chains within a given species. Nonetheless, the identification of such residues in any preparation of pooled immunoglobulin polypeptide chains would add greatly to their significance.

It has been difficult to search for such residues in light chain pools because of the multiple variable region subgroups characteristic of both kappa and lambda chains (Dreyer and Bennett, 1965). However, Novotny, Dolejs, and Franek (1972), in a study of pooled porcine lambda chains, have shown that the sequence of the amino terminal 23 residues of these proteins is highly uniform and contains several phylogenetically associated residues, as well as a deleted residue in position 5 that is unique to the pig. These authors concluded that their data were most compatible with a somatic process of variable region diversification.

The close structural similarities of the V_HIII heavy chains in the immunoglobulin pools analyzed in the present study have several important implications. As noted previously, it indicates that at least one basic V_HIII gene existed before mammalian speciation. Further, it has made possible the localization, in pools, of the same phylogenetically associated residues that have previously been identified in various heavy chains from myeloma proteins. The greater degree of preservation of primary structure, outside the hypervariable regions, seen in these heavy chains as compared with many light chains, may be a consequence of a predominant role for heavy chains in the formation of the antibody combining site. Comparison of the present data on mammalian and avian species with data available for the shark (Sledge et al., 1974) suggests that the V_HIII subgroup is ancient and has been well preserved during evolution.

The residue alternatives for the seven phylogenetically associated positions described in mammals all represent single base changes in the genetic code (e.g., position 1-Glu:Asp; 2-Val:Ile:Glu; 3-Gln:Lys; 10-Gly:Asp; 19-Lys:Arg; 21-Ser:Thr; 23-Ala:Val). In general, the

alternative amino acids are very similar in structure and properties. The alternatives seem to have been introduced independently, since there does not appear to be any particular linkage between the various positions. For example, the following associations can be observed for positions 10 and 23: Gly:Ala (man, mouse, rat); Gly:Val (guinea pig); Asp:Ala (mink, seal, sea lion); and Asp:Val (dog, cat).

B. The amino acid sequence of the variable regions of two homogeneous canine immunoglobulins

The complete amino acid sequence of the variable regions of alpha chain Gom and mu chain Moo are shown in Figure IV-1. The differences in the two sequences have been boxed. Figure IV-2 shows the sequence of the amino terminal twenty-four residues of pooled canine heavy chains and the seven homogeneous canine heavy chains that have been sequenced. There are a total of 10 out of 168 positions at which the sequence of the myeloma proteins differ from the pooled sequence. Proteins Ga and St have the same substitution of alanine for glycine at position 15 and proteins Gom and Moo both have lysine instead of glutamine at position 13. The prototype sequence, derived from all seven myelomas, is identical to the sequence of the pool. It is important to recognize that information derived from the sequence of myelomas reflects the normal statistically, and the statistical significance of such information is directly related to the number of myelomas proteins examined. For example, if only two proteins had been sequenced, (Ga and St, or Gom and Moo) the prototypic sequence derived from myelomas would be different from normal pooled material.

Figures IV-3a, IV-3b, and IV-3c display the two canine sequences

										10										20										30
Gom	GLU	VAL	GLN	LEU	VAL	GLU	SER	GLY	GLY	ASP	LEU	VAL	LYS	PRO	GLY	GLY	SER	LEU	ARG	LEU	SER	CYS	VAL	ALA	SER	GLY	ILE	THR	PHE	SER
Moo	GLU	VAL	LYS	LEU	VAL	GLU	SER	GLY	GLY	ASP	LEU	VAL	LYS	PRO	GLY	GLY	SER	LEU	ARG	LEU	SER	CYS	VAL	ALA	SER	GLY	PHE	THR	PHE	SER

										40									50										60
Gom	GLY	TYR	ASP	MET	GLN	( )	TRP	VAL	ARG	GLN	ALA	PRO	GLY	LYS	GLY	LEU	GLN	LYS	VAL	ALA	TYR	PHE	ASN	ASP	ALA	LEU	SER	ALA	GLN
Moo	SER	ASN	GLY	MET	SER	( )	TRP	VAL	ARG	GLN	ASP	PRO	GLY	GLU	GLY	LEU	GLN	TRP	VAL	ALA	ASP	ILE	SER	SER	SER	GLY	GLN	THR	( )

										70										80										90
Gom	GLY	TYR	ALA	ASP	ALA	VAL	LYS	GLY	ARG	PHE	THR	ILE	SER	LYS	ASP	ASN	ALA	LYS	ASP	SER	LEU	TYR	LEU	GLN	MET	ASN	SER	LEU	ARG	ALA
Moo	TYR	TYR	ALA	ASP	ALA	VAL	LYS	GLY	ARG	PHE	SER	ILE	SER	ARG	ASP	ASN	ALA	LYS	ASN	THR	LEU	TYR	LEU	GLN	MET	GLU	ASP	LEU	ARG	VAL

										100							110								
Gom	GLU	ASP	THR	ALA	VAL	TYR	TYR	CYS	ALA	PRO	(			)	TRP	GLN	PHE	GLU	TYR	TRP	GLY	GLN	GLY	THR	LEU
Moo	GLU	ASP	THR	ALA	VAL	TYR	TYR	CYS	ALA	THR	GLU	GLY	ASP	ILE	GLU	ILE	PRO	ARG	TYR	PHE	GLY	GLN	GLY	THR	ILE

FIGURE IV-1

Complete amino acid sequence of two canine heavy chain variable regions. The differences between the two sequences are boxed.

FIGURE IV-2

					5					10					15					20				
Canine Pool	E	V	Q	L	V	E	S	G	G	D	L	V	Q	P	G	G	S	L	R	L	S	C	V	A
Ga															A									
Le																				V				
Co																								
Lo												A	V											
St															A			I	G					
Gom													K											
Moo			K										K											

FIGURE IV-2

Amino terminal sequences of canine heavy chains. The sequence of pooled heavy chain is listed at the top. For the myeloma proteins, only those residues that differ from the pool have been shown.

FIGURE IV-3a

```
                          10        20        30        40
Human

    Tie   E V Q L V E S G G G L V Q P G G S L R L S C A A S G F T F S T S A V Y ( ) W V R
    Was   _______ L ________________________________________ S _____ D _ M _ ( ) _____
    Jon   D ____________________ K __________________________________ A W M K ( ) _____
    Zap   _______________ A ________________ G ______________________ T S R F ( ) _____
    Tur   _______ L ______________________________________________ R V L S S ( ) _____
    Nie   Z ________Q _______ V _______ R __________________________ R Y T I H ( ) _____
    Gal   _______________ D ___________ R ____________________(B V L B B F)M T ( ) _____

Canine

    Gom   _________________ D ___ K _________________ V _____ I _____ G Y D M Q ( ) _____
    Moo   ___ K ___________ D ___ K _________________ V _______________ N G M S ( ) _____
```

FIGURE IV-3b

```
                                 50                  60                  70                  80
Human

    Tie  Q A P G K G L E W V G W R Y E G S S L T H Y A V S V Q G R F T I S R N D S K N T

    Was  ___________________ A _ K _ Q E A _ N S _ F _ D T _ N _________________________

    Jon  ___________________ V ___ V _ Q V V E K A F _ N ___ N _________________________

    Zap  ___________________ E F _ V Q ___ A I S _____ D _____ A _______________________

    Tur  ___________________ S G _ L N A ___ N L _ F _____ A ___________________________

    Nie  ___________________ A V M S Y B G B B K _____ D ___ N _________________________

    Gal  ___________________ A N I K Z B G _ Z Z B _ V D ___ K _____________ D N A ___ S

Canine

    Gom  _____________ Q K _ A Y F N D A L _ A Q G ___ D A _ K ___________ K D N A _ D S

    Moo  _ D ___ E ___ Q ___ A D I S S S G Q T[ ]Y ___ D A _ K _____ S _____ D N A _____
```

FIGURE IV-3c

```
                              90                  100                 110
Human

    Tie   L Y L Q M L S L E P E D T A V Y Y C A R V T P A A A S L T F S A V W G Q G T L
    Was   _________ N R ___ A ___________________ F R Q P F V Q [ ] _ F D _ F _________
    Jon   _________ I _ V T ______________________ V V S T [     ] S M D ___________ P
    Zap   _________ N T G _ A ___________________ T R _ G G Y [  ] ___ D _____________
    Tur   _______________ Q A ___________________ L S V T _ V [  ] A F D ___________ K
    Nie   _____ N _ N ___ R ______________________ I R D T _ M [  ] _ F _ H ___________
    Gal   _________ N ___ R V ___________________ G W G [        ] G G D Y ___________

Canine

    Gom   _________ N ___ R A _________________ P W [            ] Q F E Y ___________
    Moo   _________ E D _ R V _________________ T E G D I E [    ] I P R Y F _______ I
```

FIGURE IV-3a, b, c

Heavy chain variable region sequences of seven human and two canine myeloma proteins of the V_HIII subgroup. The sequence of the human protein Tie is listed at the top. For the other proteins, only differences from Tie are shown.

along with the sequences of the only seven human V_HIII proteins completely sequenced (Capra and Kehoe, 1974a; Ponstingl *et al.*, 1970; Watanabe *et al.*, 1973). The human and canine myeloma proteins were unselected and have no known antibody activity.

The two dog proteins are identical at 76 of 113 positions. Of the 37 differences, 23 are in hypervariable regions and 11 are in framework residues. This degree of difference in both the hypervariable regions and the framework portions of the variable region is quite similar to the degree of difference found for any two unselected human myelomas of the V_HIII subgroup (Capra and Kehoe, 1974b). The locations of clustered differences in the two dog proteins are the same as the clusters of differences in human V_HIII proteins which define the hypervariable regions and, like the human V_HIII, differ from the human V_HII in the precise location of the second hypervariable region (V_HIII:52-61; V_HII:61-65).

In the framework portion of the variable region, all the differences reflect conservative, one base changes except the lysine/tryptophan interchange at position 49. Similar variation is seen in human myeloma proteins. Additionally, each canine sequence shows a high degree of homology to a prototypic framework sequence derived from the seven human myelomas (Gom 71 of 87 and Moo 72 of 87 framework residue identities). Positions 10, 13, 23, 48, 65, 67, 75, 76, and 77 are the same in both dog proteins and are different from the human prototype. Two of the differences, at positions 10 and 23, have been defined as phylogenetically associated residues based on the sequence of pooled dog heavy chains (vide supra, IV-A). These substitutions were also found in the 5 other

dog proteins for which there is data available (Kehoe and Capra, 1972). Position 13, however, while identical in Gom and Moo, is different in the other five proteins and the pool, and is therefore not a phylogenetically associated residue. The other six positions may represent phylogenetically associated residues but definite assignment requires more data, either on pooled material or several other myeloma proteins. It is interesting to note, however, that there are no paired differences (i.e. where the dog proteins contain the same difference relative to the human) after position 91, the end of the third hypervariable region, and only two of fifteen framework residues in this region have differences from the human. Thus, to date, neither subgroup specific (Kehoe and Capra, 1971) nor phylogenetically associated residues have yet been identified in the carboxy-terminal portion of the heavy chain variable region.

As has been previously noted, the only clusters of variability (less than 50% identity in a five residue sequence) in the two canine variable regions are located within the previously defined hypervariable regions. Each canine hypervariable region contains at least one amino acid difference between the two proteins requiring a two nucleotide interchange in the DNA codon for that position. The greatest variability is seen in the first, second and fourth hypervariable regions which have been shown to contribute to the antibody combining site (Padlan *et al*., 1973; Poljak *et al*., 1973).

The first hypervariable region of both proteins contains the relatively conserved methionine at position 34. The significance of this finding, even when considered with similar findings in other species (Capra and Kehoe, 1975), remains unclear.

The second hypervariable regions of the two canine proteins differ extensively and contain several amino acid differences, which each necessitate two nucleotide interchanges in the DNA. Protein Moo has a deletion at position 61. The finding of this deletion in the second hypervariable region confirms the finding, first suggested by the sequence of a murine heavy chain (Bourgois et al., 1972), that the size of the second hypervariable region is not subgroup specific. The lack of subgroup specific hypervariable region length poses significant problems for both the multi- and pauci-gene models of antibody diversity. Additional ad hoc mechanisms would be required to account for insertions or deletions in the middle of a gene by the mutational mechanisms suggested for either of these models.

Extreme variability and insertions or deletions are characteristic of the fourth hypervariable region. X-ray crystallographic studies (Padlan et al., 1973; Poljak et al., 1973) have shown that this hypervariable region forms a loop that is relatively distinct from the beta-sheet framework of the heavy chain variable region. This hypervariable region probably plays a major role in defining the size of the antibody combining site. The variation in length seen between the two canine proteins is consistent with variations seen in other species and is consistent with the presumed function of this portion of the molecule.

The third hypervariable region is the least variable of the hypervariable regions in the two canine proteins as well as in the human V_HIII proteins (Capra and Kehoe, 1974a). This hypervariable region does not contribute to the combining site of those proteins which have been studied by x-ray crystallography (Padlan et al., 1973; Poljak et al.,

1973). Sequence studies on the pooled heavy chains of rabbits and guinea pigs (Pratt and Mole, 1975; Cebra et al., 1974) have suggested that this hypervariable region is an expression of genetic polymorphism rather than a contributor to the combining site. The variability of the two canine proteins in this area is consistent with such an explanation, since the proteins came from animals of two distinct breeds.

C. Conclusions

There is a high degree of sequence homology in the V_HIII subgroup among mammals. Such conservation probably reflects the tertiary structural requirements necessary to support a functional antibody combining site. Despite structural requirements, both defined (i.e.: by evolutionary relationships) and random variations, unrelated to the combining site, exist.

The presence of the phylogenetically associated residues observed in heavy chain pools can be more easily reconciled with a pauci-gene than a multi-gene theory of antibody diversity. One would assume the existence of a single V_HIII gene, early in chordate evolution, which has been faithfully passed from one species to another along the phylogenetic scale. During this progression, point mutations within this gene occurred to generate those residues identifiable now as phylogenetically associated. For example, one could assume that for polypeptide position 10 of the V_HIII gene, a codon for aspartic acid existed in the precursor of both marsupials and other mammals. This codon has remained in all modern marsupials and carnivores, since members of these groups still have an aspartic acid at position 10 (Fig. III-4). However, this precursor gene underwent a point mutation at some time during the divergence of rodents and primates, and

all members of these two orders now have a glycine rather than aspartic acid in this position of their V_HIII proteins. Similar changes have presumably occurred in other regions of the V_HIII gene to yield the other phylogenetically associated residues now seen in other animals.

The data on pooled heavy chains seem especially difficult to reconcile with a pure germ line theory of antibody diversity. As soon as one postulates many genes in a progenitor, some ad hoc mechanism must be invoked to explain how, in certain instances, all of the genes in closely related modern species can contain information for different amino acids at a particular phylogenetically associated position. For example, both "multiple gene expansion and contraction" (Hood and Talmage, 1970) and "democratic gene conversion" (Gally and Edelman, 1970) have been proposed to explain the existence of variable region allotypes in rabbits and could equally well explain phylogenetically associated residues in the V region. However, both "multiple gene expansion and contraction" and "democratic gene conversion" assume that a large number of genes can be replaced by, or induced to change by, a single gene. An explanation of the pooled heavy chain data presented here would require these processes to have occurred in the immunoglobulin system on multiple occasions, and in relatively short stretches of evolutionary time, during emergence of the various higher animal species that can be studied today. This would seem a less likely development than the generation of the phylogenetically associated residues by simple mutational events operating on single V_HIII genes, under the assumption of a pauci-gene basis for antibody diversification.

The distribution of phylogenetically associated residues over much of the framework portion of the variable region, suggested by the finding

of six paired differences in the canine heavy chains beyond position 30, is significant. Phylogenetically associated residues reflect the forces of change to which the framework portion of the variable region is subject. The presence of phylogenetically associated residues in different portions of the framework is evidence that those portions of the framework are all controlled by the same genetic mechanism and influenced by evolutionary forces in the same manner. The absence of possible phylogenetically associated residues in the carboxyterminal portion of the variable region suggests that a structural feature, either of the DNA coding for this portion of the molecule or of the protein itself, exerts a negative selection pressure on such variability. Confirmation of these findings on the distribution of phylogenetically associated residues in certain portions of the variable region requires the examination of pooled heavy chains. Based on the results of the sequences of the Gom and Moo heavy chain variable regions and the subgroup distribution of pooled heavy chain variable regions, such studies would be most easily and most fruitfully undertaken on pooled canine immunoglobulin.

As has been previously stated, phylogenetically associated residues are explained in a multi-gene model only with difficulty. The identification of phylogenetically associated residues beyond the first hypervariable region increases these difficulties. However, while a pauci-gene model can easily account for phylogenetically associated residues in the framework portion of the variable region, an explanation for the absence of such markers in a particular part of the framework is not easily derived without invoking a special suppressor of variation for that region. In addition, there is no structural feature of the area from position 91 to

115, obvious from the x-ray crystallographic model, which makes it different from the area around the cysteine at position 28. (Phylogenetically associated residues have been described in the area around cysteine 22, vide supra, IV-A).

Length variation in the second hypervariable region (essentially the middle of the variable region) poses special problems for both the multi-gene and pauci-gene models of antibody diversity. Length variations in the first hypervariable region appear to be subgroup specific and the fourth hypervariable region is close to the putative V/C bridge. Subgroup specific length variation poses no problems for either model of antibody diversity. In both the pauci-gene and the multi-gene models an insertion or deletion occurring early in evolution (e.g.: at the time of the gene duplication which lead to the subgroups) would be expressed in all the genes which evolved from the altered gene. The length variation seen in the fourth hypervariable region can also be explained since this hypervariable region is so close to the putative V/C bridge. The V/C bridge may, in fact, be at the end of the fourth hypervariable region if one postulates that the low level of variability seen between positions 115 and 122 is a by-product of the V/C joining mechanism. Since the length variation in the second hypervariable region cannot be accounted for by either of these explanations, a new ad hoc mechanism is necessary to reconcile this finding with either model of antibody diversity.

The absence of phylogenetically associated residues after residue 91 and the confirmation of length variability in the second hypervariable region within the V_HIII subgroup require the introduction of additional mechanisms of genetic control into systems already burdened by theoretical explanations for which there is no biologic precedent.

These data resolve none of the questions relating to the generation of antibody diversity. They do help to further define the nature of immunoglobulin variability and the primary structural features of a unique system of broad capability and singular function.

APPENDIX

One letter and three letter amino acid abbreviations

Amino Acid	One Letter Code	Three Letter Code
Alanine	A	ALA
Arginine	R	ARG
Asparagine	N	ASN
Aspartic acid	D	ASP
Asparagine/aspartic acid unassigned	B	ASX
Cysteine (CM-Cysteine)	C	CYS
Glutamic acid	E	GLU
Glutamine	Q	GLN
Glutamic acid/Glutamine unassigned, Pyrrolidone-carboxylic acid	Z	GLX
Glycine	G	GLY
Histidine	H	HIS
Isoleucine	I	ILE
Leucine	L	LEU
Lysine	K	LYS
Methionine	M	MET
Phenylalanine	F	PHE
Proline	P	PRO
Serine	S	SER
Threonine	T	THR
Tryptophan	W	TRP
Tyrosine	Y	TYR
Valine	V	VAL

REFERENCES

Blomberg, G., Geckeler, W. R., and Weigert, M. (1972). Science 177, 178.

Bourgeois, A. and Fougereau, M. (1970). FEBS Lett 8, 265.

Bourgeois, A., Fougereau, M., and De Preval, C. (1972). Eur. J. Biochem. 24, 446.

Brient, B. W. and Nisonoff, A. (1970). J. Exp. Med. 132, 951.

Burnet, M. (1959). "The Clonal Selection Theory of Acquired Immunity," Vanderbilt University Press, Nashville, Tenn.

Capra, J. D. (1971). Nature (London) New Biol. 230, 61.

Capra, J. D. and Kehoe, J. M. (1974a). Proc. Nat. Acad. Sci. U. S. 71, 845.

Capra, J. D. and Kehoe, J. M. (1974b). Proc. Nat. Acad. Sci. U. S. 71, 4032.

Capra, J. D. and Kehoe, J. M. (1975). Adv. Immunol. 20, 1.

Capra, J. D. and Kindt, T. J. (1975). Immunogenetics 1, 417.

Capra, J. D., Tung, A. S., and Nisonoff, A. (1975). J. Immunol. 114, 1548.

Cebra, J. J., Koo, P. H., and Ray, A. (1974). Science 186, 263.

Cohn, M., Blomberg, B., Geckeler, W., Roschke, W., Riblet, R., and Weigert, M. (1974). In "The Immune System" (E. Sercarz, A. Williamson, and C. F. Fox, eds.), Academic Press, New York.

Crumpton, M. J. and Wilkinson, J. M. (1965). Biochem. J. 94, 545.

Cunningham, A. and Pilarski, L. (1974). In "The Immune System" (E. Sercarz, A. Williamson, and C. F. Fox, eds.), Academic Press New York.

Dray, S., Young, G. O., and Gerard, L. (1963). J. Immunol. 81, 142.

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

Edelman, G. M. (1959). J. Amer. Chem. Soc. 81, 3155.

Edelman, G. M. and Gall, W. E. (1969). Ann. Rev. Biochem. 38, 415.

Eichmann, K. and Berek, C. (1973). Eur. J. Immunol. 3, 599.

Feinstein, A. (1963). Nature (London) 199, 1197.

Fitch, W. M. and Margoliash, E. (1967). Science 155, 279.

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

Franek, F. (1971). Symposium on Developmental Aspects of Antibody Formation and Structure, Prague.

Gally, J., Edelman, G. M. (1970). Nature (London) 227, 341.

Grey, H. M. and Kunkel, H. G. (1964). J. Exp. Med. 120, 253.

Harris, Harry. (1970). The Principles of Human Biochemical Genetics. American Elsevier Publishing Company, Inc., New York.

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

Hochman, J., Inbar, D., and Givol, D. (1973). Biochemistry 12, 1130.

Hood, L., Eichmann, K., Lackland, H., Krause, R. M., and Ohms, J. (1970). Nature (London) 228, 1040.

Hood, L., Grant, J. A., and Sox, H. C., Jr. (1971). In "Developmental Aspects of Antibody Formation and Structure" (J. Sterzl and I. Riha, eds.), p. 283, Academia, Prague, Czechoslovakia.

Hood, L. and Prahl, J. (1971). Adv. Immunol. 14, 291.

Hood, L. and Talmage, D. W. (1970). Science 168, 325.

Hood, L., Waterfield, M. D., Morris, J., and Todd, C. W. (1970). Ann. N. Y. Acad. Sci. 190, 26.

Inbar, D., Hochman, J., and Givol, D. (1972). Proc. Nat. Acad. Sci. U. S. 69, 2659.

Inman, J. K., Hannon, J. E., and Appella, E. (1972). Biochem. Biophys. Res. Commun. 46, 2075.

Johnson, W. H., Delanney, L. E., Williams, E. C., and Cole, T. A. (1969). Principals of Zoology, Holt, Rinehart and Winston, New York.

Kehoe, J. M., Bourgois, A., and Fougereau, M. (1969). Nature (London) 224, 1212.

Kehoe, J. M. and Capra, J. D. (1971). Proc. Nat. Acad. Sci. U. S. 68, 2019.

Kehoe, J. M. and Capra, J. D. (1972). Proc. Nat. Acad. Sci. U. S. 69, 2052.

Kohler, H., Shimizu, A., Paul, C., Moore, V., and Putnam, F. W. (1970). Nature (London) 227, 1318.

Kuettner, M. G., Wang, A., and Nissonoff, A. (1972). J. Exp. Med. 135, 579.

Kunkel, H. G. (1954). Meth. Biochem. Anal. 1, 141-155.

Kunkel, H. G., Mannik, M., and Williams, R. C. (1963). Science 140, 1218.

Kunkel, H. G. and Prendergast, R. A. (1966). Proc. Soc. Exp. Biol. Med. 122, 910.

Leder, P., Honjo, T., Packman, S., Swan, D. and Nau, M. (1974). Proc. Nat. Acad. Sci. U. S. 71, 5109.

Lieberman, R., Potter, M., Mushinski, E. B., Humphrey, W., and Rudikoff, S. (1974). J. Exp. Med. 139, 983.

Milstein, C. (1966). Biochem. J. 101, 352.

Milstein, C. (1967). Nature (London) 216, 330.

Milstein, C. and Pink, J. R. L. (1970). Prog. Biophys. Mol. Biol. 21, 209.

Milstein, C. and Svasti, J. (1971). Prog. Immunol. 1, 33.

Milstein, C. P., Steinberg, A. G., McLaughlin, C. L., and Solomon, A. (1974). Nature (London) 248, 160.

Mole, L. E., Jackson, S. A., Porter, R. R., and Wilkinson, J. M. (1971). Biochem. J. 124, 301.

Murgola, E. J. and Yanofsky, S. (1974). J. Mol. Biol. 86, 775.

Natvig, J. B. and Kunkel, H. G. (1973). Adv. Immunol. 16, 1.

Novotny, J., Dolejs, L., and Franek, F. (1972). Eur. J. Biochem. 31, 277.

Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S., and Potter, M. (1973). Nature (London) New Biol. 245, 165.

Pawlak, L. L., Mushinski, E. B., Nisonoff, A., and Potter, M. (1973). J. Exp. Med. 137, 22.

Pisano, J. J. and Bronzert, T. J. (1969). J. Biol. Chem. 244, 5597.

Poljak, R. J., Amzel, L. M., Avery, H. P., Chen, B. L., Phizackerley, R. P., and Saul, F. (1973). Proc. Nat. Acad. Sci. U. S. 70, 3305.

Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P., and Saul, F. (1974). Proc. Nat. Acad. Sci. U. S. 71, 3440.

Ponstingl, H., Schwarz, J., Reichel, W., and Hilschmann, N. (1970). Hoppe-Seyler's Z. Physiol. Chem. 351, 1591.

Porter, R. R. (1959). Biochem. J. 73, 119.

Pratt, D. and Mole, L. E., Biochem. J., in press.

Premkumar, E., Shoyab, M., and Williamson, A. R. (1974). Proc. Nat. Acad. Sci. U. S. 71, 99.

Putnam, F. W., Titani, K., and Whitley, E. (1966). Proc. Roy. Soc. (Lond.) Ser. B. 166, 124.

Ray, A. and Cebra, J. J. (1972). Biochemistry 11, 3647.

Rudikoff, S. and Potter, M. (1974). Biochemistry 13, 4033.

Scheidegger, J. J. (1955). Int. Arch. Allergy. 7, 103.

Sercarz, E. E., Williamson, A. R., and Fox, C. F., eds. (1974). The Immune System, Academic Press, New York.

Schiechl, H. and Hilschmann, N. (1971). Hoppe-Seyler's Z. Physiol. Chem. 352, 111.

Sher, A. and Cohn, M. (1972). Eur. J. Immunol. 2, 319.

Sledge, C., Clem, L. W., and Hood, L. (1974). J. Immunol. 112, 941.

Smith, G. P. (1973). The Variation and Adaptive Expression of Antibodies, Harvard University Press, Cambridge, Mass.

Smithies, O., Gibson, D. M., Fanning, E. M., Goodfliesch, R. M., Gilman, I. G., and Ballantyne, D. L. (1971). Biochemistry 10, 4912.

Stelos, P. (1967). In "Handbook of Experimental Immunology," (D. M. Weir, ed.), F. A. Davis Co., Philadelphia.

Summers, M. R., Smythers, G. W., and Oroszlan, S. (1973). Anal. Biochem. 53, 624.

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

Todd, C. W. (1963). Biochem. Biophys. Res. Commun. 11, 170.

Tomasi, T. B., Jr. (1972). New Eng. J. Med. 287, 500.

Tonegawa, S., Sternberg, C., Dube, S., and Bernardi, A. (1974). Proc. Nat. Acad. Sci. U. S. 71, 4027.

Temin, H. and Baltimore, D. (1972). Adv. Virus. Res. 17, 129.

Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W., and Weigele, M. (1972). Science 178, 871.

Watanabe, S., Barnikol, H. V., Horn, J., Bertram, J., and Hilschmann, N. (1973). Hoppe-Seyler's Z. Physiol. Chem. 354, 1505.

Wood, A. E. (1972). Science 176, 437.

Wu, T. T. and Kabat, E. A. (1970). J. Exp. Med. 132, 211.

Yasmeen, D., Ellerson, J. R., Dorrington, K. J., and Painter, R. H. (1973). J. Immunol. 110, 1706.