ANTIBODY DIVERSITY

Thesis by

James Walter Schilling, Jr.

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(submitted June 26, 1980)

To my parents

for everything

they have done

ACKNOWLEDGEMENTS

Many thanks are due to all of those people who have made this work possible:

to my many teachers over the years who have imparted knowledge and answered so many questions;

to Lee Hood for five years of enthusiasm, concern, support, advice, and scientific rigor;

to Mike Hunkapiller and Vince Farnsworth for instruction in the techniques of protein chemistry;

to Brian Clevinger, Dan Hansburg, and Joe Davie for a long and fruitful collaboration on dextran antibodies;

to Martin Weigert and his coworkers for $V_{\kappa}21$ proteins;

to Bertha Jones, for keeping the Hood lab running so smoothly;

to the many other members of the Hood lab for their help and humor;

and finally to all the people in the stockroom and offices at Caltech on whom we are all dependent every day.

This work has been supported by an Earle C. Anthony Fellowship and NIH Training Grants GM00086 and 07616. Preparation of this thesis was supported by the Jean Weigle Memoral Fund.

Abstract

Vertebrate organisms possess a large and diverse repertoire of antibody variable regions. A number of different genetic mechanisms have been proposed to account for immunoglobulin variable (V) region diversity, including multiple germline genes, somatic mutation, somatic recombination, and multiple small gene segments which are joined to form a complete variable region gene segment. Analyses of variable region amino acid sequences demonstrate the relative contribution of each of these mechanisms to antibody diversity.

Twenty-four $V_\kappa 21$ chains have been examined. They suggest that the kappa chain variable region is encoded in two separate gene segments: V_κ and J_κ which are rearranged and joined during B cell differentiation. Diversification of the N terminus of the J_κ segment occurs as a consequence of V_κ-J_κ joining and has been explained by a site-specific recombination model. The amino acid sequence data are consistent with the existence of a minimum of six V_κ and five J_κ germline gene segments. Possible cases of somatic mutation are also observed. These conclusions are supported by nucleic acid sequence analyses performed by others.

Complete variable region amino acid sequences have been determined for twenty-one heavy chains from dextran binding antibodies. These sequences suggest that the heavy chain variable region is encoded by three gene segments: V_H, D, and J_H. Nucleic acid sequence analyses are consistent with this conclusion. The existence of a minimum of two V_H and four J_H germline gene segments is suggested by these sequences. Possible examples of somatic mutation of V_H and J_H gene

segments have also been found. Diversification of the N-terminal residue of the J_H segment may occur as a consequence of D-J_H joining by a mechanism analogous to that observed in kappa chains. Although comprised of only two residues, the D segment is the most diverse portion of dextran binding heavy chains.

Combinatorial joining of V_κ and J_κ gene segments and V_H, D, and J_H gene segments contributes significantly to antibody diversity.

Precise molecular locations of idiotypic determinants can be established in the dextran heavy chains. A cross-reactive idiotypic determinant (IdX) is located in the second hypervariable region of the V_H segment. Individual idiotypic determinants (IdIs) correspond to particular D segments.

TABLE OF CONTENTS

INTRODUCTION

The antibody molecule plays a major role in protecting vertebrate organisms against infection and disease. A single antibody exhibits a high degree of specificity for binding to a corresponding antigenic determinant or hapten. The specificity of antibody-antigen interaction combined with the large number of different antigens to which a single organism can raise an antibody response implies that the organism's repertoire of antibodies is extremely diverse. For a number of years immunologists have concerned themselves with attempting to construct theories to explain the origin of antibody diversity. This is the problem with which I have been concerned for the past five years. During that time the techniques of protein and nucleic acid sequence analysis have provided many insights into the multiple mechanisms involved in generating antibody diversity. I will now i) briefly review antibody structure and genetics, ii) review the major alternative theories of the origin of antibody diversity which have been proposed and iii) summarize briefly the approach I have taken to contribute to a resolution of this problem.

Antibody Structure and Genetics (1)

Antibody molecules are composed of two different polypeptide chains: heavy (H) chains of 50,000-70,000 daltons molecular weight and light (L) chains of approximately 25,000 daltons. Comparisons of amino acid sequences of both types of chains revealed that each chain is composed of an N-terminal variable (V) region approximately 110 amino acids long and a C-terminal constant (C) region 110 amino acids long in light chains and 440 or 550 amino acids long in heavy chains. The V_H and V_L regions combine to form the antigen binding site while the

C_H region mediates various effector functions which result in the elimination of antigen from the organism.

In the mouse there are two types of light chains: kappa (κ) and lambda (λ). Each type of light chains is encoded by a separate and unlinked gene family. As originally proposed by Dreyer and Bennett (2) and subsequently confirmed by nucleic acid analysis of antibody genes (3,4) V_L and C_L regions are encoded in germline DNA by separate V_L and C_L gene segments. During differentiation of the antibody synthesizing cell (the B cell) one V_L gene segment is rearranged to a new location adjacent to a C_L gene segment to create an actively transcribing light chain gene.

The mouse heavy chains are encoded by a third unlinked immunoglobulin gene family. This family is presently known to contain a minimum of eight different C_H gene segments: μ, δ, $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$, α, ε. The particular type of heavy chain constant region present in an antibody molecule determines the class of that molecule. Thus a molecule of the IgM class contains μ heavy chains, an IgG class molecule γ heavy chains, IgA class molecules α chains, etc. The IgG class is a tetramer with the composition $\gamma_2 L_2$. Other classes exist as multimers of H_2L_2 subunits. For example, the IgM class is a pentamer with the composition $(\mu_2 L_2)_5$.

Comparisons of heavy and light chain variable region amino acid sequences demonstrate that both types of variable regions contain four regions of relatively conserved amino acid sequence, termed framework regions, which are separated by three regions of highly variable amino acid sequence, termed hypervariable regions (5,6). Three-dimensional structures of antibody molecules derived by X-ray crystallographic analysis demonstate that the hypervariable regions fold to form the

walls of the antigen binding site (12,13).

Theories of Antibody Diversity

A number of different genetic mechanisms had been postulated by 1975 to contribute to antibody diversity (9). I have divided them into four general types and summarized them briefly here. As will be discussed below each of these four types of theories makes specific predictions for the diversity patterns which might be observed when a number of different antibody variable region sequences are compared.

Classical germline theory (10). This is the simplest of the four theories and states that each different heavy and light chain variable region sequence is encoded by a different germline gene. The existence of several thousand germline variable region genes was predicted in this theory.

Somatic mutation theories (11-14). A number of different versions of somatic mutation theories have been advanced. They all share a common element in the belief that the number of antibody variable region amino acid sequences is greater than the number of germline antibody gene sequences. The amino acid sequences not directly encoded by germline genes are postulated to arise by either ordinary mutation or special mutational mechanisms during proliferation of the precursors to antibody secreting cells during ontogeny. The precise number of germline genes postulated varies from less than ten to a few hundred in different versions of the theory.

Germline genes and somatic mutation can be identified from an analysis of amino acid sequence diversity patterns. It was generally agreed by classical germline and somatic mutation theorists that a germline variable region gene would be defined by finding two inde-

pendently isolated heavy or light chains with identical variable region amino acid sequences. This relies upon the parallel mutation argument which states that two independent cell lineages are unlikely to follow parallel mutation pathways which would yield identical variant genes in both lineages. Variant sequences derived from a particular germline sequence are expected to differ from one another and their common germline precursor by one or more amino substitutions, with each variant sequence being unique. Mutations which occur in hypervariable regions might directly influence the affinity of the antibody for antigen. Cells synthesizing variant antibodies with higher antigen binding affinity could be selectively stimulated to proliferate by virtue of this increased affinity. Such selection could potentially permit the observation of variant proteins even if somatic mutation occurs at a low frequency.

Somatic recombination (15, 16). This theory assumes the existence of a moderate number, perhaps several hundred, tandemly arrayed variable region genes. Unequal crossing over among these genes in the precursors to antibody secreting cells during ontogeny would result in novel variable region gene sequences. This theory predicts an unique diversity pattern for variable region amino acid sequences. Multiply repeated sequences will indicate unrecombined germline genes. Variant sequences resulting from unequal crossing over would be identified as having the sequence of one germline gene from the N terminus to some arbitrary position and the sequence of a second germline gene from that point to the C terminus. The precise recombination point could occur anywhere in the variable region sequence.

Multiple gene segment or minigene theories (17-19). These theories postulate that the complete variable region does not exist as a single

contiguous sequence of nucleotides in germline DNA but rather is encoded by a number of short gene segments or minigenes. Each of these gene segments encodes the complete sequence of one framework or one hypervariable region. During ontogeny framework and hypervariable region gene segments are rearranged and joined to form one complete variable region gene which is joined to a constant region gene and transcriptionally active. The amino acid sequence diversity pattern predicted by these theories is similar to that predicted by somatic recombination. However, gene segment theories predict multiple apparent recombinations at precise locations (framework-hypervariable region boundaries) and can thus be distinguished from the somatic recombination theory.

Molecular mechanisms contributing to diversity (10). Two non-genetic mechanisms may contribute significantly to antibody diversity. If any heavy chain can associate with any light chain then combinatorial association of heavy and light chains will greatly amplify antibody diversity. Thus 10^8 different antibody binding sites could be created from combinatorial association of 10^4 heavy and 10^4 light chains. If a single antibody molecule can bind more than one antigen such multispecificity could also greatly increase the antigen binding repertoire of a given antibody population.

At the time I came to Caltech partial N-terminal amino acid sequences had been determined for a number of heavy and light chains obtained from homogeneous antibodies from myeloma tumors (9,20-22). These tumors were available in large numbers from the NZB and BALB/c strains of mice. These early sequence studies demonstrated the extensive sequence diversity of mouse antibodies and the existence of multiple germline heavy and light chain variable region genes.

Several other laboratories have examined mouse light chain diversity by nucleic acid sequence analysis (16,23-25). Their conclusions are consistent with those derived from our analysis of the diversity patterns of kappa chain amino acid sequences.

As work on the $V_{\kappa 21}$ system progressed it became apparent that an analysis of a closely related set of heavy chains could also be rewarding. We settled on the set of heavy chains of BALB/c antibodies which bind the simple hapten α(1,3)-dextran for several reasons. The immune response of mice to this antigen had been well characterized (26-29). A special immunization protocol had been developed which elicited high titers of serum anti-dextran antibodies (27). Quantities of anti-dextran antibody sufficient for amino acid sequence analysis could be obtained from one or a few mice, a property not found with other BALB/c antibodies. Chapter 3 presents the results of N-terminal sequence analysis of four pools of BALB/c anti-dextran antibodies which demonstrate that dextran heavy chains do represent a closely related set of proteins.

For further sequence analysis work it was clearly desirable to obtain homogeneous dextran binding antibodies. Two BALB/c myelomas were available: M104E and J558. With the advent of the hybridoma technique in which antibody secreting spleen cells are fused with myeloma cells (30) an unlimited number of immortal cell lines secreting homogeneous dextran binding antibodies could be obtained. Chapters 4 and 5 present complete variable region amino acid sequences for two myeloma and nineteen hybridoma heavy chains. The diversity patterns observed suggested that the heavy chain variable region is encoded in three gene segments: V_H, D, and J_H. This conclusion has subsequently been supported by nucleic acid sequence analysis

Having superficially examined the scope of the mouse repertoire the next logical step seemed to be a detailed analysis of sets of closely related variable region sequences. In choosing a closely related set of proteins for study it is possible to limit one's study to a small number of germline genes. With sufficient data it should be possible to identify all germline genes for the set of proteins, even those which differ by only a few amino acid substitutions. Once the germline contribution to diversity is known it should be possible to discover the relative contributions of other mechanisms of diversification to total antibody diversity. My experimental approach to the problem has therefore been extensive amino acid sequence analysis of two closely related sets of proteins: the $V_{\kappa 21}$ group of light chains of the NZB mouse and the α-1,3-dextran binding heavy chains of the BALB/c mouse.

Chapters 1 and 2 present results of sequence analysis of 24 different NZB $V_{\kappa 21}$ myeloma light chains. Analysis of the amino acid sequence diversity patterns as discussed above led to a number of conclusions regarding the origin of diversity in these proteins. First, it was apparent that the classic variable region is encoded in two separate germline gene segments, V_κ and J_κ. Second, there are multiple germline V_κ and J_κ gene segments. Third, combinatorial joining of V_κ and J_κ can contribute significantly to antibody diversity. Fourth, diversity at the N terminus of the J_κ segment can be produced as a consequence of the rearrangement of V_κ and J_κ and can be explained by a site specific recombination model. Fifth, several potential examples of somatic mutation of V_κ and J_κ segments have been found.

(31). Multiple germline V_H and J_H gene segments appear to exist and possible examples of somatic mutation in both these gene segments have been observed. Diversification of the N-terminal residue of the J_H segment may occur as a result of gene segment rearrangement, analogous to what has been observed in kappa chains (Chapter 2). The D segment is only two amino acid residues long and is the most diverse portion of dextran binding heavy chains with twelve different sequences observed in the twenty-one proteins examined. Combinatorial joining of V_H, D, and J_H gene segments makes a major contribution to antibody diversity. One of the most striking features of the dextran binding heavy chains is that we have found twenty-one different variable region sequences in twenty-one heavy chains. Since all dextran antibodies examined appear to have identical λ light chains combinatorial association of heavy and light chains does not appear to contribute to the diversity of anti-dextran antibodies.

Idiotypes are antigenic determinants located on antibody variable regions. They have been used as markers for genetic mapping of antibody genes (32,33). Jerne (34) has suggested in his network theory that idiotypes may be involved in the regulation of the immune system. In Chapter 4 and Appendix 1 the first molecular characterization of idiotype is presented for idiotypes found on dextran binding antibodies. The finding that idiotypes are located on both V_H and D segments may have important implication for future interpretation of genetic mapping data.

All of the studies presented here have depended heavily on advances made in amino acid sequence analysis during the past five years at Caltech (35,36). I have been involved in this effort during this time.

The paper of Appendix 2 is presented as an example of the application of this technology to other, non-immunoglobulin proteins.

References

1. Hood, L. E., I. L. Weissman and W. B. Wood (1978) Immunology Menlo Park, CA: Benjamin/Cummings Publishing Co., Inc.
2. Dreyer, W. J. and J. C. Bennett (1965) Proc. Natl. Acad. Sci. USA 54: 864.
3. Brack, C., M. Hirama, R. Lenhard-Scheller and S. Tonegawa (1978) Cell 15: 1.
4. Seidman, J. G., E. E. Max and P. Leder (1979) Nature 280: 370.
5. Wu, T. T. and E. A. Kabat (1970) J. Exp. Med. 132: 211.
6. Capra, J. D. and J. M. Kehoe (1974) Proc. Natl. Acad. Sci. USA 71: 845.
7. Amzil, L. M., R. J. Polhak, F. Saul, J. M. Varga and F. F. Richard (1974) Proc. Natl. Acad. Sci. USA 71: 1427.
8. Padlan, E. A., D. M. Segal, T. F. Spande, D. R. Davies, S. Rudikoff and M. Potter (1973) Nature New Biol. 245: 165.
9. Hood, L., E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg and J. Schilling (1977) Cold Spring Harb. Symp. Quant. Biol. 41: 817-836.
10. Hood, L. and D. Talmage (1970) Science 168: 325.
11. Cohn, M., B. Blomberg, W. Geckeler, W. Raschke, R. Riblet and M. Weigert (1974) In The Immune System: Genes Receptors, Signals (ed. E. Sercarz et al.) p. 89, New York: Academic Press.
12. Jerne, N. K. (1971) Eur. J. Immunol. 1: 1.
13. Brenner, S. and C. Milstein (1966) Nature 211: 242.
14. Baltimore, D. (1974) Nature 248: 409.
15. Gally, J. A. and G. M. Edelman (1970) Nature 227: 341.

16. Seidman, J. G., A. Leder, M. Nau, B. Norman and P. Leder (1978) Science 202: 11.

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

18. Kindt, T. J. and J. D. Capra (1978) Immunogenetics 6: 309.

19. Kabat, E. A., T. T. Wu and H. Bilofsky (1978) Proc. Natl. Acad. Sci. USA 75: 2429.

20. Barstad, P., J. Hubert, M. Hunkapiller, A. Goetze, J. Schilling, B. Black, B. Eaton, J. Richards, M. Weigert and L. Hood (1978) Eur. J. Immunol. 8: 497.

21. Loh, E., J. M. Hood, R. Riblet, M. Weigert and L. Hood (1979) J. Immunol. 122: 44.

22. Loh, E., B. Black, R. Riblet, M. Weigert, J. M. Hood and L. Hood (1979) Proc. Natl. Acad. Sci. USA 76: 1395.

23. Max, E. E., J. G. Seidman and P. Leder (1979) Proc. Natl. Acad. Sci. USA 76: 3450.

24. Sakano, H., K. Hüppi, G. Heinrich and S. Tonegawa (1979) Nature 280: 288.

25. Seidman, J. G., A. Leder, M. H. Edgell, F. Polsky, S. M. Tilghman, D. C. Tiemeier and P. Leder (1978) Proc. Natl. Acad. Sci. USA 75: 3881.

26. Geckeler, W., B. Blomberg, C. de Preval and M. Cohn (1977) Cold Spring Harb. Symp. Quant. Biol. 41: 743.

27. Hansburg, D., D. E. Briles and J. M. Davie (1976) J. Immunol. 117: 569.

28. Hansburg, D., D. E. Briles and J. M. Davie (1977) J. Immunol. 119: 1406.

29. Hansburg, D., R. M. Perlmutter, D. E. Briles and J. M. Davie (1978) Eur. J. Immunol. 8: 352.

30. Köhler, G. and C. Milstein (1975) Nature 256: 495.

31. Early, P., H. Huang, M. Davis, K. Calame and L. Hood (1980) Cell 19: 981.

32. Riblet, R., B. Blomberg, M. Weigert, R. Lieberman, B. A. Taylor and M. Potter (1975) Eur. J. Immunol. 5: 775.

33. Weigert, M. and R. Riblet (1977) Cold Spring Harb. Symp. Quant. Biol. 41: 837.

34. Jerne, N. K. (1974) Ann. Immunol. (Inst. Pasteur) 125C: 373.

35. Hunkapiller, M. W. and L. E. Hood (1978) Biochemistry 17: 2124.

36. Hunkapiller, M. W. and L. E. Hood (1980) Science 207: 523.

CHAPTER 1

This paper was published in Nature.

Rearrangement of genetic information may produce immunoglobulin diversity

M. Weigert & L. Gatmaitan

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

E. Loh*, J. Schilling & L. Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125

The nearly complete amino-acid sequences of 22 closely related immunoglobulin κ variable ($V_κ$) regions from the inbred NZB mouse are presented. This group of $V_κ$ regions is encoded by at least six germline $V_κ$ genes. These data also suggest that the mouse κ gene is divided into three segments termed V or variable (residues 1 to 98 or 99), J or joining (residues 99 or 100 to 112) and C or constant (residues 113–219). Tonegawa et al. *have recently described a similar J segment for mouse λ chains. Inbred mice contain multiple $V_κ$ and $J_κ$ gene segments. Therefore, different combinations of V and J gene segments may be joined at the DNA level during the differentiation of individual lymphocytes to contribute to antibody diversity.*

RECENT structural studies on antibody molecules, mRNAs and genes have contributed substantially to our understanding of antibody diversity. Antibody polypeptides are divided into two regions, an N-terminal variable (V) region and a C-terminal constant (C) region. The organisation of antibody genes seems unusual in two respects: (1) the V and C regions are encoded by separate genes in the germline of each organism[1,2] and (2) these genes are rearranged with respect to one another during the differentiation of the antibody-producing cell[3,4]. Variable regions, which encode the antigen-binding sites, are the focus for three theories of antibody diversity. The germline theory contends that antibody variability is encoded by large numbers of germline V genes[5]. The somatic mutation theory postulates that diversity arises by somatic variation of relatively few germline genes[6]. A third model suggests that the rearrangement of antibody gene segments during differentiation may also generate antibody variability[7–9]. The studies described here on mouse κ chains stress the importance of multiple germline V genes and suggest that antibody diversity may be increased by the combinatorial joining of multiple V and J gene segments.

Diversity in mouse light chains

Myeloma tumours artificially induced in two inbred strains of mice, BALB/c and NZB, have provided homogeneous immunoglobulins, mRNAs and chromosomal DNA for the analysis of the diversity patterns of V regions in the two light-chain immunoglobulin families, λ and κ[10]. The overall pattern of V-region diversity in mouse light chains shows that most of the variability resides in three hypervariable regions. These fold to constitute the light-chain contribution to the walls of the antigen-binding crevice. The remainder of the V region provides a scaffold for the binding site and is termed the framework region[7,11–13]. The extent and nature of diversity patterns in $V_λ$ and $V_κ$ are quite different.

The sequences of mouse λ chains suggest that somatic mutation may have an important role in generating antibody diversity. Amino acid sequence analyses have established that 12 of 18 $V_λ$ regions studied are identical and that the remaining six differ by one to three amino acid substitutions, all of which occur in one of three hypervariable regions[15]. It has been proposed that the 12 identical $V_λ$ regions ($V_{λ0}$) are encoded by a single germline gene and that the variants arise by somatic mutations[6]. This interpretation is supported by nucleic acid hybridisation studies which suggest that $V_λ$ regions are encoded by one or a few germline genes[16,17]. In addition, the DNA sequence of the putative germline $V_λ$ gene, $V_{λ0}$, has been determined for a $V_λ$ gene isolated from presumably undifferentiated mouse embryo DNA and a variant sequence has been determined on another $V_λ$ gene isolated from myeloma DNA[18]. (A 19th λ chain, MOPC 315, seems to differ by about 10% of its V sequence from the other 18 λ chains. This V region is probably encoded by a second $V_λ$ gene.) Thus, most mouse λ chains exhibit a simple pattern of diversity at the protein and nucleic acid levels that is consistent with one germline V gene and variants arising by somatic mutations.

Variability in mouse $V_κ$ regions is extensive in both hypervariable and framework regions[11,19]. A relatedness tree for the N-terminal 23 residues, a portion of the framework region, of 61 BALB/c and NZB κ chains depicts 55 different sequences (Fig. 1). We will refer to a set of $V_κ$ regions which are very similar over their N-terminal 23 residues as a group. (Potter has suggested that each set of mouse $V_λ$ regions differing by three or fewer residues at their N terminus be given number (that is, group) designations[20]. By his criterion, 64 BALB/c $V_κ$ sequences fall into 26 groups and 31 NZB $V_κ$ sequences fall into 14 groups.) The complete V-region sequences from distinct groups may differ by up to 50% of their sequence. The genetic interpretation of V-region groups is that each is encoded by at least one distinct germline V gene because multiple, identical somatic mutations would be required to produce the immunoglobulin V regions of two or more groups from a single $V_κ$ gene[21,22]. This is described as the argument of parallel mutation.

Nucleic acid hybridisation data support the supposition that different groups are encoded by distinct V genes[23]. Hence, the minimum number of germline genes in a particular antibody family can be estimated by counting V-region groups. In the mouse κ family, the number of groups has been estimated to be about 50 for either BALB/c or NZB mice[15]. If the myeloma data analysed do not constitute a random selection of the normal $V_κ$ genes, the number of groups could be larger. In striking contrast, most λ chains fall into a single group.

The variability of mouse κ chains within a single group, $V_κ21$, does not resemble that of the predominant λ group (for example, with variability only in hypervariable regions). This has been shown for eight BALB/c κ chains in the $V_κ$ 21 group that have identical N-terminal sequences except for a single substitution. Three pairs of these $V_κ$ regions fall into closely related sets or subgroups, designated $V_κ21A$, $V_κ21B$ and $V_κ21C$, which are defined by multiple shared or subgroup-associated amino acid residues. The remaining two V regions differ from $V_κ$ regions of the three subgroups and from each other[24,25]. This survey suggests that the $V_κ21$ group in BALB/c mice is coded for by at

*Present address: Biochemistry Department, Stanford Medical School, Stanford, California 94305.

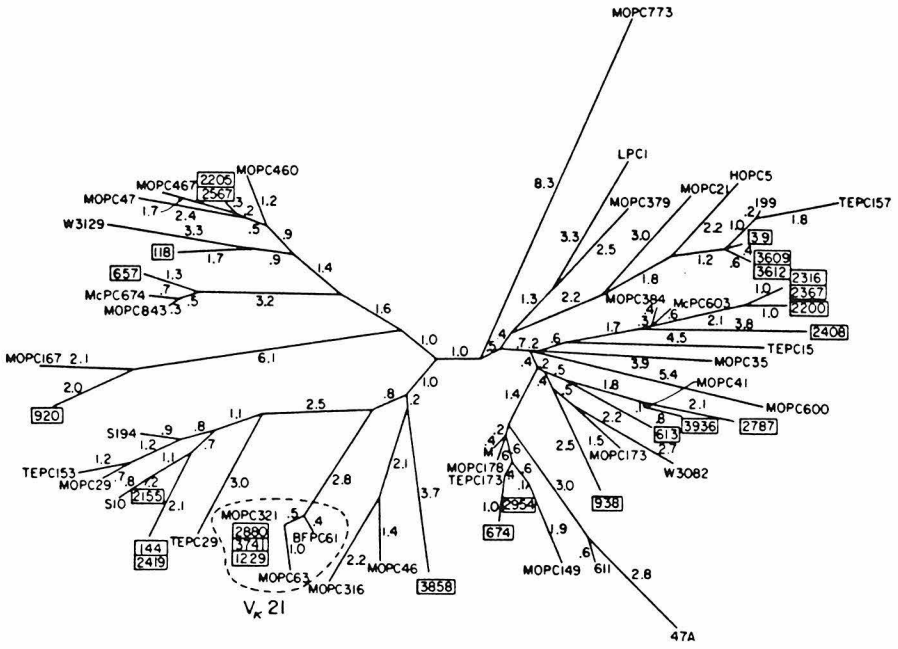


Fig. 1 A relatedness tree for N-terminal 23 residues of 55 different V_κ sequences of the mouse. These V_κ regions are derived from NZB (boxed) and BALB/c mice. The V_κ21 group is indicated by a dotted circle. The numbers on the lines indicate the number of nucleotide substitutions that separate two successive junction (node) points on the tree. Thus, the minimal extent to which individual κ chains differ from one another at the nucleotide level can be determined by summing the lengths of the lines between the two V_κ regions (adapted from ref. 19)

least three germline V genes by the argument of parallel mutation.

The object of this study is to analyse further the extent and pattern of diversity in the V_κ21 group of NZB mice. In a screen of the N-terminal sequences of 31 NZB κ chains, three V_κ21 chains, PC3741, PC2880 and PC1229, were found[26]. Using the NZB κ chains, PC3741 and PC2880, or the BALB/c V_κ21 chains, T124 and M70, antisera were made which specifically identify V_κ21 chains and distinguish a set reacting with antisera to PC2880 or M70 from a second set reacting with antisera to PC3741 or T124 (Fig. 2). Of 700 NZB myeloma immunoglobulins screened with these antisera, 5.7% fell into the first set and 3.1% into the second. In addition, the normal light chains from the sera of several inbred strains of mice crossreact with these antisera to levels of 7.2% and 1·4%, respectively.[27] Thus, the V_κ21 group constitutes about 8–9% of normal serum and myeloma light chains. This group is an ideal model system for studying the nature of diversity within the κ family of the mouse because many members of the V_κ21 group can be readily identified by rapid serological assays. Furthermore, the high frequency of these κ chains in normal sera suggests that the V_κ21 light chains are used not only in myeloma antibodies, but also in normal antibodies. In this article we report on the V-region sequences of 22 NZB κ chains belonging to the V_κ21 group. These data suggest that the V_κ21 group is encoded by multiple germline genes and that the V region is encoded by two separate DNA segments, V and J.

V_κ21 regions fall into six subgroups

Twenty-two NZB V_κ21 regions are compared. A summary of the methodology used to sequence these V regions is given in the legend to Fig. 2 and the complete data will be presented in detail elsewhere[28–30]. Of these V_κ21 regions, 18 fall into six subgroups, denoted by V_κ21A, B, C, D, E and F, in which two or more closely related V_κ regions share at least three subgroup-associated residues. Subgroup-associated residues distinguish the members of a particular subgroup from members of most of the other subgroups (boxed residues in Fig. 2); 29 subgroup-associated residues occur in positions 1–98. Variable regions from different subgroups differ from one another by 3 to 20 residues over their N-terminal 98 residues (Fig. 3). The reason for limiting our subgroup analysis to the N-terminal 98 residues will become apparent later. Again, even closely related subgroups (such as V_κ21E and V_κ21F) seem to be encoded by separate germline V genes by the argument of parallel mutation, namely, that it is unlikely that somatic mutation could generate in two or more independent lymphocyte lineages the three identical substitutions that separate V_κ regions of these two subgroups. Thus, at least one distinct germline V gene seems to encode each of the six subgroups (Fig. 3).

The minimal number of six V_κ21 germline genes deduced from V region amino acid sequences is larger than the one to three V_κ21 deduced from nucleic acid hybridisation experiments[31]. However, saturation-hybridisation studies using DNA probes derived from several V_κ21 mRNAs suggest that four to six germline genes encode the V_κ21 regions (see accompanying paper[32]). This number is consistent with the number of V genes we have identified so far by subgroup analysis.

It is possible that additional V_κ21 subgroups will be found. Four of the NZB V_κ21 regions (that is, PC2154, PC2413, PC7461 and PC2960) do not fall into any of the six subgroups and are indicated as distinct branches on the relatedness tree (Fig. 3). These V regions cannot yet be designated separate subgroups because a similar reasoning is needed to invoke the argument that separate germline genes for each subgroup avoid multiple identical somatic mutations (see above). If new V_κ21 sequences are found which pair with these examples, additional V_κ21 subgroups would be defined. In addition, currently defined subgroups may split into two or more new subgroups. For example, PC2485 was a member of the V_κ21E subgroup until the sequence of PC4039 was determined. These two identical sequences then became subgroup V_κ21F, sharing three distinct subgroup-associated residues (Fig. 2).

If additional V_κ21 subgroups are defined, then two possibilities must be considered. The saturation-hybridisation experiments have underestimated the number of V_κ21 genes. Alternatively, a somatic mechanism may exist which permits a single germline V gene to generate two or more subgroups of V_κ21 regions.

Genetic rearrangement during differentiation

The V_κ21 regions seem to have two distinct segments which we have designated the V segment (residues 1 to 98 or 99) and the J

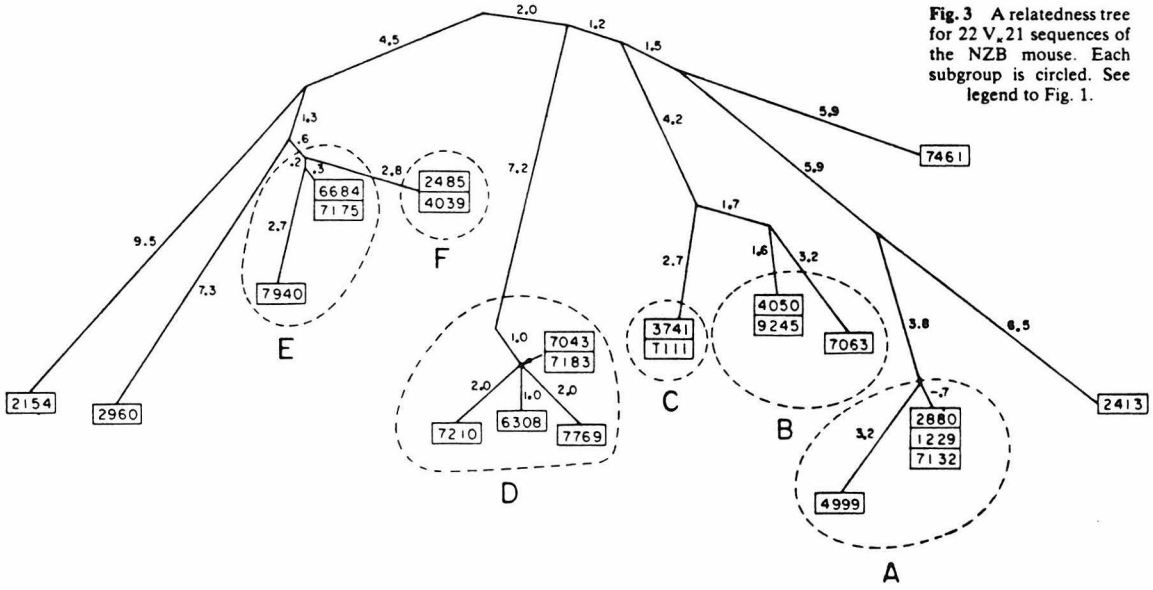


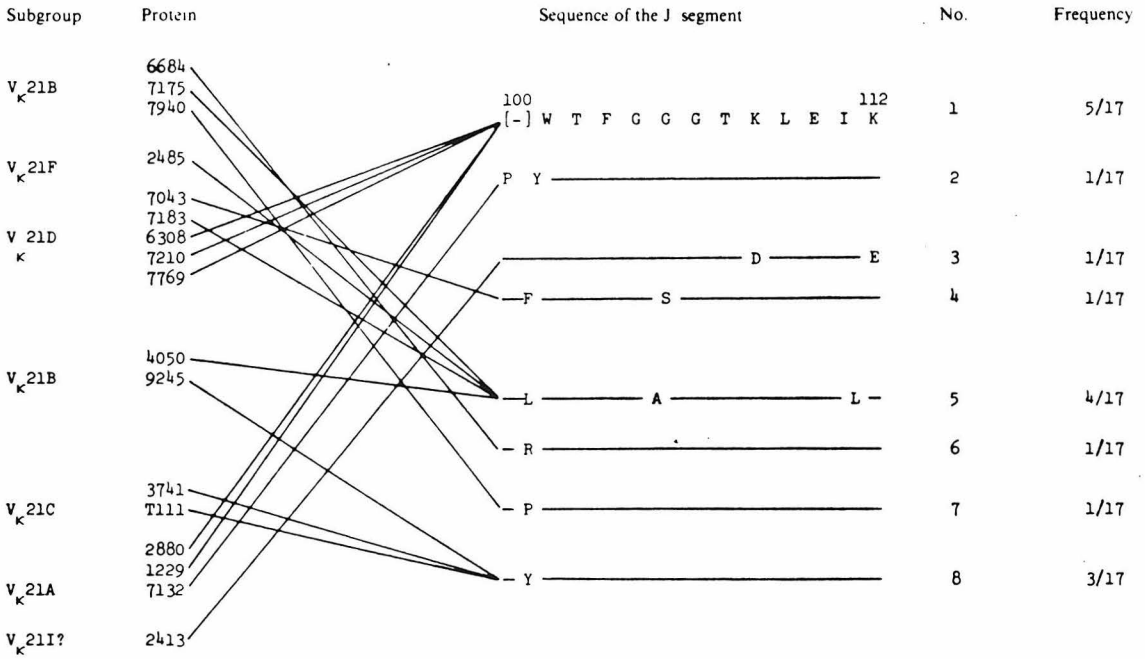
Fig. 3 A relatedness tree for 22 $V_{\kappa}21$ sequences of the NZB mouse. Each subgroup is circled. See legend to Fig. 1.

single J segment may be joined in a combinatorial fashion to each V segment to provide another stage in the amplification of antibody diversity. We describe this process as combinatorial joining. Indeed, position 101 is among the most variable in the $V_{\kappa}21$ regions examined (Fig. 2). This diversity might include sequence gaps (insertions or deletions) as well as nucleotide substitutions. The sequence gaps would result from J regions of different length (for example, proline at position 100 in PC7132).

The precise size of the J segment of $V_{\kappa}21$ genes must wait the DNA sequence analysis of $V_{\kappa}21$ genes from differentiated and undifferentiated DNA. If J does not include part of the third hypervariable region, then combinatorial joining, as such, would not generate diversity in the combining site. In this case, invariant associations such as the leucine at 101 with alanine at 105 and leucine at 111 as well as the pattern of substitutions at position 101 require an alternative explanation.

The J segment may also mediate the joining of V, J and C gene information. DNA rearrangements and mRNA splicing[36–38] must occur before the V, J and C gene information is joined into a single cytoplasmic mRNA (Fig. 5). The joining of V and J regions seems to occur at the DNA level[18]. The intervening

Fig. 4 A diagram illustrating the independent association of the V segments (residues 1–99) and the J segments (residues 100–112).



segment (residues 99–112). With this nomenclature, we can distinguish the V segment from the classical V region (residues 1–112).

The N- and C-terminal boundaries of the V and J segments are defined as follows. The J segment seems to have started by position 101 because of the invariant association of a leucine at this position with an alanine and leucine at positions 105 and 111, respectively, in the J segment of six different $V_\kappa 21$ regions (Fig. 2). This repeated invariant association of three residues suggests that positions 101–111 behave as a discrete genetic unit. The last subgroup-associated residue of the V segment is at position 98. Accordingly, the N-terminal boundary of the J segment may start somewhere between residues 99 and 101. Position 100 is an amino acid insertion in one κ chain (PC7132) and we have arbitrarily chosen the J segment to include this identical V segments may be associated with different J segments (for example, PC6684 and PC7175, PC7043 and PC7183, and PC4050 and PC9245). The implication is that the V and J segments associate freely with one another and that particular V segments are not generally associated with certain J segments (Fig. 4). Indeed, some of these same J segments are associated with non-$V_\kappa 21$ V segments of widely differing sequence[11]. Thus, a particular J segment of the κ family may be capable of associating with any V segment of the κ family.

Nucleic acid studies on mouse λ and κ genes also suggest that there are V and J segments. Two mouse V_λ genes isolated from embryonic DNA terminate their V genes at a position homologous to residue 101 (refs 2, 18). In addition, a J_λ segment consisting of approximately 36 nucleotides and separated from the V_λ segment by intervening DNA sequences has been

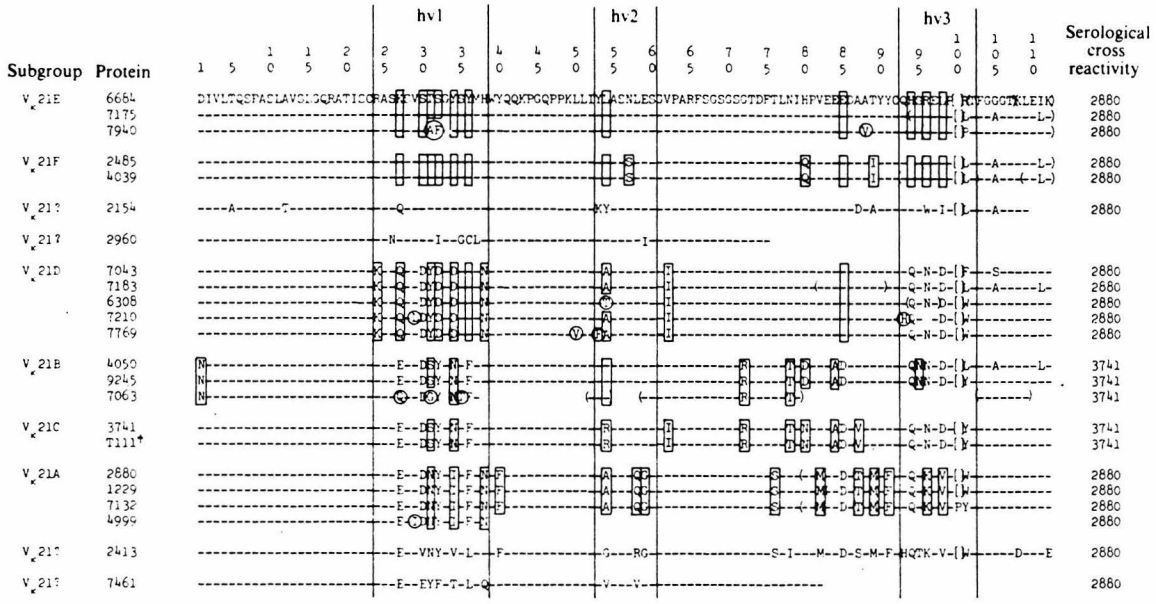


Fig. 2 The amino-acid sequences of V regions from the mouse $V_\kappa 21$ group. The sequences are divided into subgroups which are designated $V_\kappa 21A$–$V_\kappa 21F$. Additional potential subgroups are indicated with question marks. The T111 sequence is a BALB/c protein[25], identical to PC3741, which permits the definition of subgroup $V_\kappa 21C$. The subgroup-associated residues unique to three or fewer subgroups are boxed. The intra-subset substitutions are circled. Amino-acid compositional data are indicated by brackets. The three hypervariable regions are indicated as hv1, hv2 and hv3. The J segment extends along residues 100–112. The serological crossreactivity of individual proteins to antisera directed against two screening κ chains is given on the right. The following general strategy was used for sequencing these V regions. The intact light chain was analysed on the automatic sequenator for 40 residues. The procedures for automatic sequence analysis are given in ref. 28 where the data for a 78-residue run on PC4050 are presented. Methionine fragments were prepared; all V_κ regions but one have a methionine at position 37. Methionine fragments starting at position 38 sequenced for 40–65 residues. Only the $V_\kappa 21A$ subset has additional methionine residues and these fragments were isolated and sequenced. In certain cases mild acid cleavage was used to split aspartic acid–proline bonds. Arginine fragments extending over positions 65–113 were prepared for certain κ chains and sequenced for 35–40 residues. Thermolysin, tryptic and chymotryptic peptides were isolated from certain of these methionine or arginine fragments and compositions were determined. In some cases these peptides were sequenced in the presence of polybrene. A detailed analysis of these data will be presented elsewhere[28].

position, suggesting that the extra residue comes from a J segment one residue longer than the others. The C-terminal boundary of the J segment occurs at position 112, the C-terminal-most variable residue in κ chains (Fig. 2. PC2413). Moreover, as the C region seems to be encoded by one or a few C genes[17,32–34], the variability in the J region is too great to be accounted for by multiple C_κ genes beginning at position 100.

The V and J segments associate independently with one another (Fig. 4). In the 18 $V_\kappa 21$ regions with relatively complete sequence data between residues 100 and 112, there are 16 different V segments and eight different J segments (Fig. 2). A particular J segment may be associated with V_κ segments from four different subgroups (sequence 5 in Fig. 4). Conversely, described[4]. Thus, the mouse λ-gene family also seems to be composed of V, J and C segments. Several mouse V_κ genes isolated from myeloma DNA also terminate at approximately position 101 (ref. 35). These DNA studies taken together with our $V_\kappa 21$ sequence data suggest that the κ-gene gamily has two important organisational features (Fig. 5): (1) The κ chain is encoded by three noncontinuous gene segments, V, J and C; (2) multiple V-gene and J-gene segments are present in the κ-gene family of each mouse.

The most attractive function for the J segment is that it may be involved in generating antibody diversity. Residues 100 and 101 constitute a portion of the third hypervariable region[14]. Accordingly, if the J segment occurs at positions 100–112, a

DNA sequence between the J and C regions could either lack other J regions (Fig. 5*a*) or contain them, in which case they would be eliminated by RNA processing (Fig. 5*b*).

Intra-subgroup variability

The analysis of the intra-subgroup diversity in mouse λ chains presented a simple mutational pattern consistent with somatic mutation[15]. Likewise, if each of the six $V_\kappa 21$ subgroups is encoded by a single germline V gene, the nature and patterns of substitutions within each subgroup should reflect the mechanism of somatic mutation. The 12 intra-subgroup substitutions within the $V_\kappa 21$ regions are indicated by circles in Fig. 2. Several features distinguish these V_κ substitutions from their V_λ counterparts discussed earlier. (1) Two of 12 substitutions are found outside the hypervariable regions. (2) Four of six variant sequences have two or more substitutions. If these differences arise by somatic mutation, the mutational mechanism operates outside as well as within hypervariable segments and shows a tendency to generate multiple substitutions in individual variant sequences. Moreover, somatic mutation followed presumably by antigen selection seems to favour diversity at certain sites in the hypervariable regions.

Again, this analysis of intra-subgroup variability must be qualified by the possibility that one variant differing by two or more residues from the prototype sequence will become a distinct subgroup when a second identical sequence is determined. Thus, perhaps the reason the intra-subgroup variability in the $V_\kappa 21$ chains and the λ chains differ is that in the former some of the variants are encoded by additional germline V genes.

Sources of antibody diversity

The analysis of the V_κ 21 group of mouse V_κ regions offers certain insights into each of four basic mechanisms for producing antibody diversity.

Multiple germline V genes: According to a statistical analysis of N-terminal data, there are about 50 κ groups[15,39]. This number may be an underestimate because it is based on the sequences of myeloma proteins which may represent only a subset of the total V genes. Our studies suggest the $V_\kappa 21$ group is encoded by at least six V genes. In addition, studies at the DNA level on three V_κ groups suggest that four to eight germline V genes are present in each group[23,32]. If we assume that seven germline V genes are present in each group defined by N-terminal analysis, then approximately 350 germline genes encode the V_κ family. Less is known about the diversity in the V_H gene family, but a similar estimate of at least 200 V_H genes seems reasonable.[39] In contrast, the mouse λ gene family is apparently encoded by one or two germline genes.

Combinatorial association: That most light chains may combine with most heavy chains is suggested by several types of observations[19,40,41]. This non-genetic mechanism for generating antibody diversity is termed combinatorial association. In principle, 350 V_L and 200 V_H regions could, through combinatorial association, generate 7×10^4 different types of antibody molecules. If the mammalian antibody repertoire is greater than 10^6 (see ref. 42), and if we assume that there are only 350 V_L and 200 V_H germline genes, then more than 90% of the antibody repertoire must be generated by somatic mechanisms.

Somatic mutation: All the intra-subgroup substitutions noted in the $V_\kappa 21$ regions are single-base substitutions, predominantly in the hypervariable regions. Obviously, some or all of this variation could arise by spontaneous somatic mutation or by a hypermutational mechanism.

Combinatorial joining: One potential source of somatic diversification is the combinatorial joining of multiple J segments and V segments (Fig. 5). Thirteen different J segments have already been identified in mouse κ chains. For example, if 10 different J segments may be joined to any of 350 different V segments, then 3,500 different V regions may be generated. If combinatorial joining of a similar magnitude occurs in light and heavy variable regions, 7×10^6 antibody molecules could be generated by this process.

Combinatorial joining could generate identical variants among mice of the same inbred strain. In contrast, most somatic theories have assumed that the generation of diversity is a random process occurring by mutation, recombination or gene conversion. A reproducible somatic process could explain the

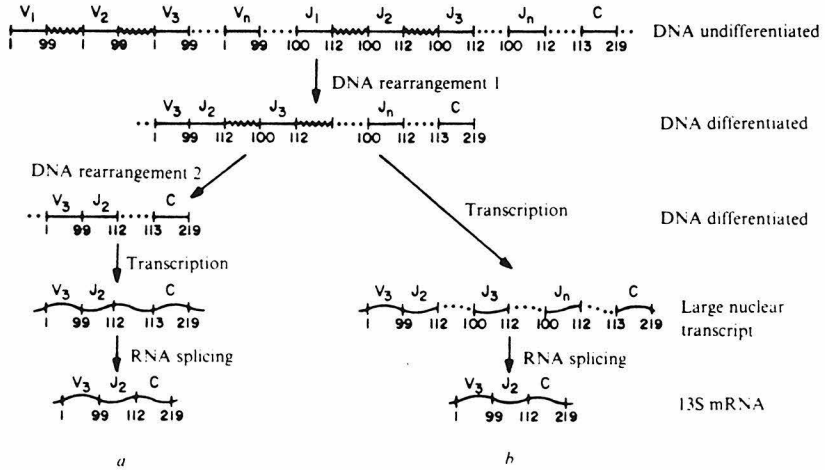


Fig. 5 Hypothetical models for the DNA segments of mouse κ genes. These gene segments are rearranged at the DNA level, presumably as the lymphocyte differentiates. As with λ genes, DNA rearrangement is depicted as joining the V and J segments[4]. There are at least two pathways for joining the J and C segments at the nucleotide level. *a*, A second DNA rearrangement moves the J_2 segment closer to the C segment, possibly deleting the intervening J sequences. A recent report suggests that two or more DNA rearrangement events can occur in lymphocytes during the switch in synthesis from one to a different immunoglobulin class[48]. Final joining occurs by mRNA splicing. *b*, A large nuclear RNA transcript complete with the intervening J segments is processed by mRNA splicing enzymes.

patterns of diversity seen in certain immune responses. For example, the responses to α1,3-glucosyl linkages of BALB/c mice[43] and to the 4-hydroxy-3-nitrophenylacetyl hapten (NP) of C57BL/6 mice[44] are heterogeneous by isoelectric focusing, but the heterogeneity is limited to V_H and is reproducible in each case[45,46]. Perhaps this reproducible heterogeneity arises from the combinatorial joining of V_H–DEX or V_H–NP^b genes with a variety of different J_H segments. A similar possibility could explain the closely related heavy-chain sequences produced in response to the immunisation of guinea pigs against several haptens[47].

Other investigators have proposed combinatorial joining as a mechanism for antibody variability[7,8], but their models arose from the conviction that certain hypervariable regions may be associated with many different framework regions. For example, based on a computer analysis of all available V-region sequences, Kabat *et al.* have suggested the mini-gene model, which proposes that each of the three hypervariable regions and the four intervening framework regions are encoded by mini-genes which are joined during differentiation[9]. The correlation of specific hypervariable and framework segments within certain V_κ21 subgroups (Fig. 2) argue against this general model. Our data suggest that Kabat's model is correct in principle, but that the combinatorial process occurs only within the third hypervariable region. The identification of a J segment in λ chains raises the possibility that combinatorial joining of V and J segments will be used by all three antibody families, λ, κ and H.

Conclusion

Antibody variability may be generated by a variety of mechanisms, some genetic (multiple V genes, combinatorial joining and somatic mutation) and others non-genetic (combinatorial association). Combinatorial joining is both a somatic and a germline mechanism in that it presumably rearranges germline information during somatic differentiation. The critical question that remains for the future is what are the relative contributions of these four basic mechanisms for generating diversity to the functional repertoire of antibody molecules? The answer to this intriguing question will obviously require combined chemical, biological and genetical approaches.

This research was supported by a grant from the NSF grants from the NIH and an appropriation from the Commonwealth of Pennsylvania. We thank Drs Michael Potter and David McKean who developed the serological methods essential for identifying V_κ21 light chains.

Received 22 August, accepted 31 October 1978.

1. Tonegawa, S., Hozumi, N., Mathyssens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877–889 (1976).
2. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
3. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976).
4. Brach, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* (in the press).
5. Hood, L., Campbell, J. H. & Elgin, S. C. R. *A. Rev. Genet.* **9**, 305–353 (1975).
6. Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Nature* **228**, 1045–1047 (1970).
7. Wu, T. & Kabat, E. *J. exp. Med.* **132**, 211–250 (1970).
8. Capra, J. D. & Kindt, T. J. *Immunogenetics* **1**, 417–427 (1975).
9. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429–2433(1978).
10. Potter, M. *Physiol. Rev.* **52**, 631–719 (1972).
11. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Variable Regions of Immunoglobulin Chains, Tabulations and Analyses of Amino Acid Sequences* (Bolt, Bearanek and Newman, Massachusetts, 1976).
12. Amzel, L., Polyjak, R., Saul, F., Varga, J. & Richard, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1427–1430 (1974).
13. Segal, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4298–4302 (1974).
14. Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. *Biochemistry* **12**, 5400–5416 (1973).
15. Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 837–846 (1976).
16. Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 203–207 (1976).
17. Leder, P., Honjo, T., Swan, D., Peckman, S., Nau, M. & Norman, B. in *Molecular Approaches to Immunology, Miami Winter Symposia*, Vol. 9 (eds Smith, E. E. & Ribbons, D. W.) 173–188 (Academic, New York, 1975).
18. Bernard, O., Hosumi, N. & Tonegawa, S. (submitted).
19. Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 817–836 (1976).
20. Potter, M. *Adv. Immun.* **25**, 141–212 (1978).
21. Hood, L., Gray, W., Sanders, E. & Dreyer W. *Cold. Spring Harb. Symp. quant. Biol.* **32**, 133–146 (1967).
22. Milstein, C. *Nature* **216**, 330–332 (1967).
23. Rabbitts, T. H. *Immun. Rev.* **36**, 29–50 (1977).
24. McKean, D., Potter, M. & Hood, L. *Biochemistry* **12**, 749–759, 760–771 (1973).
25. McKean, D. J., Bell, M. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
26. Loh, E. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
27. Julius, M., Potter, J., McKean, D. & Weigert, M. (submitted).
28. Hunkapiller, M. & Hood, L. *Biochemistry* **17**, 2124–2133 (1978).
29. Loh, E., Schilling, J., Weigert, M. & Hood, L. (submitted).
30. Schilling, J., Loh, E., Weigert, M. & Hood, L. (submitted).
31. Tonegawa, S., Steinberg, C., Dube, S. & Bernardini, A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4027–4031 (1974).
32. Valbuena, O., Marcu, K. B., Weigert, M. & Perry, R. P. *Nature*, **276**, 780–784 (1978).
33. Leder, P., Honjo, T., Seidman, J. & Swan, D. *Cold Spring Harb. Symp. quant. Biol.* **41**, 855–862 (1976).
34. Rabbitts, T. & Milstein, C. *Cont. Topics molec. Immun.* **6**, 117–143 (1977).
35. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* (in the press).
36. Gilmore-Herbert, M. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 342–345 (1978).
37. Schibler, U., Marcu, K. & Perry, P. (submitted).
38. Rabbitts, T. H. & Forster, A. *Cell* **13**, 319–327 (1978).
39. Cohn M. *et al.* in *The Immune System, Genes, Receptors, Signals* (eds. Sercarz, E. E., Williamson, A. R. & Fox, C. F. 89–118 (Academic, New York, 1974).
40. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerly, R. P. & Saul, F. *Immunogenetics* **2**, 393–394 (1975).
41. Milstein, C. *et al. Cold Spring Harb. Symp. Quant. Biol.* **41**, 793–803 (1976).
42. Williamson, A. R. *A. Rev. Biochem.* **45**, 467–500 (1976).
43. Blomberg, B., Geckeler, W. & Weigert, M. *Science* **177**, 178–180 (1972).
44. Imanishi, T. & Mäkelä, O. *J. Exp. Med.* **140**, 1498–1510 (1974).
45. McMichael, A. J., Phillips, J. M., Williamson, A. R., Imanishi, T. & Mäkelä, O., *Immunogenetics* **2**, 161–173 (1975).
46. Hansburg, D., Briles, D. E. & Davie, J. M. *J. Immun.* **119**, 1406–1412 (1977).
47. Cebra, J. J., Koo, T. H. & Ray, A. *Science* **186**, 263–265 (1974).
48. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2140–2144 (1978).

CHAPTER 2

This paper was published in Nature.

The joining of V and J gene segments creates antibody diversity

Martin Weigert, Robert Perry & Dawn Kelley

The Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Tim Hunkapiller, James Schilling & Leroy Hood

California Institute of Technology, Division of Biology, Pasadena, California 91125

The variable regions of mouse kappa (κ) chains are coded for by multiple variable (V) gene segments and multiple joining (J) gene segments. The V_κ gene segments code for residues 1 to 95; the J_κ gene segments code for residues 96 to 108 (refs 1–3). This gene organisation is similar to that encoding the V_λ regions[4]. Diversity in V_κ regions arises from several sources: (1) there are multiple germ-line V_κ gene segments and J_κ gene segments; (2) combinatorial joining of V_κ gene segments with different germ-line J_κ gene segments; and possibly, (3) somatic point mutation, as postulated for V_λ gene segments[5]. Also, from a comparison of the number of germ-line J_κ gene segments and amino acid sequences, it has been suggested that J_κ region sequences may be determined by the way V_κ and J_κ gene segments are joined[2,3]. This report supports this model by directly associating various J_κ sequences with given J_κ gene segments.

The number of germ-line J_κ gene segments has been estimated in two ways. A region of DNA containing J_κ gene segments has been characterised by DNA sequence analysis and was found to include five J_κ (or J_κ-like) gene segments, only four of which have been found to be expressed as J segments[2,3]. Additional J_κ gene segments have not been found, although their presence cannot be excluded rigorously. Alternatively, the number of J_κ gene segments may be determined by counting the number of distinctive RNA precursor patterns associated with different κ chain mRNAs[6]. Since the κ-mRNA precursors contain the intervening sequence between the J_κ and C_κ segments, they have a characteristic size depending on which J segment is used, the largest precursor corresponding to the J_κ segment most distal to the C_κ gene and the smallest to the most proximal J_κ segment. On this basis the RNA precursor patterns for 20 different myeloma tumours were classified into four different categories, corresponding to the four different J_κ gene segments that these chains use. The distances between the C_κ gene and the various J_κ gene segments deduced from the analysis of precursor patterns in NZB plasmacytomas agree with the distances determined by restriction mapping and DNA sequence analysis of BALB/c germ-line DNA. A previously suspected discrepancy[6] has been resolved by a more accurate calibration of the mobility data.

We have analysed κ chains of the V_κ 21 group from the NZB mouse strain. To date ten different J_κ amino acid sequences have been observed in twenty V_κ 21 examples (Fig. 1). Thus the

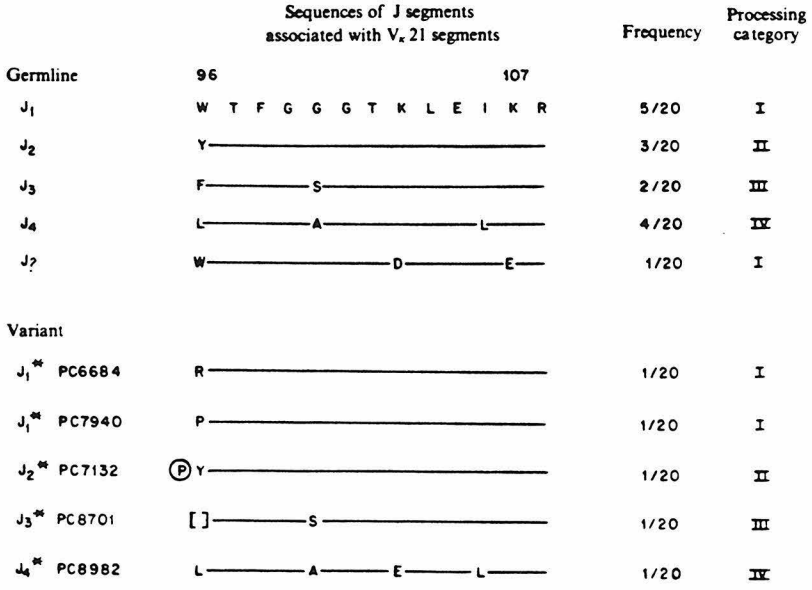


Fig. 1 J segments associated with $V_\kappa 21$ chains from NZB myeloma antibodies. J segments are numbered according to their processing category[6]. This is based on the relative sizes of the κ-mRNA precursors, as determined by gel-electrophoretic analysis of poly(A)$^+$ nuclear RNA. These precursors are approximately 5.3, 5.0, 4.4 and 4.1 kilobases for categories I, II, III and IV, respectively. The amino-terminal sequences (40 residues) of PC8701, PC8982 and PC10916 were determined on a modified Beckman sequenator as described previously[11] and were found to be representatives of the $V_\kappa 21$ subgroups, D, C and C, respectively. The J segment sequence of these κ chains were determined as follows; each κ chain was subjected to mild acid cleavage[12] of the aspartylproline between residues 94 and 95. The peptide mixture was sequenced without separation after the amino-terminal peptide was blocked by addition of fluorescamine[13]. The rest of the sequences are taken from ref. 1. The germ-line J segments are defined as described in the text. J_1, J_2, J_3 and J_4 are coded for by the BALB/c embryo J gene segments termed J_1, J_2, J_4 and J_5 in ref. 2 or those termed J_5, J_4, J_2 and J_1 in ref. 3. The PC2413 J segment may be coded for by a germ-line J gene (or variant of a J gene) unique to the NZB mouse.

number of J_κ sequences appears to be greater than the number of germ-line J_κ gene segments. There are three types of J_κ sequence. First, four distinct types (J_1, J_2, J_3 and J_4) correspond to those coded for by four of the germ-line J_κ gene segments in BALB/c DNA. Each of these J_κ sequences has been observed two or more times in the NZB $V_\kappa 21$ regions (Fig. 1). These J_κ sequences also have been observed in BALB/c κ chains[7] indicating that NZB and BALB/c germ-line J_κ gene segments code for the same J_κ regions. A second type of J_κ sequence differs by a single residue from those of the germ-line J_κ gene segments; these are termed variant J_κ sequences. There are five J_κ sequences of this type and each has been observed only once. There are two possible explanations for the variant J_κ sequences: (1) they arise from somatic diversification which occurs at the junctional region during the joining of V_κ and J_κ gene segments or (2) they are encoded by additional germ-line J_κ gene segments that are absent from or have not yet been detected in the BALB/c genome. The former model predicts that the precursors of the mRNAs coding for these variant J_κ segments would belong to a category typical of one of the four germ-line examples; the latter model predicts a unique mRNA precursor for each variant. A third type of J sequence which occurs in the exceptional PC2413, differs from the other germ-line J_κ gene segments by at least two nucleotide substitutions.

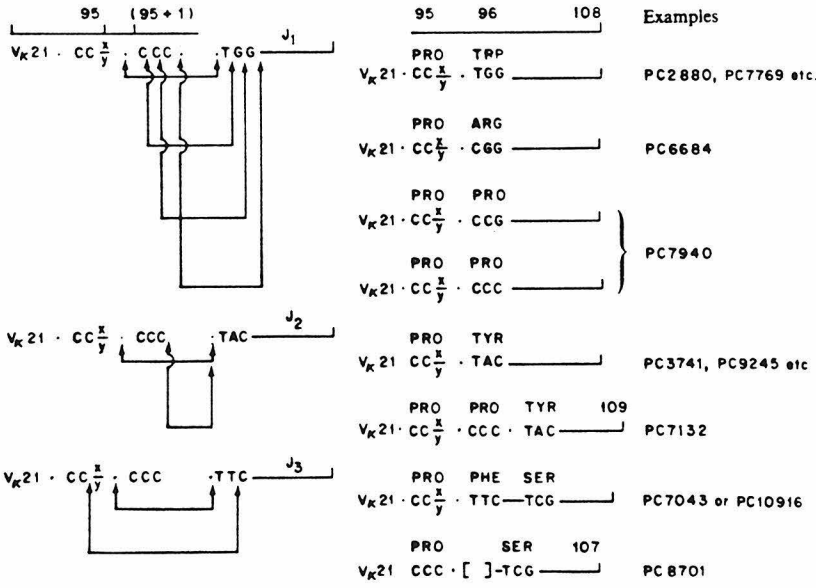


Fig. 2 Fusion of $V_\kappa 21$ genes and J gene segments. The fusion can occur to join V and J gene segments exactly to form a contiguous V–J gene. Such genes will have germ-line J region sequences as in the case of PC2880, PC3741 or PC7043 and other examples[1]. Fusion to J genes might include nucleotides in the intervening sequence (95 + 1). If this codon is CCC for all $V_\kappa 21$, then the variant sequences, PC6684, PC7940 and PC7132 can be explained by fusion to a J gene (PC7132), within a J gene (PC6684 and PC7940) or within a V gene (PC8701).

Differences of this extent raise the issue of whether the PC2413 J_κ sequence might be encoded by a fifth germ-line J_κ gene segment in the NZB mouse.

An analysis of the RNA processing patterns for the tumours producing κ chains with variant J_κ sequences indicates that each pattern belongs to one of the four categories characteristic of germ-line J_κ gene segments (Fig. 1). Thus it appears that these variant J sequences arise during the process of V_κ–J_κ joining.

Knowing the size of the mRNA precursor for a J_κ variant allows us to infer the kind of somatic mutation that could have produced the variant sequence. For example, both PC6684 and PC7940 mRNAs exhibit the J_1 type prescursor. Their J segments differ from the germ-line J_1 segment at position 96, presumably due to modifications of the tryptophan codon 96. PC7132 arose from the insertion of a proline codon before the J_2 gene segment. PC8701 resulted from a deletion of the phenylalanine codon at the beginning of J_3 gene segment. In addition to this junctional diversity, random point mutations either in the germ line or during somatic differentiation may occasionally occur (for example, PC8982).

The κ-mRNA precursor in PC2413 appears to belong to the category characteristic of the germ-line J_1 segment. Thus, if the J sequence of this tumour is actually encoded by a unique NZB germ-line J_κ segment, thus segment would have to be located very close to the J_1 segment, or alternatively, be associated with a second, as yet unidentified, C_κ gene.

The nature of most J_κ variants suggests that they may be due to variations arising from the V–J joining process because the substitutions, mutations, insertions or deletions generally occur at the beginning of J_κ segments. These variations may be explained by two base substitutions (for example, tryptophan to proline) as well as the insertion or deletion of codons. Thus this variation does not appear to arise merely by single-base somatic mutation. Leder and Tonegawa and their coworkers have recently observed identical inverted repeat sequences at the 3′ end of the V gene segment and at the 5′ end of the J gene segment. They have hypothesised that these inverted repeats may form paired stem structures that juxtapose the V and J gene segments and that recombination may occur across the base of this stem at slightly different points to generate V–J junctional diversity (refs 2 and 3 and Fig. 2). This site-specific recombinational model appears to explain the V–J junctional diversity that we have noted in the $V_\kappa 21$ chains.

By whatever mechanism V gene segments are translocated to J gene segments, a continuous DNA sequence coding for the entire V region must be generated[4]. Since all mouse κ chains studied to date have proline at residue 95 (CC^{a}_{γ}), a V_κ gene segment translocation to the J_1 gene segment (tryptophan (TGG) at residue 96) will result in the contiguous DNA sequence, CC^{a}_{γ} TGG at the V–J joining site (Fig. 2). Such a translocation will create a V_κ gene segment that is associated with the J_1 germ-line gene segment. This type of translocation explains five out of the twenty $V_\kappa 21$ regions observed (Fig. 1). Likewise, similar translocations between $V_\kappa 21$ gene segments and the J_2, J_3 and J_4 gene segments can explain the nine additional examples of germ-line J segments (Fig. 1).

The variant J segments may arise by recombination within rather than between the V gene segment codons for positions 95 or 95 + 1 and their counterparts in the J gene segments (Fig. 2). Codon 95 + 1 is CCC in several V_κ gene segments[2,8]. If the (95 + 1) codon is CCC for all $V_\kappa 21$ gene segments, the type of variation that may occur at the V–J joining site is limited. Amino acid substitutions at residue 96 for a $V_\kappa 21$–J_1 translocation can only be arginine (CGG) or proline (CCG) (Fig. 2). An amino acid insertion would result in an extra proline codon before the J gene segment. In addition, deletions of residue 96 could occur. As shown in Fig. 2, the variants observed so far fit the predictions of this model. If there is variation at the (95 + 1) codon, a greater variety of substitutions could occur at residue 96[3]. Obviously similar variation could occur in the (96 − 1) codon of the J gene segment. Additional analyses of $V_\kappa 21$ chains are under way to define more completely the amino acid sequence variability at the V–J junctional site.

In our survey of $V_\kappa 21$ chains, the frequency of κ chains expressing variant J segments relative to those expressing germline J segments is high (5 out of 20). This frequency could be high because the N-terminus of the J_κ segment is important for combining-site size and shape[9] and antigen selection of specific antibodies amplifies mutations that occur at a low frequency. (The uniformity of J_λ segments[10] may mean that variation in this region does not affect the complementarity of antibodies with λ chains.) Alternatively, this frequency may be a true reflection of the rate of variation by this mechanism.

This research was supported by a grant from the NSF, grants from the NIH and an appropriation from the Commonwealth of Pennsylvania.

Received 4 September; accepted 11 November 1979.

1. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276**, 785–790 (1978).
2. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
3. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454 (1979).
4. Brack, C., Hirama, M., Lenhard-Schuller & Tonegawa, S. *Cell* **15**, 1–14 (1978).
5. Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Nature* **228**, 1045–1047 (1970).
6. Perry, R. P., Kelley, D. E. & Schibler, U. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3678–3682 (1979).
7. Kabat, E. A., Wu, T. T. & Bilofsky, H. (eds) *Variable Regions of Immunoglobulin Chains, Tabulations and Analysis of Amino Acid Sequences* (Bolt, Baranek & Newman, Massachusetts, 1976).
8. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11–17 (1978).
9. Padlan, E. A., Davies, D. R., Rudikoff, S. & Potter, M. *Immunochemistry* **13**, 945–949 (1976).
10. Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol* **41**, 837–846 (1976).
11. Hunkapiller, M. & Hood, L. *Biochemistry* **17**, 2124–2133 (1978).
12. Paulsen, K., Fraser, K. J. & Haber, E. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2495–2499 (1972).
13. Udenfriend, S. *et al. Science* **178**, 871–872 (1972).

CHAPTER 3

This paper was published in The Journal of Immunology.

ANALYSIS OF THE DIVERSITY OF MURINE ANTIBODIES TO DEXTRAN B1355: N-TERMINAL AMINO ACID SEQUENCES OF HEAVY CHAINS FROM SERUM ANTIBODY[1]

J. SCHILLING, D. HANSBURG,[2] J. M. DAVIE, AND L. HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125 and the Departments of Microbiology and Immunology and of Pathology, Washington University School of Medicine, St. Louis, Missouri 63110

The N-terminal amino acid sequences of two γ and two μ chains from normally induced serum antibodies to dextran in BALB/c mice are presented. These heavy chains are derived from antibodies with three distinguishable idiotypes. These variable region (V_H) sequences are all identical as far as they have been analyzed (27 to 53 residues). The light chains from these antibodies are all of the λ type and are identical by isoelectric focusing analysis. Accordingly, the diversity of dextran antibodies appears to reside primarily in the heavy chains. The implications of these observations for antibody diversity are discussed.

The generation of antibody diversity is an intriguing biologic problem that has fascinated immunologists for the past 20 years. The murine immune response to B1355 dextran (Dex)[3] has several features that afford an excellent opportunity to gain insights into the molecular basis of antibody diversity in a normal immune response. First, three myeloma proteins have been found that bind dextran, M104 (IgM), J558 (IgA), and U102 (IgA). These myeloma proteins each exhibit unique idiotypic determinants (1). Moreover, the μ-chain of M104 has been completely sequenced (2) and the V_H region of J558 is nearly sequenced (3) (J. Schilling and C. Oken, unpublished data). These three myeloma immunoglobulins serve as useful structural models for antibodies binding dextran. Second, the immune response to B1355 dextran exhibits limited and reproducible antibody heterogeneity by two criteria—isoelectric focusing and idiotypic analysis (1, 4, 5). Isoelectric focusing (IEF) analyses demonstrate several spectrotypes of antibody, most of which are shared by individual immunized mice. The antibodies to dextran can be divided into five distinct molecular species by idiotypic analysis: 1) antibodies bearing IdI(M104) determinants [individual idiotypic determinant(s) characteristic of myeloma protein M104]; 2) antibodies bearing IdI(J558) determinants; 3) antibodies bearing IdI(U102) determinants; 4) antibodies bearing IdX determinants (idiotypic determinants shared by all three myeloma proteins) but not any of the three IdI determinants; and 5) antibodies not bearing any of the four defined idiotypic determinants (Id neg). Although the degree of heterogeneity in the fourth and fifth classes of antibodies (IdX and Id neg) is unknown, all of these analyses taken together suggest the dextran system is a relatively simple one whose individual components may be analyzed in molecular terms. Third, certain mouse strains give high responses upon immunization with dextran and others give low responses (6). Idiotypic analyses (probably employing the IdX specificity) demonstrate that the high-low response is a simple Mendelian trait that maps to the heavy chain locus of mouse (6–8). These observations suggest that a single gene or a closely linked cluster of genes controls the expression of this idiotypic determinant. Accordingly, the dextran system affords an opportunity to analyze antibody diversity in a system that appears to be regulated by a limited number of V genes. Finally, mice that give high responses to dextran appear to have λ but not κ light chains in their dextran antibodies (6, 7). The λ-chain of the mouse exhibits highly restricted heterogeneity (8, 9). Accordingly, most, if not all, of the diversity of the dextran system will likely be contributed by the heavy chans. This allows us to focus our attempts at diversity analysis on the heavy chains and to analyze the extent and nature of variability in the hypervariable and framework portions of these V_H regions.

In this paper we analyze the N-terminal amino acid sequences of heavy chains derived from dextran serum antibodies that exhibit several distinct idiotypes. The N-terminal sequences of these antibody heavy chains appear homogeneous. Moreover, the individual sequences are identical to one another and to those of the corresponding heavy chains from myeloma proteins binding dextran. These observations support the supposition that the immune responses to dextran in high responder mice are highly restricted in their molecular heterogeneity and that they afford an excellent system for analyzing antibody diversity.

MATERIALS AND METHODS

Immunization with B1355 dextran. Immunization with dextran fraction S from *Leuconostoc mesenteroides* NRRL B1355 has been described (4). Briefly, BALB/c mice were primed i.p. with soluble dextran B1355 in complete Freund's adjuvant, boosted at 3 to 4 weeks with soluble dextran in incomplete Freund's adjuvant and again at 8 weeks with *Escherichia coli* strain B. Bleedings were made at 10 and 15 days after the final immunization. Total anti-dextran antibody was quantitated in a modified Farr assay (4). Idiotypes were determined by using tube binding assays (1).

Received for publication January 22, 1979.

Accepted for publication April 4, 1979.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by United States Public Health Service Grants AI-10781 and AI-11635 and National Science Foundation Grant PCM76-09719.

[2] D. H. was supported by United States Public Health Service Fellowship AI-05344.

[3] Abbreviations used in this paper: Dex, 1355 dextran; IEF, isoelectric focusing; PTH, phenylthiohydantoin; Id, idiotype.

Preparation of antibody pools. Animals were selected for either high levels of both IdI(M104) and IdX or low levels of IdX; sera from five animals constituted the "high Id" pool, and sera from six animals were chosen for the "low Id" pool. Antibodies were precipitated with dextran B1355 and the IgM and IgG antibodies were separated by gel filtration on Ultrogel AcA22 (LKB, Stockholm, Sweden) in 3 M guanidine. Each antibody was completely reduced and alkylated and the light and heavy chains were separated over AcA34 (LKB).

N-terminal amino acids sequence analysis of purified heavy chains. The isolated μ- and γ-chains from the high Id pool were sequenced on a Beckman 890B spinning-cup sequenator. Phenylthiohydantoin (PTH) amino acids were identified by gas chromatography and high pressure liquid chromatography. μ- and γ-chains from antibodies from the low Id pool were sequenced on a modified Beckman sequenator and residues identified by high pressure liquid chromatography (3). Data for the N-terminal 54 residues of the J558 sequence have been published previously (3). Data for the M104 sequence will be published elsewhere (2).

IEF of light chains. IEF analysis of light chains was performed with the LKB Multiphor flat plate apparatus employing an Ampholine gradient from pH 3.5 to 10.

RESULTS AND DISCUSSION

The light chains of myeloma proteins and antibodies binding dextran appear to be identical by IEF criteria. Figure 1 presents an IEF analysis of the light chains from myeloma proteins and antibodies binding dextran. The myeloma proteins and antibodies are derived from BALB/c mice. The λ light chains of M104 and J558 are identical by sequence analysis (9). A comparison of the IEF patterns of these two light chains (Fig. 1) shows their major bands co-focus. In addition, several acidic bands are present in the J558 λ-chains that are missing in the M104 λ-chains. These extra acidic bands may result from deamidation. All of the major bands of the light chains from the antibody pools co-focus with the J558 λ-chain bands. These observations suggest that the light chains are very similar, if not identical, in structure because 40% of the random amino acid substitutions involve charge differences. Accordingly, even with just a few amino acid substitutions, one should see charge heterogeneity. Only a complete amino acid sequence analysis can determine if these λ-chains are, in fact, identical in sequence. Nevertheless, these data to allow us to conclude that the different spectrotypes of dextran antibody must be produced by charge variations in the heavy chains. Some of the individual idiotypes and the IdX and Id neg idiotypes differ in charge from one another. Accordingly, the diversity of the dextran system that has been observed to date residues primarily in the heavy chains.

The heavy chains from myeloma proteins binding dextran are very similar to one another in amino acid sequence, but they are quite distinct from the sequences of other myeloma heavy chains. In Figure 2 are presented the amino acid sequences of the heavy chains from dextran-binding myeloma proteins M104 and J558. These polypeptides are identical for their N-terminal 53 residues by N-terminal amino acid analysis. Tryptic peptides derived from the cyanogen bromide fragment of J558 extending from 35 to 81 are identical in amino acid composition to the corresponding tryptic peptides from M104 (J. Schilling and C. Oken, unpublished data). Moreover, N-terminal amino acid sequence analysis of the cyanogen bromide fragment starting at position 82 in the V_H region of J558

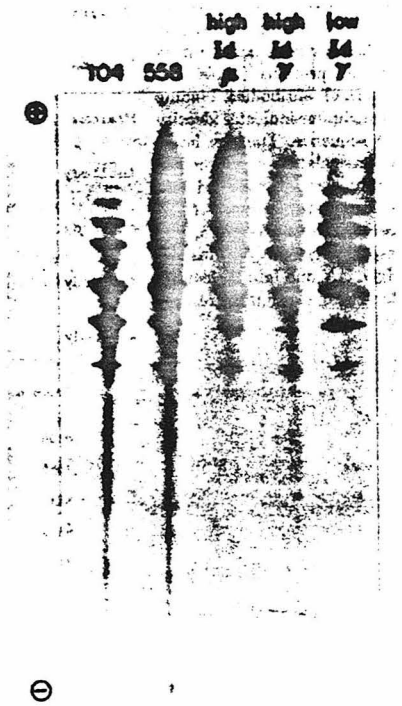


Figure 1. Isoelectric focusing gels of light chains derived from myeloma proteins and antibodies binding dextran. High Id μ and high Id γ denote λ-chains isolated from μ and γ fractions of the "high Id" dextran antibody pool; low Id γ designate λ-chains isolated from the γ fraction of the "low Id" dextran antibody pool. All light chains were completely reduced and alkylated with iodoacetamide before focusing. The gel was stained with Coomassie Brilliant Blue R250.

demonstrates that it is identical to that of M104 for residues 82–99. These observations suggest that the V_H regions of M104 and J558 are identical up to the third hypervariable region. Accordingly, the substitutions that confer idiotypic individuality on the V_H regions of these two myeloma proteins must lie in or beyond the third hypervariable region.

In Figure 3 the N-terminal sequences of the V_H regions from M104 and J558 are compared with the corresponding portions of 13 other V_H regions derived from myeloma proteins that do not bind dextran. The N-terminal 53 residues of V_H regions from dextran-binding myeloma proteins differ by 23 (43%) to 27 (50%) amino acid substitutions from the other myeloma V_H regions. Accordingly, the V_H regions from myeloma proteins binding dextran are quite distinct from other myeloma V_H regions.

μ And γ V_H regions from dextran antibodies exhibiting IdX and IdI(M104) determinant(s) are homogeneous and identical to their myeloma counterparts at their N-termini. Table I summarizes the idiotypic composition of the antibody pools that were used to derive heavy chains for amino acid sequence analyses. The "high Id" pool is essentially equivalent to J558 and M104 in anti-IdX reactivity and reacts half as well as M104 with anti-IdI(M104). In Figure 2 are given the N-terminal 27 residues of the V_γ and V_μ regions derived from dextran antibodies with high levels of IdX and IdI(M104) determinants. Several aspects of these analyses are interesting. First, yields at the first

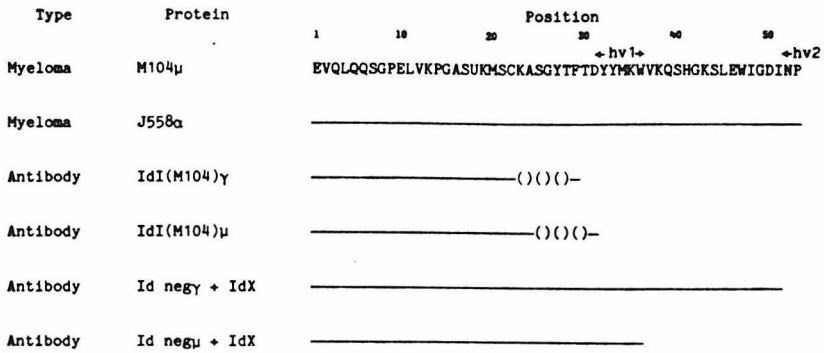


Figure 2. The N-terminal amino acid sequences of heavy chains derived from murine myeloma proteins (M104 and J558) and normal antibodies binding dextran. A straight line indicates identity to the M104 sequence. The one letter amino acid nomenclature of Dayhoff is used. The abbreviations hv1 and hv2 denote the first and a portion of the second hypervariable regions. γ, α, and μ denote, respectively, heavy chains isolated from IgG, IgA, and IgM molecules. ( ) indicate some uncertainty in residue assignment.

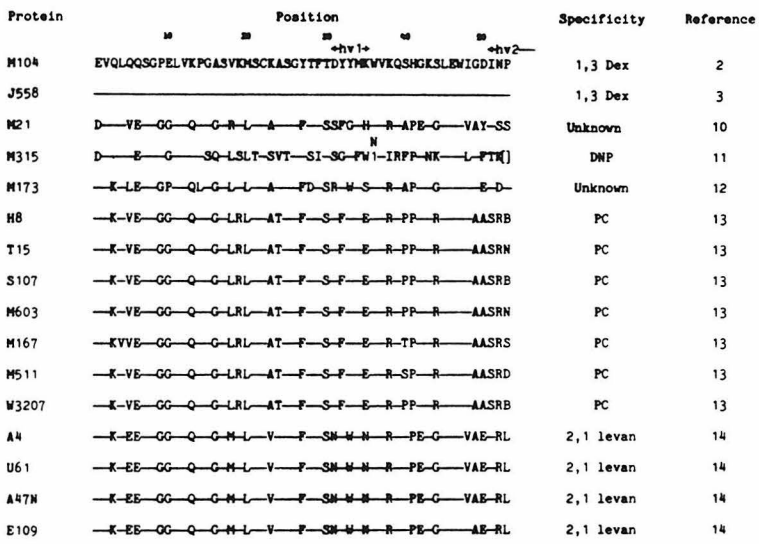


Figure 3. The N-terminal sequences of 16 V_H regions from myeloma proteins of the mouse.

TABLE I

Idiotype composition of anti-dextran antibody pools

Antibody Pools		Idiotypic Determinants[a]		
		IdX	IdI(M104)	IdI(J558)
High Id Pool[b]	IgG	1.73	0.46	<0.01
	IgM	1.08	0.58	<0.01
Low Id Pool[c]	IgG	0.20	<0.01	<0.01
	IgM	0.37	<0.01	<0.01
J558		1.00	<0.01	1.00
M104		1.00	1.00	<0.01

[a] Relative capacity of antibody compared to M104 (or J558) proteins to inhibit binding of ^{125}I-M104 (or ^{125}I-J558) to idiotypic antisera absorbed to microtiter wells. Variability within a 2-fold range is not unusual for this assay.

[b] Pool of five sera.

[c] Pool of six sera.

TABLE II

Yields at the first cycle of sequence analysis for the V_H regions from homogeneous myeloma proteins and antibodies binding dextran[a]

Source	Protein	Yield at Cycle 1
		%
Myeloma	M104μ	32
Myeloma	J558α	38
Antibody	High Id γ	30
Antibody	High Id μ	35
Antibody	Low Id γ	76
Antibody	Low Id μ	57

[a] The yields were calculated by dividing the nanomoles of PTH-glutamic acid recovered at the first cycle by the number of nanomoles of heavy chain employed for sequence analysis.

step of these V_H sequences of dextran antibodies fall in the range of yields obtained for the heavy chains from dextran-binding myeloma proteins (Table II). This indicates that a majority of the antibody V_H regions are probably being sequenced. Approximately 70% of the normal serum heavy chains typically have blocked α-amino groups (15). Accordingly, the dextran immunization procedure has led to a population of antibody molecules that predominantly express unblocked α-amino groups. Second, both the V_γ and V_μ sequences appear to be as homogeneous as their myeloma counterparts. This homogeneity can be clearly demonstrated, for example, by the presence of single residues at positions where a majority of the pool of sequenced myeloma V_H regions generally exhibit different residue alternatives. For example, the combination of glutamine at positions three, five, and six, and glutamic acid at 10 are not found in any other known BALB/c myeloma V_H regions (Fig. 3). In the N-terminal 10 residues or so of the antibody V_H regions, minor sequences of even a few percent could be detected. No minor sequences are seen. Moreover, the residues at

positions 10, 16, 18, 20, and 23 are unique to the heavy chains from dextran-binding myeloma proteins (Fig. 3). In the γ- and μ-chains from the high Id dextran antibodies, only single residues are seen at each of these five positions. Thus, we conclude that the high Id antibody V_H regions are as homogeneous as M104 and J558 at their N-termini. Third, the sequences of the γ- and μ-chains of the high Id antibodies are identical to one another. This observation coupled with their idiotypic identity is consistent with the supposition that a single V_H region can be associated with two or more different constant genes (i.e., C_γ and C_μ) (16). Of course, complete amino acid sequence analyses of the V_H regions will be required to substantiate fully this hypothesis. Finally, the V_H sequences of the dextran antibodies are identical to those of their myeloma counterparts up to the first hypervariable region (i.e., throughout the first framework region). This observation confirms by direct amino acid sequence analysis the limited nature of the heterogeneity at the N-termini of the high Id γ- and μ-heavy chains of the dextran immune response.

V_H regions derived from dextran antibodies deficient in all four of the dextran idiotypes are homogeneous and identical to V_H regions from myeloma proteins M104 and J558 at their N-termini. In Figure 2 are presented the N-terminal amino acid sequences of γ- and μ-chains derived from dextran antibody pools labeled "low Id" pools, lacking the IdI(M104) and IdI(J558) determinants (<1%) and having low levels of the IdX determinant, 20 and 37%, respectively (Table I). These V_γ and V_μ sequences appear to be homogeneous throughout (Fig. 4) and are identical over the regions analyzed to their myeloma and high Id antibody counterparts (Fig. 2). Moreover, the yields at the first cycle are even better than their myeloma counterparts, indicating that a majority of the antibody heavy chains had unblocked α-amino groups (Table II).

Several interesting conclusions can be drawn from these data. i) Quantitative arguments allow us to conclude that the Id deficient and probably the IdX V_H regions are identical to one another and their myeloma counterparts for at least 30 or 40 residues. For example, the yield at the first cycle for the low Id γ-chains was 76% (Table II). The pool from which this chain was derived had only 20% IdX positive molecules (Table I). Accordingly, at least 74% (73/99) of the γ-chains analyzed must be idiotype deficient. ii) The IdX V_H regions were almost certainly sequenced along with the Id deficient V_H regions. Indeed, if 76% of each of these two γ populations were sequenced, then 80% of the sequenced material should be Id-deficient V_H regions and 20% IdX V_H regions. The important point is that we should be able to detect minor sequences at the 10 to 20% level for 30 or 40 residues. Accordingly, both the Id-deficient and IdX V_H regions appear identical at least through the first hypervariable region. Indeed, the Id-deficient antibodies appear to be identical to their myeloma counterparts for 53 residues (Fig. 2). These observations indicate that *all* of the analyzed myeloma proteins and serum antibodies binding dextran appear to be identical at the N-terminus of their V_H regions. The idiotypic and charge diversity must reside in the more C-terminal portion of the V_H region.

The dextran system affords a unique opportunity for studying antibody diversity. Several features combine to make the dextran system an ideal model for analyzing mechanisms of antibody diversity. i) The IEF studies on the λ light chains from the dextran-binding antibodies suggest that the amino acid sequence diversity will reside primarily in the heavy chains. ii) Because the five cases of idiotypically distinguishable dextran-binding antibodies are generally found in individual mice,

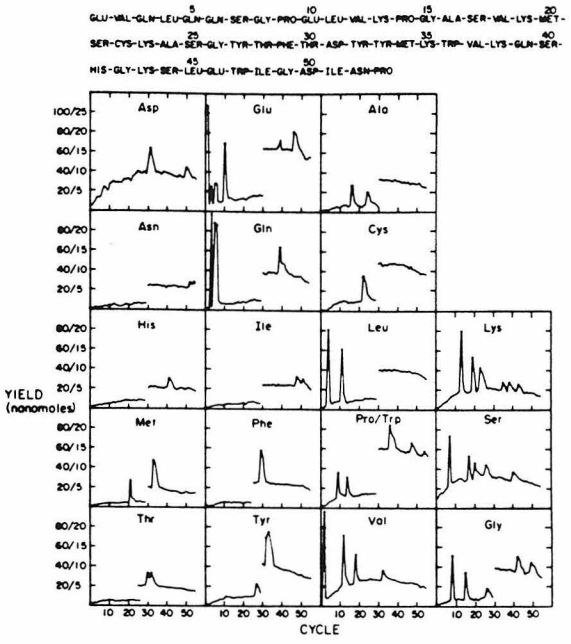


Figure 4. Yields of amino acid phenylthiohydantoins from an N-terminal analysis of 130 nmole of the IdI deficient γ-chain derived from the low Id dextran antibody pool (see *Materials and Methods*; Table I). The abscissa denotes the cycle or step in the sequenator run. The ordinate gives the nanomole yield of all amino acid derivatives for each cycle. The ordinate scale is expanded 4-fold at cycle 30 for all residues except lysine and valine, which are plotted exclusively on the reduced scale and serine and aspartic acid, which are plotted exclusively on the expanded scale. Proline and tryptophan were not resolved by high pressure liquid chromatography. Gas chromatography was used to identify proline at positions 9, 14, and 53 and tryptophan at positions 36 and 47. Cysteine was identified as the carboxyamidomethyl cysteine phenylthiohydantoin derivative.

it appears that one or more distinct germline V_H genes or gene segments (17) will encode each of the five specificities: IdI(M104), IdI(J558), IdI(U102), IdX, and Id negative. Genetic and nucleic acid analyses could be employed to determine how closely linked these V_H genes are. iii) V_H regions containing the IdI(M104), IdI(J558), and Id-deficient specificities appear to be identical for 53 or more residues (Fig. 2). Indeed, the M104 and J558 heavy chains appear to be identical up to the third hypervariable region (J. Schilling and C. Oken, published observations). Accordingly, the amino acid residues responsible for the distinct idiotypic specificities may lie in or near the third hypervariable region. iv) This diversity might be explained by any one of three basic mechanisms: 1) multiple germline V_H genes, 2) somatic mutation, and/or 3) combinatorial joining of V_H and J_H segments. With regard to this latter possibility, in mouse λ- and κ-chains, the V region is encoded by two separate DNA segments; the V segment (encoding ∽ residues 1 to 100) and the J segment (encoding ∽ residues 101 to 112) (17, 18). In mouse κ-chains, different J segments may be associated with the same V segments to produce a sequence and charge heterogeneity in the V region (17). One provocative possibility is that the diversity of the dextran system could be explained in part by existence of multiple J_H segments of varying sequence and charge being associated with an identical V_H segment.

We are now isolating homogeneous antibodies to dextran by

employing the cell-fusion technique of Köhler and Milstein (19). We plan to localize the patterns and extent of the diversity in the V_H regions of these molecules and to analyze the λ-chains for the presence of possible minor sequence diversity. The dextran system, because the diversity appears to be concentrated in or near the third hypervariable region of the heavy chain, appears to afford a unique opportunity for determining the molecular basis of diversity in the normally induced antibodies raised against dextran.

REFERENCES

1. Hansburg, D., D. E. Briles, and J. M. Davie. 1977. Analysis of the diversity of murine antibodies to dextran B1355. II. Demonstration of multiple idiotypes with variable expression in several strains. J. Immunol. 119:1406.
2. Kehry, M., J. Fuhrman, C. Sibley, J. Schilling, and L. Hood. 1979. The amino acid sequence of a mouse immunoglobulin mu chain. Proc. Natl. Acad. Sci. In press.
3. Hunkapiller, M. W., and L. E. Hood. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (Polybrene), and high pressure liquid chromatography. Biochemistry 17:2124.
4. Hansburg, D., D. E. Briles, and J. M. Davie. 1976. Analysis of the diversity of murine antibodies to dextran B1355. I. Generation of a large, pauci-clonal response by a bacterial vaccine. J. Immunol. 117: 569.
5. Hansburg, D., R. M. Perlmutter, D. E. Briles, and J. M. Davie. 1978. Analysis of the diversity of murine antibodies to dextran B1355. III. Idiotypic and spectrotypic correlations. Eur. J. Immunol. 8:352.
6. Blomberg, B., W. R. Geckeler, and M. Weigert. 1972. Genetics of the antibody response to dextran in mice. Science 177:178.
7. Riblet, R., B. Blomberg, M. Weigert, R. Lieberman, B. A. Taylor, and M. Potter. 1975. Genetics of mouse antibodies. 1. Linkage of the dextran response locus, V_H-DEX, to allotype. Eur. J. Immunol. 5:775.
8. Weigert, M., and R. Riblet. 1977. Genetic control of antibody variable regions. Cold Spring Harb. Symp. Quant. Biol. 41:837.
9. Cesari, I. M., and M. Weigert. 1973. Mouse lambda-chain sequences. Proc. Natl. Acad. Sci. 70:2112.
10. Milstein, C., K. Adetuybo, N. J. Cowan, and D. S. Secker. 1974. Clonal variants of myeloma cells. Prog. Immunol. 1:157.
11. Francis, S. H., R. G. Q. Leslie, L. Hood, and H. N. Eisen. 1974. Amino-acid sequence of the variable region of the heavy (alpha) chain of a mouse myeloma protein with anti-hapten activity. Proc. Natl. Acad. Sci. 71:1123.
12. Bourgeois, A., M. Gaugereau, and C. Depreval. 1972. Sequence of amino acids of the NH_2-terminal region of a mouse-clonal immunoglobulin heavy chain. Eur. J. Biochem. 24:446.
13. Hood, L., E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg, and J. Schilling. 1976. The structure and genetics of mouse immunoglobulins: an analysis of NZB myeloma proteins and sets of BALB/c myeloma proteins binding particular haptens. Cold Spring Harbor Symp. Quant. Biol. 41:817.
14. Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin-binding myeloma proteins. Proc. Natl. Acad. Sci. 75:1957.
15. Capra, J. D., R. W. Wasserman, and J. M. Kehoe. 1973. Phylogenetically associated residues within the V_H III subgroup of several mammalian species: Evidence for a "pauci-gene" basis for antibody diversity. J. Exp. Med. 138:410.
16. Sledge, C., D. S. Fair, B. Black, R. G. Krueger, and L. Hood. 1976. Biclonal myeloma proteins: apparent sequence identity between variable regions shared by IgA and IgG immunoglobulins. Proc. Natl. Acad. Sci. 73:923.
17. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. Nature 276:785.
18. Brach, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa. 1978. Cell 15:1.
19. Kohler, G., and C. Milstein. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256:495.

CHAPTER 4

This paper was published in Nature.

Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangments in heavy chain V-region gene segments

J. Schilling*, B. Clevinger†, J. M. Davie† & L. Hood*

* Division of Biology, California Institute of Technology, Pasadena, California 91125

† Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110

The complete variable region sequences from ten antibodies and two myeloma proteins binding α-1,3 dextran have been determined. The diversity patterns of these homogeneous antibody molecules suggest that the variable regions of heavy chains are encoded by separate variable (V) and joining (J) gene segments. The most striking feature of these data is the extensive sequence variability of a region that we denote the D (diversity) segment which is located at the junction between the V and J segments in the centre of the third hypervariable region. The D segment diversity may arise from a novel somatic mutational mechanism or may be encoded by multiple D gene segments. For the first time, the amino acid sequence correlates of several V region idiotypes are determined.

STUDIES of immunoglobins at the protein and DNA levels have led to striking advances in our understanding of the organisation and evolution of antibody genes and mechanisms of antibody diversity[1-4]. There are three unlinked families or clusters of antibody genes—the λ and κ gene families encode light (L) chains and the heavy (H) gene family encodes heavy chains[5]. The κ or λ gene is composed of four distinct gene segments, leader (L), variable (V), joining (J) and constant (C), which are separated by intervening sequences in undifferentiated or germ-line DNA[6-12]. Heavy chain genes probably have a similar organisation of gene segments (P. Early, H. Huang, M. Davis and L.H., unpublished observations). During the differentiation of antibody-producing cells, the V_L and J_L gene segments are juxtaposed and together encode the variable or antigen-binding domain of the antibody molecule. (The term V or J segment will refer to the corresponding protein sequence; the term V or J gene segment will refer to the corresponding DNA coding regions.) Comparative amino acid sequence analyses of V regions showed that for both light and heavy chains, there are three regions of extreme variability termed hypervariable regions[13,14], which fold in three dimensions to constitute the walls of the antigen-binding site[15,16]. The diversity in the hypervariable regions of L chains has been most extensively studied and explained by a variety of genetic mechanisms. (1) There are many germ-line V_κ (and J_κ) gene segments[5,10]. (2) A single V_κ gene segment may be joined to many different J_κ gene segments and conversely, a single J_κ gene segment may be joined to many different V_κ gene segments[17]. As the boundary of the V_L and J_L segments falls within the third hypervariable region, the combinatorial joining of these segments contributes to the diversity of antigen-binding sites. (3) Random single base mutations may occur throughout the V_L (and J_L) gene segments and those variants exhibiting altered and useful antigen-binding properties may be selected for clonal expansion. For example, 12 of 18 V_λ regions examined were identical to one another[18-20]. The six variant V_λ sequences differed from the 12 identical sequences by between one and three amino acid substitutions. All these substitutions occur in the hypervariable regions. These data suggest that a single V_λ gene is often expressed unchanged and less frequently is modified by somatic mutation to produce the variants. Recent DNA sequence analyses of germ-line and myeloma λ gene segments are consistent with this supposition[6,11,12]. The pattern of diversity wherein variant V region sequences exhibit few substitutions mostly concentrated in the hypervariable regions will be denoted λ-type diversity. (4) A site-specific recombinational mechanism has been postulated to explain amino acid sequence diversity at the junction of V_κ and J_κ gene segments[21-23]. This diversity is reflected as single codon substitutions at the N-terminal residue of the J_κ segment and as the insertion or deletion of single codons precisely at the V-J junction[23]. This pattern of diversity will be denoted the V-J junctional diversity.

We have analysed the diversity patterns in mouse antibodies directed against a simple antigenic-determinant, α-1,3 dextran, in order to determine the relative contributions of the diversification mechanisms discussed above. The hybridoma technique of Köhler and Milstein facilitates the analyses of antibody diversity by making large quantities of antigen-induced, homogeneous antibody molecules available[24]. This is accomplished by fusing normal antibody-secreting cells with mouse plasmacytoma cells[25]. The hybrid cells generate immortal cell lines that secrete large amounts of homogeneous antibody molecules of the two parental cell types. The light chains of the dextran antibodies are all of the λ type and appear to be very similar, if not identical, to one another. Hence, most of the diversity in the dextran system appears to reside in the heavy chains. We report here the complete amino acid sequences from the variable regions of 10 heavy chains derived from hybridoma antibodies to dextran. These data suggest that heavy chains have a D (diversity) segment located in the heart of the third hypervariable region. The D segment may be encoded by a separate germ-line gene segment or generated by a new type of somatic mutational mechanism.

Immunoglobulins binding dextran

Immunoglobulins binding dextran exhibit limited heterogeneity in several respects. First, preliminary amino acid sequence studies on two myeloma proteins binding α-1,3 dextran, M104E and J558, revealed that the V_L regions are identical[18,19] and that the V_H regions are identical through their first hypervariable regions[26]. The N-terminal sequences of V_H regions derived from

Table 1 RNA codons of positions 100 and 101 from V_H regions of myeloma and hybridoma proteins binding dextran*

V-J segments†	Proteins	Codons‡ 100	Codons‡ 101
		a b c	a b c
V_1-J_1	M104E	UAY	GAY
	J558	AGZ CGX	UAY
	Hdex2	AAY	UAY
	Hdex3	AGZ CGX	GAY
	Hdex7	GCX	GAY
	Hdex6	AGY UCX	CAY
V_2-J_1	Hdex8	UAY	GAY
V_3-J_1	Hdex9	AGZ CGX	UAY
V_4-J_1	Hdex10	GUX	AAY
V_1-J_2	Hdex4	AAZ	GAY
	Hdex5	AGY UCX	AAY
V_1-J_3	Hdex1	AAY	UAY

* These codon assignments are derived from the corresponding amino acid sequences

† The V and J segments are defined in Fig. 2.

‡ Y denotes pyrimidine; Z designates purine; and X denotes any one of the four bases.

normally-induced pooled serum antibodies to α-1,3 dextran also are identical to their myeloma counterparts[27]. Second, isoelectric focusing analyses of anti-dextran antibodies of the IgG class demonstrate limited heterogeneity with considerable sharing of bands between sera from different mice[28]. Moreover, the serum antibody light chains are identical to one another and to their myeloma counterparts by isoelectric focusing criteria[27]. Third, the light chains from the 10 hybridomas and 2 myeloma proteins reported here are all of the λ type and have identical isoelectric focusing patterns (J.S., unpublished observation). Thus, light chains from dextran antibodies appear to be very similar, if not identical, to one another. Therefore, most of the diversity in the antibodies to α-1,3 dextran must reside in the heavy chains. Fourth, the anti-α-1,3 dextran antibodies can be divided into several general categories by idiotypic analyses[29]. More than half of the anti-α-1,3 dextran antibodies from mice immunised with the branched dextran B1355 share a variable-region antigenic determinant or idiotype with both the M104E and J558 proteins. This idiotype is absent from non-dextran binding immunoglobulins and λ light chains and is denoted the 'cross-reactive' or IdX idiotype. Antibodies with the IdX idiotype are synthesised by multiple different antibody-producing clones. Additional idiotypes specific for either the M104E or the J558 immunoglobulins are present on a minor fraction of the normally-induced antibodies. These idiotypes are denoted as individual or IdI idiotypes (for example, IdIM104E and IdIJ558). Another general category of dextran antibodies is that of those lacking the IdX determinant (IdX negative). The IdX-positive and IdX-negative antibodies to dextran are structurally very similar in that their light chains are identical by isoelectric focusing analysis and the heavy chains from both the IdX-positive and IdX-negative pools of serum antibodies are identical to the heavy chains of the M104E and J558 immunoglobulins for their N-terminal 27 and 53 residues, respectively[27]. These observations show that much of the V domain diversity in normal antibodies to α-1,3 dextran resides in the heavy chain beyond residue 53. Thus, we turned to the amino acid sequence analysis of homogeneous normal antibodies to α-1,3 dextran.

Diversity patterns in V_H regions from antibodies binding dextran

Figure 1 gives the complete amino acid sequences of V_H regions from 2 myeloma proteins and 10 hybridoma antibodies that bind dextran. The hypervariable regions and the variable region site of carbohydrate attachment are indicated. These V_H regions are compared to the published sequence of the myeloma M104E V_H region[30] with amino acid substitutions indicated by the one-letter code[31].

Two features are of interest in this set of V_H regions. First, each of these 12 V_H regions differs from all of the others by 1 to 13 amino acid substitutions (Fig. 1). We have not yet detected a pair of identical V_H regions in 12 attempts. Accordingly, the V_H sequence diversity in the antibody response to α-1,3 dextran is extensive. Second, patterns of variability divide the V_H regions from dextran antibodies into three distinct sections—V(1–99), V(100–101) and V(102–117). The highly variable section, V(100–101), separates the other two sections which are relatively conserved. Section V(100–101) exhibits nine different sequences in the 12 V_H regions (Fig. 2). Indeed, position 100 has eight alternatives and position 101 has four. In contrast, nine V_H regions have identical sequences for section V(1–99) and the remaining V_H regions differ by between two and seven residues from the others. Likewise, nine V_H regions have identical sequences for the section V(102–117), two have a second sequence (Hdex4 and Hdex5) and one has a third sequence (Hdex1). Thus, the relatively conserved sections V(1–99) and V(102–117) are separated by two highly variable residues at positions 100 and 101.

This same three-section pattern is seen when 20 completely sequenced V_H regions are aligned by homology with their anti-dextran counterparts (Fig. 3). The section homologous to V(1–99) can clearly be defined in each of these 20 V_H regions and each terminates precisely at the position homologous to 99 in the heavy chains binding dextran (Fig. 3). The homology relationships among the V(102–117) sections are also obvious. The V(102–117) sections from the various V_H regions are identical in size, apart from the levan-binding immunoglobulins, and exhibit extensive amino acid sequence homology to one another (56–100% sequence identity). Accordingly, sections homologous to V(1–99) and V(102–117) in the dextran V_H regions are clearly present in all of the other V_H regions examined to date. In contrast, there is no homology in the V(100–101) regions between these two sections because of extensive sequence and size variability (Fig. 3). The sections analogous to V(100–101) can be defined only as those residues remaining after the V(1–99) and V(102–117) sections have been assigned for each V region by homology.

Implications of V_H diversity patterns for gene organisation

The patterns of amino acid diversity in the V_H regions from antibodies binding α-1,3 dextran allow us to define tentatively V_H and J_H segments with properties similar to their light chain counterparts (Figs 1–3). The V(1–99) section appears to correspond to the V_H segment and V(102–117) section appears to represent the J_H segment. The V(100–101) section will be called the D (diversity) segment and its origin will be discussed.

The extensive variability in these V_H regions at position 100 appears to mark the C-terminal end of the V_H segment, just as extensive variability at position 96 marks the C-terminal end of the V segment in kappa light chains[17]. The correspondence of V(1–99) to a V_H gene segment is confirmed by the DNA sequence analysis of a germ-line V_H gene segment which ends precisely at codon 99 (P. Early, H. Huang, M. Davis and L.H., unpublished observation).

The J_H segment includes at least positions 102–117. The C-terminal boundary of the J_H segment at position 117 has been

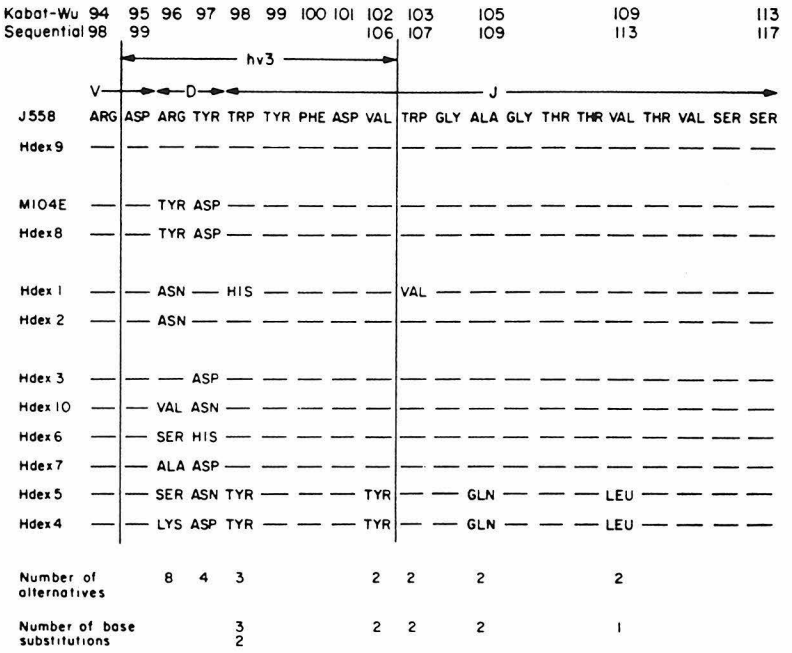


Fig. 2 The C-terminal portions of V_H regions from immunoglobulins binding dextran. See Fig. 1 legend. The number of alternative amino acids for each position at which variability exists are indicated. The number of base substitutions denotes the minimum number of nucleotide changes required to convert the most common codon at a particular position into the alternative codons at that position. The V_1 to V_4 and J_1 to J_3 segments are defined. The third hypervariable region is denoted by hv3. Three pairs of proteins with identical sequences at positions 100 and 101 are set apart from the rest. The V_H, D and J_H segments are defined in Fig. 1.

identical J_H segments may be associated with differing V_H segments (such as M104E and Hdex8; J558 and Hdex9). Conversely, identical V_H segments may be associated with different J_H segments (for example, Hdex1, Hdex2 and Hdex4) (Fig. 1). Even more striking is the observation that the J_1 segment in V_H regions from antibodies binding dextran also occurs in the very different V_H regions from myeloma proteins binding phosphorylcholine (Fig. 3). Likewise, the same J_H segment may be found in heavy chains binding phosphorylcholine and galactan (for example, M167.and T601). Thus, the combinational joining of J_H and V_H gene segments appears to be an important source of diversity in these dextran-binding antibodies.

D segment diversity. This most striking diversity pattern in these V_H regions (Figs 2, 3). Diversity in the D segments may arise from one of three general mechanisms: (1) a novel somatic mutation mechanism possibly occurring as a result of the joining of V_H and J_H gene segments; (2) a large number of germ-line J_H or V_H gene segments which include the D segment; or (3) a third gene segment (D) encoding positions 100 and 101 which is joined with the V_H and J_H gene segments during the differentiation of an antibody-producing cell. Let us consider what constraints the diversity patterns of the V_H regions impose on each of these models.

Several observations argue against many germ-line V_H gene segments extending from codons 1 to 101. First, identical D segments may be associated with different V_H segments—M104E-Hdex8 and J558-Hdex9 (Fig. 1). Thus, the D segments appear to associate independently with V_H segments. Second, the T15 V_H gene segment terminates at codon 99 (P. Early, H. Huang and L.H., unpublished observation). Accordingly, the D segment cannot be encoded as part of the V_H gene segment for the T15 V_H gene segment (Fig. 3). Identical D segments may be associated with different J_H segments—Hdex1-Hdex2 and X44-J539 (Fig. 3). A more striking example of this independent association of D and J_H segments is the observation that a J_1 segment is associated with very different D segments in immunoglobulins binding distinct haptens—T15 and M104E (Fig. 3). The independent segregation of D segments with both V_H and J_H segments suggests that the D coding regions are not merely germ-line extensions of V_H or J_H gene segments. In addition, homology comparisons among the V_H regions very clearly define the V_H and J_H segments (Fig. 3). In contrast, the D segments lack constancy both in size (0 to 7 residues) and amino acid sequence (20 different sequences in 25 different V_H regions compared) (Fig. 3). The V_H regions in Fig. 3 would require a very large number of germ-line gene segments if the V_H or J_H gene segments include the D segment.

The D segment diversity is quite distinct from the V-J junctional diversity of kappa chains in several regards. First, there are extensive size differences among the D segments whereas the κ junctional diversity shows at most the insertion or deletion of just a single codon[23]. Second, the two codons for the D segment associated with a particular $V_H(V_1)$ and a particular $J_H(J_1)$ segment exhibit three different nucleotide alternatives at three positions—100a, 100b and 101a (Table 1). This observation is important because it suggests that D segment diversity cannot be generated solely by recombination between the V_1 and J_1 gene segments as is suggested by the site-specific recombinational model of mouse kappa chains[21-23]. This model allows for only two alternatives at any nucleotide position. Accordingly, the extensive variability, the size differences and the presence of three bases at several nucleotide positions in the D segments constitute a distinctly new pattern of diversity that must be explained by a novel mechanism for somatic mutation or the presence of multiple D gene segments.

The striking base invariance at nucleotide positions 101b and 101c in the midst of the remarkable adjacent diversity raises several interesting possiblities (Table 1). Perhaps these invariant nucleotides form part of a recognition site for an enzyme that might mediate somatic mutation or join the D and J_H gene segments together. Alternatively, perhaps these invariant nucleotides are a part of the J_H gene segment.

If the D segment is encoded by a gene segment separate from the V_H and J_H gene segments, then heavy chain genes differ in a major organisational feature from light chain genes. Moreover, the presence of a putative D gene segment creates a second gene segment boundary (V_H–D and D–J_H) at which junctional diversity of the type noted in kappa chains can occur[21-23].

Antigen-binding diversity and V_H variability

The third hypervariable region forms an important part of the antigen-binding site[15,16]. Diversity in the dextran antibodies is

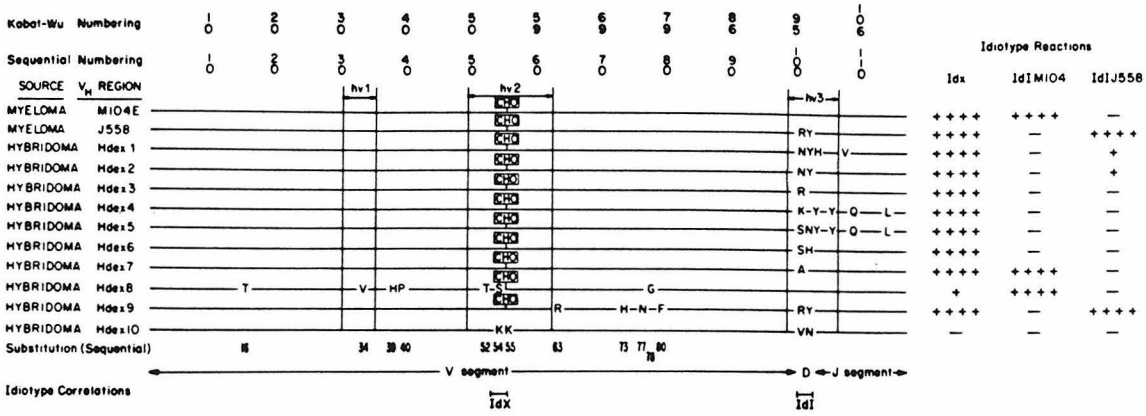


Fig. 1 The amino acid sequences and idiotypes from V regions of myeloma proteins and homogeneous hybridoma antibodies binding dextran. The myeloma protein M104E is used as a prototype sequence and variation from this sequence is indicated. The hypervariable regions are indicated by hv1, hv2 and hv3, respectively. CHO indicates the attachment of a carbohydrate moiety. The one-letter amino acid code of Dayhoff is used[31]. The V and J segments are indicated by arrows. The D segment includes positions 100 and 101. Sequential numbering denotes consecutive numbering of the V_H regions. The Kabat–Wu numbering is from ref. 36. Details of the complete heavy chain sequence of the M104E immunoglobulin are published elsewhere[30]. A similar strategy was used for the sequence analysis of the V_H regions from the hybridomas. The positions of amino acid substitutions in V segments are indicated. Eight hybrids were derived from individual BALB/c mice: Hdex6, 7, 8 and 9 were obtained at 5 or 6 days after a single injection of 100 μg dextran B1355 in complete Freund's adjuvant (CFA), and Hdex2, 3, 4 and 10 were obtained 1–3 days after hyperimmunisation with dextran in CFA followed 1–2 months later with 3 interperitoneal injections of 2×10^9 *Escherichia coli* at 2-day intervals. Hybrids Hdex1 and 5 were derived from a single (BAB-14 × BRVR)F_1 mouse at 7 days after the dextran–*E. coli* protocol. Hybrids were generated as described by Galfre *et al.*[41] using the HGPRT-deficient MPC-11 ($\gamma_{2b}\kappa$) line, 45.6TG1.7, (Hdex1, 3, 4, 5 and 10) or a non-secreting variety of the MOPC-21 ($\gamma_1\kappa$) line, NSI/1-Ag4-1, which synthesises but does not secrete κ chains (Hdex2, 6, 7, 8 and 9). Anti-dextran secreting hybrids were detected by direct binding of ^{125}I-dextran by culture supernates either in a tube-binding assay or after isoelectric focusing in polyacrylamide gels. Hybrids were cloned in soft agar[42] over feeder layers and grown in BALB/c mice as ascites tumours. The antibodies produced by the hybrids appeared representative of those found in the serum at the time of fusion. Idiotypes were determined as described by ref. 29. ++++ Indicates a reaction 10 times as strong as +; − indicates no reaction.

defined by the DNA sequence analysis of an expressed V_H-J_H gene (H. Huang, P. Early, M. Davis and L.H., unpublished observations). This analysis demonstrates that the C_H region begins at a position homologous to residue 118 in the heavy chains from antibodies binding dextran, in agreement with the J_H and $C_{\gamma 1}$ gene segment boundaries defined by DNA sequence analysis of a $C_{\gamma 1}$ gene[32]. The N-terminal boundary of the J_H segment is unclear. Figure 2 shows three J_H sequences which differ from each other by two or three two-base substitutions (J_1, J_2 and J_3). Position 102 is the first position in the J_H segment at which linkage relationships among amino acid residues appear. A tryptophan at position 102 is invariably linked to valine, alanine and valine residues at positions 106, 109 and 113, respectively, in the J_1 segments, whereas the tyrosine at position 102 is linked to tyrosine, glutamine and leucine at these positions in the J_2 segments (Fig. 2). These linkage relationships imply that the J_1 and J_2 segments are encoded independently in the germ-line. The D segment comprises positions 100 and 101 and may or may not be a part of the J segment, depending on how V–J junctional diversity arises. These assignments of the V_H, J_H and D segment boundaries also are supported by the homology relationships exhibited among the other completely sequenced V_H regions (Fig. 3).

The D and J_H segments comprise almost the entire third hypervariable region (Fig. 2). This is in striking contrast to the J_κ segment which just extends into the third hypervariable region by a single residue[17]. Moreover, the size variation seen in the D segment of heavy chains is far more extensive than that seen at the V–J boundary of kappa chains.

Rudikoff and coworkers have used more limited V_H region data to suggest the V_H segment extends from 1 to 98 and the J_H segment from 105 to 117. They conclude that positions 99 to 104 may be part of the V segment or, alternatively, part of the J segment[33]. These assignments are different from those made here.

Implications of V_H diversity patterns for mechanisms of antibody diversity

The V_H regions from dextran antibodies all differ from one another. Let us analyse possible contributions to this diversity by multiple germ-line gene segments, somatic mutation and the combinational joining of V_H and J_H segments.

Germ-line V_H and J_H gene segments. Four different amino acid sequences corresponding to the V_H gene segment are present in the 12 antibodies binding dextran and these are denoted V_1 to V_4 in Figs 1 and 2. The V_1 segment is expressed in nine different immunoglobulins binding dextran and, accordingly, appears to be encoded by a germ-line gene. This supposition follows from the improbability of generating the same V_H sequence many times by somatic mutation[4]. The V_2, V_3 and V_4 segments differ from V_1 by seven, four and two residue substitutions, respectively. The V_2, V_3 and V_4 segments may be encoded by distinct germ-line V_H gene segments or these sequences may arise from one or more somatic mutational mechanisms. However, the diversity patterns among the V_2, V_3 and V_4 segments do not resemble either of the two somatic mutational patterns described earlier for light chains, λ-type diversity and V_κ-J_κ junctional diversity.

The three different J_H segments differ from one another by two to five residue alternatives (Fig. 2). Moreover, most of these J_H segment substitutions require two or even three base substitutions and repeats of two of these sequences have been seen. These observations strongly suggest that the three J_H sequences are directly encoded by three germ-line J_H gene segments. The J_H diversity noted in Fig. 3 appears to be significantly greater than the corresponding J_L diversity[17] and, accordingly, may be encoded by a larger number of germ-line J_H gene segments.

Combinational joining. The combinational joining of V_H and J_H gene segments to encode many distinct V_H regions provides yet another source of antibody diversity (Fig. 4). For example,

Protein	Specificity	V segment	D segment	J segment	References
		75 80 90 99	100 101	102 110 117	
M104E	1,3 Dextran	SSSTAYMQLNSLTSEDSAVYYCARD	YD	WYFDVWGAGTTVTVSS	32
Hdex8	"	---G----------	YD	----------	This paper
J558	"	----------	RY	----------	"
Hdex9	"	--N--F----------	RY	----------	"
Hdex10	"	----------	VN	----------	"
Hdex6	"	----------	SH	----------	"
Hdex3	"	----------	R-	----------	"
Hdex7	"	----------	A-	----------	"
Hdex2	"	----------	NY	----------	"
Hdex1	"	----------	NY	H----V----------	"
Hdex5	"	----------	SN	Y---Y--Q---L----	"
Hdex4	"	----------	K-	Y---Y--Q---L----	"
T15,S63,S107,Y5236	Phosphorylcholine	-Q-IL-L-M-A-RA--T-I------	YYGSSY	----------	46
H8	"	-Q-IL-L-M-A-RA--T-I------	---N--	----------	"
W3207	"	-Q-IL-F-M-A-RA--T-I-----N	--KYDL	--V----------	"
M603	"	-Q-IL-L-M-A-RA--T-I-----N	----T	----------	47
M167	"	-Q-VL-L-M-A-RA--T-T---T--	ADYGDSYF	G----------	48
M511	"	-Q-IL-L-M-A-RA--T-I------	GDYG-SY	----------	"
A4	2,1 Levan	-K-SV-L-M-N-RA--TGI---TTG	[ ]	[ ]-AY--Q--L---	49
E109	"	-K-SVFL-M-N-RA--TGIH--TTG	[ ]	[ ]-AY--Q--L---	"
U61	"	-K-SV-L-M-N-RA--TGI---TTG	[ ]	[ ]-AY--Q--L-P-	"
A47N	"	-K-SV-L-M-N-RA--T-I---STG	[ ]	[ ]-AY--C--L---	"
M315	DNP	-ENQFFLK-D-V-[ ]T-T----G-	NDH	L---Y--Q---L---	50
M21	Unknown	PKN-LFL-MT--R---T-M-----H	GNYPW	YAM-Y--Q--S-----	51
M173	Unknown	AKN-L-L-MSKVR---T-L-----S	PY	YAM-Y--Q--S-----	52
T601	Galactan	AKN-L-L-MSKVR---T-L-----G	GYY	G----------	35
X44	"	AKN-L-L-MSKVR---T-L-----L	H--	G-AAY--Q--L----A	"
X24	"	AKN-L-L-MSKVR---T-L-----G	---	G---Y--Q---L----	"
J539	"	AKNSL-L-MSKVR---T-L-----L	H--	G-NAY--Q--L----A	"

Fig. 3 A comparison of published heavy chain sequences including the C-terminal portions of V_H regions. The division of this region into V_H segments, D segments, and J_H segments is based on homology with the heavy chains binding dextran. The prototype sequence for the V_H and J_H segments is M104E. The prototypes for the D segments are the first sequence in each group. The J segments which do not extend to amino acid 117 are due to incomplete amino acid sequence data and not to shorter J segments.

concentrated in the third hypervariable region of heavy chains which includes the D segment and much of the J_H segment (Fig. 2). Studies on the binding specificites of the myeloma proteins M104E and J558 which differ only in the D segment have demonstrated that the antigen-binding properties of these two molecules are quite distinct[34,35]. Accordingly, the mechanism for producing D segment diversity does have an important role in modifying the antigen-binding properties of antibodies.

Variability in the third hypervariable region of the dextran antibodies arises from germ-line diversity (J_H segments), by combinational joining of V_H and J_H (and possibly D_H) segments and by the mechanism for producing D segment diversity. Of these, D segment diversity appears to be by far the most extensive and, presumably, important in generating a diversity of antigen-binding sites (Figs 2, 3).

One striking feature evident in Fig. 3 is that immunoglobulins binding particular haptens share similar V_H segments, exhibit D segments of similar size (for example, zero for levans, four for galactans, and so on), and to a lesser extent share similar J_H segments. Whether these correlations arise through some intrinsic properties of gene organisation, somatic mutational mechanisms or antigen selection is unknown.

Correlation of dextran idiotypes and protein structure

Several interesting conclusions can be drawn about the correlation of V_H structure and the various idiotypes of dextran-binding antibodies (Fig. 1). First, the individual idiotype appears to correlate with the D segment sequence. The V_H regions from the M104E and J558 proteins differ only by their D segments. As the light chains are identical, the individual idiotypes of M104E and J558 must be determined by positions 100 and 101. Moreover, the V_H region of Hdex9, which differs from that of J558 by four residues and is identical at positions 100 and 101, has the individual J558 idiotype. Likewise, the V_H region of Hdex8, which differs from that of M104E by seven residues and is identical in the D segment, also has the individual M104E idiotype. However, a single residue substitution can eliminate the individual idiotype as demonstrated by the observation that Hdex3 is lacking both the M104E and J558 individual idiotypes (Figs 1, 2). Partial and in one case complete IdI reactivity can be retained by D segments sharing one of two residues (Hdex1, Hdex2 and Hdex7). The overall conclusion is that the individual idiotypes of myeloma proteins M104E and J558 are determined by residues 100 and 101 and, accordingly, are either encoded by a third gene segment (D) or represent antigenic determinants generated by a novel somatic mutational mechanism. Second, the IdX idiotype correlates with several residues in the V_H segment. The V_H region from one IdX-negative protein, Hdex10, has been sequenced. Hdex10 differs from the two myeloma V_H regions by four residues—positions 54, 55, 100 and 101. The substitution at position 55 destroys the attachment site for the carbohydrate moiety, distinguishing it from the other 11 V_H regions (Fig. 1). It is attractive to postulate that the IdX idiotype depends on amino acids at positions 54 and 55. This supposition is supported by the observation that Hdex8, which has greatly reduced IdX reactivity, has a substitution at position 54. Analysis of V_H regions from several additional IdX-negative antibody molecules are in progress to determine whether they also exhibit substitutions at positions 54 and 55 and to test the intriguing possibility that the IdX idiotype may be in part associated with a carbohydrate moiety.

Thus, idiotypes can apparently arise from V segments (IdX) or D segments (IdI). These data on the hybridoma antibodies binding α-1,3 dextran do allow us to assign structural correlates for the cross-reactive (V segment) and individual (D segment) idiotypes of several dextran antibodies. However, an idiotype localised to a particular region (for example, the IdIJ558 to positions 100 and 101) may require particular sequences in still other portions of the V region for its expression (for example, V and J segments like those found in dextran-binding antibodies). Accordingly, the interpretation of genetic mapping of idiotypes is complex. Indeed, one can never be certain what is being mapped—a V segment, a D segment, a J segment, the ability to produce certain somatic variants, or some combination of these.

In an historical context, note that Kabat and coworkers predicted the existence of minigenes for each hypervariable region based on an analysis of the V region amino acid sequences[13,36]. In addition, Capra and Kindt postulated the existence of multiple V coding elements based on an analysis of V_H sequence data and idiotypes[37,38]. If the D segments are encoded by distinct gene segments, they are 'minigenes' that encode the individual idiotypes of murine antibodies binding dextran.

The evolution of split-gene segments and DNA rearrangements

The V_H regions and V_L regions are homologous to one another at the protein level[31,39]. Both V_L and V_H are encoded by V and J gene segments of similar size. This homology implies that the split-gene strategy and the mechanism for rearranging split genes must have arisen before the divergence of light and heavy chains. Indeed, it is attractive to speculate that split genes and mechanisms for their rearrangement evolved before the emergence of the vertebrate immune system, possibly in families of genes coding membrane recognition functions. It will be interesting to determine whether there are other contemporary

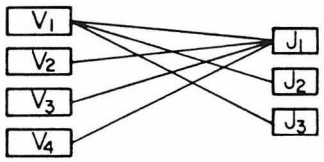


Fig. 4 A diagram indicating the combinational joining of V_H and J_H segments. V_1 to V_4 and J_1 to J_3 are defined in Fig. 2.

multigene families using strategies of split genes and DNA rearrangment for the amplification of information[5]. This is discussed elsewhere[40].

The patterns of phenotypic diversity in the V_H regions from antibodies binding α-1,3 dextran have allowed us to make several important conclusions. We can define V_H and J_H segments which are homologous to their light chain counterparts. Diversity in a third region, the D segment, can be explained by a novel somatic mutational mechanism, the existence of a third germ-line (D) gene segment, or the less likely possibility of a large number of J_H gene segments. Antibody diversity appears to arise from several sources: multiple germ-line gene segments, combinational joining of V_H and J_H gene segments and the mechanism for producing D segment diversity which includes extensive nucleotide substitutions as well as the insertion or deletion of multiple residues. Correlations between the V_H sequences and the idiotypes of the antibodies to α-1,3 dextran are possible but there should be caution in the interpretation of genetic mapping studies which use idiotypes. Studies are in progress in our laboratories to expand our analyses of the phenotypic (protein) diversity in the dextran system and to analyse the gene segments encoding the dextran response.

This work was supported by USPHS Grants AI-10781, AI-15353, AI-11635, AI-05599 and CA-09118, NSF grant PCM76-09719, and National Institute of General Medical Sciences grant 5 T32 Gm 07616. We thank Dr Michael Hunkapiller and Ms Carol Oken for assistance in the sequence analysis of the J588 protein, Alex Shaffer for assistance with making hybridomas, Dr Daniel Hansburg for defining the dextran idiotypes and for helping in the early stages of this study, Philip Early for pointing out the possiblity of a D gene segment and Drs Douglas Fambrough and John Campbell for comments on the manuscript.

Received 27 July; accepted 31 October 1979.

1. Dreyer, W. J., Gray, W. R. & Hood, L. *Cold Spring Harb. Symp. quant. Biol.* **32**, 353–367 (1967).
2. Gally, J. A. & Edelman, G. M. *A. Rev. Genet.* **6**, 1–46 (1972).
3. Cohn, M. in *The Immune System, Genes, Receptors, Signals* (eds Sercarz, E. E., Williamson, A. R. & Fox, C. F.) 89–118 (Academic, New York, 1974).
4. Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 817–836 (1976).
5. Hood, L., Campbell, J. Y. & Elgin, S. C. R. *A. Rev. Genet.* **9**, 305–353 (1975).
6. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
7. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3881–3885 (1978).
8. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 857–861 (1979).
9. Smith, G. P., Hood, L. & Fitch, W. *A. Rev. Biochem.* **40**, 969–1012 (1971).
10. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11–17 (1978).
11. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
12. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133–1144 (1978).
13. Wu, T. & Kabat, E. *J. exp. Med.* **132**, 211–250 (1970).
14. Capra, J. D. & Kehoe, J. M. *Proc. natn. Acad. Sci. U.S.A.* **71**, 845–849 (1974).
15. Amzel, L., Poljak, R., Saul, F., Varga, J. & Richard, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1427–1430 (1974).
16. Padlan, E. A., Segal, D. M., Rudikoff, S., Potter, M., Spande, T. & Davies, D. R. *Nature, new Biol.* **245**, 165–167 (1973).
17. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276**, 785–790 (1978).
18. Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Nature* **228**, 1045–1047 (1970).
19. Appella, E. *Proc. natn. Acad. Sci. U.S.A.* **68**, 590–594 (1971).
20. Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **61**, 837–846 (1977).
21. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
22. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
23. Weigert, M. *et al. Nature* (in the press).
24. Köhler, G. & Milstein, C. *Nature* **256**, 495–497 (1975); *Eur. J. Immun.* **6**, 511–519 (1976).
25. Littlefield, J. N. *Science* **145**, 709–710 (1964).
26. Barstad, P. *et al. Eur. J. Immun.* **8**, 497–503 (1978).
27. Schilling, J., Hansburg, D., Davie, J. M. & Hood, L. *J. Immun.* **123**, 384–388 (1979).
28. Hansburg, D., Perlmutter, R. M., Briles, D. E. & Davie, J. M. *Eur. J. Immun.* **8**, 352–359 (1978).
29. Hansburg, D., Briles, D. E. & Davie, J. M. *J. Immun.* **117**, 569–575 (1976); **119**, 1406–1412 (1977).
30. Kehry, M., Sibley, C., Fuhrman, J., Schilling, J. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2932–2936 (1979).
31. Dayhoff, M. O. *Atlas of Protein Sequence and Structure 1972* (National Biomedical Research Foundation, Maryland, 1972).
32. Sokano, H. *et al. Nature* **277**, 627–633 (1979).
33. Rao, D. N., Rudikoff, S., Krutzach, H. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2890–2894 (1979).
34. Leon, M. A., Young, N. M. & McIntire, K. R. *Biochemistry* **9**, 1023–1030 (1970).
35. Lundblad, A. *et al. Immunochemistry* **9**, 535–544 (1972).
36. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Variable Regions of Immunoglobulin Chains* (Bolt, Beranek & Newman, Cambridge, Mass., 1976); *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429–2433 (1978).
37. Capra, J. D. & Kindt, T. *Immunogenetics* **1**, 417–427 (1975).
38. Kindt, J. G. & Capra, J. D. *Immunogenetics* **6**, 309–321 (1978).
39. Edelman, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **63**, 78–85 (1969).
40. Hood, L., Huang, H. & Dreyer, W. J. *J. supremolec. Struct.* **7**, 531–559 (1977).
41. Galfre, G., Howe, S. C., Milstein, C., Butcher, G. W. & Howard, J. C. *Nature* **266**, 550–552 (1977).
42. Coffino, P., Baumel, R., Laskov, R. & Scharff, M. D. *J. cell. Physiol.* **79**, 429–440 (1972).
43. Hubert, J., Johnson, N., Barstad, P. & Hood, L. (in preparation).
44. Rudikoff, S. & Potter, M. *Biochemistry* **13**, 4033–4038 (1974).
45. Rudikoff, S. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2109–2112 (1976).
46. Vrana, M., Rudikoff, S. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1957–1961 (1978).
47. Francis, S. H., Leslie, R. G. Q., Hood, L. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1123–1127 (1974).
48. Milstein, C., Adetugbo, K., Cowan, N. J. & Secker, D. S. *Prog. Immun.* **2**, 157–168 (1974).
49. Bourgeois, A., Fougereau, M. & Depreval, C. *Eur. J. Biochem.* **24**, 446–455 (1972).

CHAPTER 5

This paper will be submitted to

The Journal of Experimental Medicine.

AMINO ACID SEQUENCE ANALYSES OF ANTIBODIES TO α-(1→3) DEXTRAN: MECHANISMS OF DIVERSIFICATION

James Schilling, Brian Clevinger, Martha Bond, Joseph Davie and Leroy Hood

Division of Biology
California Institute of Technology
Pasadena, California 91125
and
Department of Microbiology and Immunology
Washington University School of Medicine
St. Louis, Missouri
63110

The immunoglobulin molecule is coded for by three unlinked gene families: kappa (κ), lambda (λ), which encode the light chains, and heavy (H) which encodes the heavy chains (1,2). The kappa and lambda families each contain four types of gene segments: leader (L), variable (V), joining (J) and constant (C) (3-7). These are separated in the germline by intervening sequences. During the B-cell differentiation one V_L and one J_L gene segment are joined with apparent deletion of the intervening sequence to form a contiguous transcriptionally active variable region gene. In addition to L, V_H, J_H and C_H gene segments heavy chains possess a fifth gene segment, diversity (D), located between V_H and J_H (8). DNA rearrangements join one V_H and one J_H gene segment to a single D gene segment to form a transcriptionally active V_H region gene during B-cell differentiation. The V_H region gene remains separated from the C_H gene segments by an intervening DNA sequence. Both protein and nucleic acid sequence analysis of joined V_H region and germline V_H and J_H gene segments have demonstrated the existence of the D segment (8,9). However, D gene segment sequences have not yet been obtained from germline DNA.

The joined V_H and V_L genes encode the antigen binding portion of the antibody molecule (1). Each variable region is composed of four regions of relatively conserved amino acid sequence (framework regions) and three regions of highly variable sequence (hypervariable regions) (10,11). X-ray crystallographic studies have demonstrated that the hypervariable regions fold to form the walls of the antigen-binding site (12,13). Amino acid substitutions in these regions therefore may influence antigen binding.

Several mechanisms have been shown to contribute to antibody diversity. i) Multiple V_H, J_H, V_L, and J_L gene segments exist in the germline (2,5,8,15,16). ii) These gene segments can be joined in combinatorial fashion to form many different variable regions (9,14). Since part of the J segment is in the third hypervariable region, this combinatorial joining can contribute significantly to antigen-binding diversity in both heavy and light chains. iii) The joining of V_κ and J_κ can produce additional diversity at the N-terminal portion of the J segment. This type of diversity has been termed junctional diversity and has been explained by a site-specific recombination model (21-23). iv) Simple somatic mutation of both V and J segments also can potentially produce variant sequences which may be selected by virtue of their higher antigen-binding affinity. For example, the majority of myeloma λ chains which have been sequenced (18-20) exhibit a V_λ region sequence identical to that of a germline V_λ gene segment (7). Several variant V_λ region sequences have been found which differ from the germline sequence by one to three amino acid substitutions, all located in hypervariable regions. These sequences are consistent with the hypothesis that the variant V_λ sequences were produced by somatic mutation of a single germline V_λ gene segment.

The anti-1,3-dextran (Dex) antibodies of mice have been well studied genetically, serologically, and structurally. Two BALB/c myeloma tumors which bind dextran, M104E and J558, have been used as prototype dextran antibodies. Both of these proteins have λ light chains with identical sequences (18,19). Mouse strains can be classified as either high or low responders to dextran based on the rapidity and magnitude of their antibody response to dextran immunization

(21,22). The high/low response trait, as well as an idiotypic determinant common to M104E and J558 but not other immunoglobulins which do not bind dextran have been shown by genetic analysis to map adjacent to the heavy chain constant region locus (23). Antibodies to dextran obtained from high responder mouse serum following dextran immunization invariably carry λ light chains which comigrate on isoelectric focusing with those of M104E and J558, suggesting that there is little or no contribution to dextran antibody diversity from the light chain (24). In addition to the common idiotypic determinant described above, termed IdX, individual idiotypic determinants unique to M104E (IdIM104) and J558 (IdIJ558) have been described (25,26). Antisera directed against these determinants have demonstrated the presence of all three determinants in serum antibodies. An idiotype negative class of serum antibody which lacks all these three idiotypic determinants also is present in serum antibodies. Amino acid sequence studies of pools of serum antibodies which express either very high levels of IdIM104 and IdX or low levels of IdX demonstrate a single homogeneous heavy chain N-terminal sequence identical to that of M104E and J558 (24).

The availability of hybridomas which secrete large quantities of homogeneous dextran-binding immunoglobulin greatly expedited protein sequence and idiotype analyses. We have previously reported complete heavy chain variable region sequences for two myeloma and ten hybridoma proteins (9). From those sequences came the first evidence that the heavy chain V region is encoded by three separate gene segments: V_H, D, and J_H. The gene segments of the V_H region have subsequently been confirmed by nucleic acid analysis (8). The twelve V_H regions were

constructed from four V_H, nine D, and three J_H segments. Combinatorial joining of these segments contributes significantly to dextran antibody diversity. In this paper we present complete V_H region sequences from ten additional dextran-binding hybridomas. These molecules contain V_H, D, and J_H segments not previously found in dextran antibodies. The new sequences serve to strengthen and extend our previous conclusions regarding the origin of diversity of dextran antibodies. Idiotypic studies of the hybridomas had previously allowed identification of specific sites as contributing to idiotypic determinants (8,27). Further idiotypic analyses are presented in the accompanying paper by Clevinger et al. (manuscript in preparation).

MATERIALS AND METHODS

IgM Purification and Peptide Preparation. Kehry et al. (28) present complete details of the purification of the M104E IgM immunoglobulin, H and L chain separation, and cyanogen bromide (CNBr) digestion and peptide separation. The same procedures have been used for all IgM class dextran antibodies presented here.

Sequence Strategy of IgM Antibodies. A detailed report on the sequence analysis of M104E will be presented elsewhere (28). The sequence analysis of IgM class hybridomas V_H regions generally followed the strategy described there for the V_H region and summarized here with specific exceptions to the general strategy noted.

Residues 1-35. The intact H chain was sequenced from the N-terminus for 35 steps. If necessary, tryptic peptides containing residues 20-23 and 24-34 were isolated from cyanogen bromide peptide

(CN 1-2; see ref. 28) containing residues 1-34 and sequenced.

Residues 35-81. The cyanogen bromide peptide containing residues 35-81 (CN3) was sequenced from its N-terminus to residue 68 or further. Two tryptic peptides containing residues 68-74 and 75-81 were isolated and sequenced. Compositions were obtained for all tryptic peptides from CN3 to confirm the sequence. In some cases CN3 was succinylated to block the N-terminus and digested with protease V8 from Staphylococcus aureus which yields a single specific cleavage following glutamic acid at residue 46. The mixture of one blocked and one unblocked peptide was loaded on the sequencer without further purification. One hybridoma heavy chain, Hdex8, lacks a methionine at residue 34. In this case a cyanogen bromide peptide containing residues 20-81 was sequenced to residue 69. Residues 68-71 were determined by sequence analysis of two tryptic peptides as above.

Residues 82-117. A cyanogen bromide peptide containing residues 82-151 (CN4) was sequenced from its N-terminus to at least residue 108. CN4 was succinylated to block α and ε amino groups and digested with trypsin which now cleaves only at arginine residues yielding a simple specific cleavage after residue 98. This mixture of two fragments, one block and one unblocked, was sequenced without further purification from residues 99 to 117. Thus the sequence of all residues in the third hypervariable region has been determined twice for every protein. Tryptic digestion of succinylated Hdex13 CN4 yields cleavages after residues 98 and 103. This resulted in a mixture of one blocked and two unblocked peptides being sequenced. The sequence of residues 99-103 was known from the N-terminal sequence of Hdex13 CN9 and thus could be subtracted from the mixed sequence obtained to provide an unambiguous

sequence for residues 104-117. Hdex14 heavy chain has a methionine at residue 104 not found in other dextran antibodies. Cyanogen bromide cleavage of this protein generates peptides containing residues 82-104 and 105-151. The former was completely sequenced and the latter was sequenced to residue 118.

Purification and Sequence Analysis of IgG3 antibodies. Ascites fluid containing IgG3 immunoglobulin was dialyzed against distilled water, resulting in the specific precipitation of the IgG3 protein. This was resuspended in phosphate buffered saline (0.12 M sodium chloride -- 0.10 M sodium phosphate pH 7.8) and mildly reduced under N_2 with 10 mM dithiothreitol (DTT) for 90 min at room temperature and alkylated with 23 mM iodoacetamide (IA) for 60 min on ice. Heavy and light chains were precipitated with an equal volume of saturated ammonium sulfate at 4°C followed by at 5,000 rpm for 10 min. The resulting pellet was resuspended in 3 M guanidine·HCl-0.12 M ammonium bicarbonate for chain separation by gel filtration on a 2.5 x 100 cm Ultrogel AcA54 (LKB) column equilibrated in the same buffer. Heavy chains were desalted over G-25 (Pharmacia) in 0.2 M ammonium hydroxide and lyophilized. Dry heavy chain was resuspended in 70% formic acid, digested with cyanogen bromide (50 mg/ml) for 18 h at 4°C, and lyophilized. The digest was resuspended in 8 M guanidine·HCl and subsequently diluted with 0.2 M ammonium bicarbonate to yield a 3 M guanidine·HCl solution which was applied to a 2.5 x 150 cm Ultrogel AcA 54 column equilibrated and eluted with 3 M guanidine-HCl-0.2 M ammonium bicarbonate. A peptide containing residues 35-81 and corresponding to CN3 from IgM molecules is recovered in a pure state,

desalted, and lyophilized. A second peptide found in the excluded volume of the column was desalted, lyophilized, resuspended in 6 M Gu·HCl - 0.2 M Tris base pH 8.2, fully reduced under N_2 with 12 mM DTT at 37°C for 90 min, alkylated with 25 mM IA on ice for 60 min, diluted with an equal volume of 0.2 M ammonium bicarbonate and applied to the same AcA54 column previously described. A peptide corresponding to CN4 of the IgM molecule with residue 82 as its N-terminus can be recovered from the column.

Sequence Strategy for IgG3 Antibodies. Residues 1-35 were determined by sequence analysis of intact heavy chain for 35 steps. Residues 35-81 were determined by complete sequence analysis of the corresponding cyanogen bromide peptide. Residues 82-117 were determined from a single 37-cycle sequence analysis of the corresponding peptide.

Sequence Analysis. Automated amino acid sequence analysis was performed on extensively modified Beckman 890B sequenators with Polybrene carrier (29,30). Phenylthiohydantoin amino acids were identified by high performance liquid chromatography as previously described (29,30,31). Amino acid analyses were performed with a Durrum D-500 amino acid analyzer with a ninhydrin detection system.

RESULTS

The complete variable region amino acid sequences of the twelve previously reported and nine new dextran-binding heavy chains are presented in Fig. 1. Fig. 2 presents the sequences of residues 98-117.

The heavy chain variable region is comprised of three segments: V_H (residues 1-99), D (residues 100-101) and J_H (residues 102-117). All of the 21 V_H segments have identical sequences from residue 81 to residue 99. However, eight different amino acids are found at residue 100, the most diverse position in these proteins. Proteins with different sequences for residues 1-99 can have identical sequences for residues 100 and 101 (e.g., M104E, Hdex8 and Hdex16; J558 and Hdex9). These data are consistent with residue 99 being the C terminus of the V_H segment and residue 100 the N terminus of the D segment. A phosphorylcholine-binding germline V_H gene segment has recently been sequenced and found to end with the residue homologous to residue 99 of dextran-binding H chains, supporting this conclusion (8). Similar arguments suggest that residue 102 is the N terminus of the J_H segment. Thus some proteins with identical sequences for residues 1-101 have different sequences for residues 102-117 (e.g., J558, Hdex 14 and Hdex18; Hdex1 and Hdex2) and some proteins with different sequences for residues 100 and 101 have identical sequences for residues 102-117 (e.g., M104E, J558, Hdex2, Hdex6, Hdex7, Hdex10, Hdex12, Hdex17). Furthermore, when all known sequences of BALB/c H chains are compared (Figure 3) proteins with very different V_H and D segments are found to have identical J_H segments (e.g., M104E and T15; Hdex14, M21 and M173).

Each of the three segments of dextran-binding H chains exhibits a strikingly different pattern of diversity. Six different V_H segment sequences have been found. One, termed V_1 (see Figure 2), occurs in 15 of 21 heavy chains. The V_4 sequence differs from the V_1 sequence by two amino acid substitutions, each of which requires a single nucleotide

substitution and has been found twice in the 21 dextran-binding heavy chains. The V_2, V_3, V_4 and V_5 sequences differ from V_1 by seven, four, one, and one amino acid substitutions, respectively, and occur once each. All amino acid substitutions in these proteins require only single nucleotide substitutions. Twelve different D segment sequences have been found, making this the most diverse of the three dextran heavy chain segments. The D_1 and D_2 segments have been found four times each, D_3 and D_4 twice each, and the remaining light D segments only once each. Of the six different J_H segments found in dextran-binding heavy chains, J_1, J_2, J_4 and J_5 sequences differ from one another by multiple amino acid substitutions, a number of which require two nucleotide substitutions in a single codon. The J_1 and J_3 sequences are more closely related, differing by two amino acid substitutions which require five nucleotide substitutions. The J_2 and J_4 sequences differ only at their N-terminal residue and this amino acid substitution requires only a single nucleotide substitution.

Each hybridoma antibody has a λ light chain. These have been compared to the M104E and J558 λ chains by isoelectric focusing (data not shown). All of the λ chains examined focus at identical positions, suggesting that they are all very similar if not identical in sequence. A more detailed characterization of these proteins is now under way.

Discussion

Each of the 21 V_H regions from dextran-binding antibodies that we have examined to date has a unique amino acid sequence indicating the

existence of an enormous diversity of dextran-binding antibodies. We will discuss the diversity patterns seen for each of the three heavy chain gene segments with regard to possible mechanisms of the generation of diversity which these patterns suggest.

V_H Segments. Germline gene segments are defined by multiple appearances of identical amino acid sequences in independently derived myelomas and hybridomas. Two identical variant sequences arising by somatic mutation of the same germline gene segments requires two independent B-cell lineages to follow identical mutation pathways. Such parallel mutation is unlikely and therefore two identical variant sequences are generally considered to indicate the existence of a separate germline gene segment (32). By this argument, the V_1 and V_4 sequences represent two separate germline gene segments. The four sequences which have been found only once each might have arisen in any one of several ways. 1) Each of these sequences may represent a separate germline V_H gene segment. If so, the small extent of divergence of the sequences suggests that they have arisen from fairly recent gene duplication events. 2) These variants may arise by somatic mutation of the germline V_1 gene segment. If this is the case, then somatic mutation in V_H gene segments has somewhat different properties compared to the somatic mutational pattern of the V_λ gene segment. First, ten of thirteen amino acid substitutions occur outside of hypervariable regions in the dextran V_H segments, in contrast to V_λ where all substitutions are in hypervariable regions (18-20). Second, dextran V_H segments with four and seven amino acid substitutions have been found while in V_λ segments the maximum number of amino acid substitutions is three. If these variants do in fact arise from somatic

mutation, it is still necessary to ask whether they create an increased antigen-binding affinity and therefore are selected in vivo during an immune response. Third, variants may have arisen from unequal crossing over among tandemly arrayed dextran V_H gene segments which would produce new hybrid V_H gene segments. Such a process would be promoted by flanking sequence homology.

Complete nucleic acid sequence analysis of all germline dextran V_H gene segments should elucidate the relative contributions of each of these three mechanisms. If, for example, each of the six V_H amino acid sequences was found to correspond to a separate germline gene segment then somatic variational mechanisms would be unnecessary. If there are fewer germline V_H gene segments than protein sequences, a comparison of the germline gene segment sequences may indicate the most probable mechanism for producing each of the variants observed. Since the C terminal sequence of all dextran V_H sequences is invariant, junctional mechanisms of diversification do not appear to operate on the V_H segment at this boundary.

D Segments. Of the three V_H region gene segments the origin of diversity in the D segment is the most difficult to assess. The parallel mutation argument discussed above suggests that at least those four D segments (D_1, D_2, D_3 and D_4) which occur multiple times represent germline gene segments. It is possible that all twelve D segments found in dextran antibodies may represent germline gene segments. However, a sufficiently high mutation rate acting on six nucleotide-long D gene segments combined with strong selection for certain amino acids at residues 100 and 101 imposed by dextran binding could result in the generation of the same variant D segments in independent cell lineages (i.e., parallel mutation).

Both residues 100 and 101 are junctional residues, raising the possibility that junctional mechanisms may contribute to diversity at both of these positions. Several dextran D segment sequences can be found within the larger D segment sequence of other H chains. For example, both D_{10} and D_{12} are present in the T15 D segment, D_3 and D_{12} are present in the T15 D segment, D_3 and D_{12} in the H8 D segment, and D_3 and D_4 in the M21 D segment (Figure 3). This observation raises the intriguing possibility that there will be a relatively small number of germline D segments of relatively large size which will have appropriate flanking sequences to allow joining with V_H and J_H gene segments. If joining can occur at variable positions in the 3' and 5' ends of the D segment, then proteins with a large diversity of D segment sizes and sequences will result. As seen in Figure 3, D segments do exhibit extensive size and sequence diversity. The data of Figure 3 suggest that selection for binding to a specific antigen limits D segments to a particular length. Thus dextran antibodies have D segments two residues long; galactan antibodies have D segments of three residues; and levan antibodies appear to lack D segments entirely.

Nucleic acid sequence analyses of germline D gene segments will be required for a complete understanding of the origin of D segment diversity. Our laboratory is presently synthesizing a D segment DNA probe for this purpose.

We have previously pointed out the conservation of nucleotide sequence at codon 101 so that the second nucleotide of codon 101 is always adenine and the third nucleotide is always a pyrimidine (9). We have now found a single exception to this rule in D_{12} of Hdex15, which has a serine at

position 101 (mRNA codons UCX or AGY). The significance of this generally conserved sequence is unknown at present. It is frequently but not invariably observed in the D segments of other heavy chains (Figure 3). If the conservation does reflect some requirement of the joining process it serves to limit the junctional diversification of residue 101.

J_H Segments. Three of the dextran J_H sequences (J_1, J_2, J_6) occur multiple times in dextran and other BALB/c antibodies (Figures 2 and 3), suggesting that they represent germline gene segments. Indeed, DNA sequence analysis has demonstrated that the J_1 and J_2 gene segments are germline DNA (8). The J_3 sequence differs from J_1 by two amino acid substitutions requiring five nucleotide substitutions. While this could be produced by somatic mutation, it seems more probable that J_3 also is a germline gene segment. The J_4 segment differs from J_2 by only the N-terminal residue. A second hybridoma derived independently from the same mouse as Hdexll has a heavy chain variable region sequence identical with Hdexll, making it unlikely that the variation in this protein occurred in the cell fusion and cloning process. It is interesting to note that when the J_H segments of all BALB/c heavy chains are compared (Figure 3), there are four J_2-like segments which differ from one another only at their N-terminal residues (Figure 4). It is possible that each of these represents a separate germline J_H gene segment. It should be noted that there are two germline J_κ gene segments which differ only at their N-terminal residues (15, 16). A J_H segment which differs from J_1 by only the N-terminal residue also has been found in two different proteins (Figure 4). The location of these differences at the D-J_H boundary also raises the possibility that they may be generated by a junctional mechanism analogous

to that observed in kappa chains. When the J_5 sequence is compared with the J_H sequence of levan-binding antibodies A4, E109, and A47N (Figure 4), they are found to differ in length, with the dextran J_5 sequence having two additional residues at its N terminus. There are two possible explanations for this length difference: i) The germline J_5 gene segment begins with residue 102 and residues 102 and 103 have been deleted by the V_H-J_H joining event in levan antibodies. Deletions and insertions of single nucleotides have been observed at the V_K-J_K junction (20). ii) The J_5 gene segment begins with residue 104. In this instance, residues 102 and 103 would arise from a four-residue-long D segment in dextran antibodies. This also suggests that variation at residue 104 in galactan antibodies (Figure 3) could arise from junctional mechanisms. It should be remembered that insertions and deletions of a single residue at the V_κ-J_κ junction have been observed (20). Complete DNA sequence analyses of the germline J_H gene segments should resolve the questions raised here.

As discussed above, the protein sequence data suggest that residue 102 is the N-terminus of all six dextran J_H segments. DNA sequence analysis suggests that the J_1 gene segment encodes residue 101 as well while the J_2 gene segment begins with the codon for residue 103 (8). Thus it may be that in different proteins, codons for residues 101 and 102 will be derived either from germline D or germline J_H gene segments. D segments of dextran antibodies, therefore, may not be of a unique length.

Combinatorial joining of dextran V_H, D, and J_H gene segment contributes significantly to dextran antibody diversity. Combinatorial joining of six V_H, twelve D, and six J_H segments produces a potential population

of 432 different dextran antibodies. Our failure to find a repeated V_H region sequence in the 21 antibodies examined suggests that the pool of dextran antibodies in the BALB/c mouse is indeed very diverse. As data have accumulated, we continue to find new dextran V_H, D, and J_H segments, suggesting that we have not yet determined the full diversity existing for any of these segments.

Of the three dextran variable region gene segments the D segment diversity makes the greatest contribution to antibody diversity. If dextran antibodies contained no D segment, combinatorial joining of six V_H and six J_H gene segments would produce only 36 dextran antibodies. Multiplicity of germline V_H and J_H gene segments, potential somatic mutation of V_H gene segments and potential junctional diversity produced at the N-terminus of J_H gene segments all make lesser contributions to dextran antibody diversity. If all dextran antibodies do in fact possess identical λ chains, then combinatorial association of heavy and light chains does not contribute to the diversity of these antibodies.

The J_1 gene segment is found in 60% of all dextran antibodies examined. If all possible combinations of dextran V_H, D, and J_H gene segments initially occurred in differentiating B-cells with the same frequency, then the predominance of J_1 is presumably due to selection for the ability to bind dextran. However, dextran binding is clearly compatible with five other J_H sequences. Studies of antigen binding by M104E and J558 (33, 34) which differ only in their D segments, demonstrated different antigen-binding properties for these two proteins and suggested a role for the third hypervariable region in dextran binding. However, dextran antibodies can differ by six of eight third hypervariable

region residues (e.g., M104E and Hdex14). Further studies of dextran affinity and antigen binding by these antibodies may elucidate a role for D and J_H segments in determining the fine specificity of antigen binding.

The diversity of IgM class and two IgG3 class dextran antibodies (Hdex18 and Hdex19) can be compared. We find V_H, D, and J_H segments in these antibodies which also are found in the IgM class antibodies. We have suggested above that all of the segments observed in these IgG3 molecules are germline in origin. This is consistent with the current model of B-cell differentiation in which V_H and D segments are translocated and joined to a J_H segment located 5' to the C_μ gene segment. During the IgM to IgG class switch the complete V_H coding segment is translocated to a site adjacent to the C gene segment without further alteration of the V_H region.

Summary

The V_H region of the dextran-binding antibody molecule is assembled from three separate gene segments: V_H, D, and J_H. We have analyzed the possible sources of diversity in each of these three gene segments. Multiple germline genes certainly exist for the J_H and probably also the V_H and D gene segments. Diversity of V_H gene segments also may arise by either ordinary somatic mutation or by homologous unequal crossing over. Whether diversification of D and J_H segments may arise during joining from special junctional mechanisms will be known only when complete DNA sequences for each of the three gene segments are available for comparison with the protein sequences. Combinatorial joining of V_H, D, and J_H segments contributes significantly to diversity as demonstrated by our failure to find a repeated V_H region in 21 dextran antibodies examined.

References

1. Gally, J. A., and G. M. Edelman. 1972. The genetic control of immunoglobulin synthesis. Ann. Rev. Genet. 6: 1-46.

2. Hood, L., J. Campbell, and S. C. R. Elgin. 1975. Ann. Rev. Genet. 9: 305-353.

3. Brack, C., M. Hirama, R. Lenhard-Scheller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. Cell 15: 1-14.

4. Seidman, J. G., A. Leder, M. H. Edgell, F. Polsky, S. M. Tilghman, D. C. Tiemeier, and P. Leder. 1978. Multiple related immunoglobulin variable region gene identified by cloning and sequence analysis. Proc. Natl. Acad. Sci. USA. 75: 3881-3885.

5. Seidman, J. G., A. Leder, M. Nau, B. Norman, and P. Leder. 1978. Antibody diversity: The structure of cloned immunoglobulin genes suggests a mechanism for generating new sequences. Science 202: 11-17.

6. Tonegawa, S., A. M. Maxam, O. R. Tizard, O. Bernard, and W. Gilbert. 1978. Sequence of a mouse germline gene for a variable region of an immunoglobulin light chain. Proc. Natl. Acad. Sci. USA 75: 1485-1489.

7. Bernard, O., N. Hozumi, and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell 15: 1133-1144.

8. Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. Cell 19: 981-992.

9. Schilling, J., B. Clevinger, J. Davie, and L. Hood. 1980. Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. Nature 283, 35-40.

10. Wu, T. T., and E. A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. 132: 211-250.

11. Capra, J. D., and J. M. Kehoe. 1974. Variable region sequences of five human immunoglobulin heavy chains of the V_H III subgroup: Definitive identification of four heavy chain hypervariable regions. Proc. Natl. Acad. Sci. USA 71: 845-849.

12. Amzel, L. M., R. J. Poljak, F. Saul, J. M. Varga, and F. F. Richard. 1974. The three-dimensional structure of a combining region-ligand complex of immunoglobulin NEW at 3.5 Å resoltuion. Proc. Natl. Acad. Sci. USA 71: 1427-1430.

13. Padlan, E. A., D. M. Segal, T. F. Spande, D. R. Davies, S. Rudikoff, and M. Potter. 1973. Structure at 4.5 Å resoltuion of a phosphoryl-choline-binding Fab. Nature, New Biol. 245: 165-167.

14. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. Nature 276: 785-790.

15. Max, E. E., J. G. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin K constant region gene. Proc. Nat. Acad. Sci. USA 76: 3450-3454.

16. Sokano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature 280, 288-294.

17. Weigert, M., R. Perry, D. Kelley, T. Hunkapiller, J. Schilling, and L. Hood. 1980. The joining of V and J gene segments creates antibody diversity. Nature 283: 497-499.

18. Weigert, M. G., I. M. Cesari, S. J. Yonkovich, and M. Cohn. 1970. Variability in the lambda light chain sequences of mouse antibody. Nature 228: 1045-1047.

19. Apella, E. 1971. Amino acid sequence of two mouse immunoglobulin lambda chains. Proc. Natl. Acad. Sci. USA 68: 590-594.

20. Weigert, M., and R. Riblet. 1977. Genetic control of antibody variable regions. Cold Spring Harb. Symp. Quant. Biol. 61: 837-846.

21. Blomberg, B., W. R. Geckeler, and M. Weigert. 1972. Genetics of the antibody response to dextran in mice. Science 177: 178-180.

22. Riblet, R., B. Blomberg, M. Weigert, R. Lieberman, B. A. Taylor, and M. Potter. 1975. Genetics of mouse antibodies. 1. Linkage of the dextran response locus, V_H-DEX to allotype. Eur. J. Immunol. 5: 775-777.

23. Geckeler, W., B. Blomberg, C. de Preval, and M. Cohn. 1976. On the genetic dissection of a specific humoral immune response to $\alpha(1,3)$ dextran. Cold Spring Harb. Symp. Quant. Biol. 41: 743-748.

24. Schilling, J., D. Hansburg, J. M. Davie, and L. Hood. 1979. Analysis of the diversity of murine antibodies to dextran B1355: N-terminal amino acid sequences of heavy chains from serum antibody. J. Immunol. 123: 384-388.

25. Hansburg, D., D. E. Briles, and J. M. Davie. 1976. Analysis of the diversity of murine antibodies to dextran B1355: I. Generation of a large, pauci-clonal response by a bacterial vaccine. J. Immunol. 117: 569-575.

26. Hansburg, D., D. E. Briles, and J. M. Davie. 1977. Analysis of the diversity of murine antibodies to dextran B1355: II. Demonstration of multiple idiotypes with variable expression in several strains. J. Immunol. 119: 1406-1412.

27. Clevinger, B., J. Schilling, L. Hood, and J. M. Davie. 1980. Structural correlates of cross-reactive and individual idiotypic determinants on murine antibodies to α(1-3) dextran. J. Exp. Med. 151: 1059-1070.

28. Kehry, M. R., J. S. Fuhrman, J. Schilling, J. Rogers, T. Hunkapiller, W. Grimes, C. H. Sibley, and L. E. Hood. Submitted. Complete amino acid sequence of a mouse mu chain: Homology among heavy chain constant region domains. Biochemistry.

29. Hunkapiller, M. W., and L. E. Hood. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (Polybrene), and high pressure liquid chromatography. Biochemistry 17: 2124-2133.

30. Hunkapiller, M. W., and L. E. Hood. 1980. New protein sequenator with increased sensitivity. Science 207: 523-525.

31. Johnson, N. D., M. W. Hunkapiller, and L. E. Hood. 1979. Analysis of phenylthiohydantoin amino acids by high-performance liquid chromatography on Du Pont Zorbax cyanopropylsilane columns. Anal. Biochem. 100: 335-338.

32. Hood, L., E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg, and J. Schilling. 1977. The structure and genetics of mouse immunoglobulins: An analysis of NZB myeloma proteins and sets of BALB/c myeloma proteins binding particular haptens. Cold

Spring Harb. Symp. Quant. Biol. 41: 817-836.

33. Leon, M. A., N. M. Young, and K. R. McIntire. 1970. Immunochemical studies of the reaction between a mouse myeloma macroglobulin and dextrans. Biochemistry 9: 1023-1030.

34. Lundblad, A., R. Steller, E. A. Kabat, J. W. Hirst, M. G. Weigert, and M. Cohn. 1972. Immunochemical studies on mouse myeloma proteins with specificity for dextran as for levan. Immunochemistry 9: 535-544.

35. Dayhoff, M. O. 1972. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, MD.

36. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1979. Sequences of Immunoglobulin Chains. National Institutes of Health, Bethesda, MD.

37. Kehry, M., C. Sibley, J. Fuhrman, J. Schilling, and L. E. Hood. 1979. Amino acid sequence of a mouse immunoglobulin μ chain. Proc. Natl. Acad. Sci. USA 76: 2932- 2936.

38. Hubert, J., N. Johnson, P. Barstad, and L. Hood. In preparation.

39. Rudikoff, S., and M. Potter. 1974. Variable region sequence of the heavy chain from a phosphorylcholine binding myeloma protein. Biochemistry 13: 4033-4038.

40. Rudikoff, S., and M. Potter. 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. Proc. Natl. Acad. Sci. USA 73: 2109-2112

41 Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from insulin-binding myeloma proteins. Proc. Natl. Acad. Sci. USA 75: 1957-1961.

42. Francis, S. H., R. G. Q. Leslie, L. Hood, and H. N. Eisen. 1974. Amino

acid analysis of the variable region of the heavy (alpha) chain of a mouse myeloma protein with anti-hapten activity. Proc. Natl. Acad. Sci. USA 71: 1123-1127.

43. Milstein, C., K. Adetgubo, N. J. Cowan, and D. S. Secker. 1974. Clonal variants of myeloma cells. Prog. Immun. 2: 157-168.

44. Bourgeois, A., M. Fougereau, and C. de Preval. 1972. Sequence of amino acids of the NH_2-terminal region of a mouse-clonal immunoglobulin heavy chain. Eur. J. Biochem. 24: 446-455.

45. Rao, D. N., S. Rudikoff, H. Krutzach, and M. Potter. 1979. Proc. Natl. Acad. Sci. USA 76: 2890-2894.

FIG. 1. Complete V_H region sequences of 21 independently derived dextran binding myelomas and hybridomas. The one-letter code of Dayhoff (35) is used to indicate differences with the M104E sequence, which is published in 37. Kabat-Wu numbering refers to the numbering scheme of references 10 and 36. hv 1,2,3 indicate hypervariable regions 1, 2, and 3. The locations of V_H, D, and J_H segments are indicated at the bottom along with the residue position (sequential numbering) of all amino acid substitutions in the V_H segments. Both Hdex9 and Hdex13 have substitutions at position 63. CHO indicates a carbohydrate moiety attached to asparagine at position 55 in all but Hdex10 and Hdex16. M104E and Hdex1 - Hdex17 are IgM class, Hdex17 and Hdex18 are IgG3, and J558 is IgA. All have λ light chains.

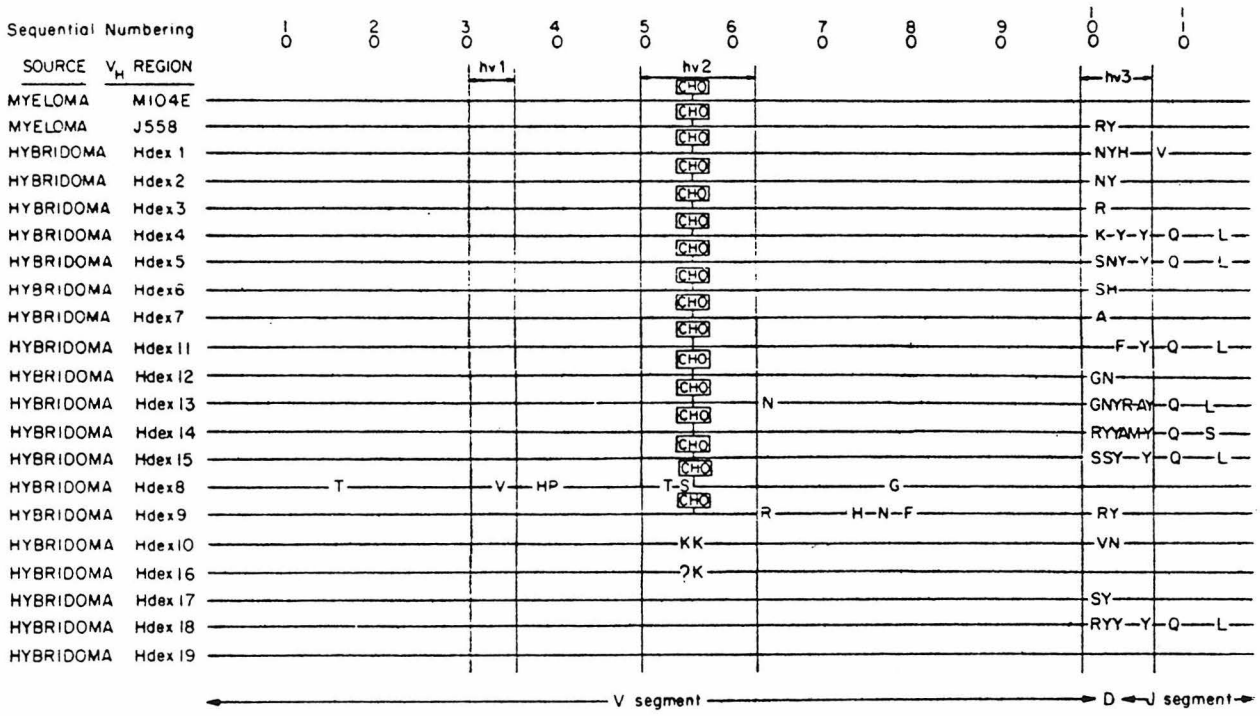

Sequential Numbering
SOURCE
V_H REGION
hv1
hv2
hv3
MYELOMA M104E
MYELOMA J558
HYBRIDOMA Hdex 1
HYBRIDOMA Hdex2
HYBRIDOMA Hdex3
HYBRIDOMA Hdex4
HYBRIDOMA Hdex5
HYBRIDOMA Hdex6
HYBRIDOMA Hdex7
HYBRIDOMA Hdex 11
HYBRIDOMA Hdex 12
HYBRIDOMA Hdex 13
HYBRIDOMA Hdex 14
HYBRIDOMA Hdex 15
HYBRIDOMA Hdex8
HYBRIDOMA Hdex9
HYBRIDOMA Hdex10
HYBRIDOMA Hdex 16
HYBRIDOMA Hdex 17
HYBRIDOMA Hdex 18
HYBRIDOMA Hdex 19
CHO
RY
NYH—V
NY
R
K-Y-Y—Q—L
SNY—Y—Q—L
SH
A
F—Y—Q—L
GN
GNYR-AY—Q—L
RYYAMY—Q—S
SSY—Y—Q—L
T
V—HP
T-S
G
R—H-N-F
RY
KK
VN
?K
SY
RYY—Y—Q—L
V segment
D
J segment

Figure 1

FIG. 2. The C-terminal portions of dextran binding heavy chain variable regions. The V_1-V_6, D_1-D_{12}, and J_1-J_6 segments are defined here. See legend to Figure 1 for other terms.

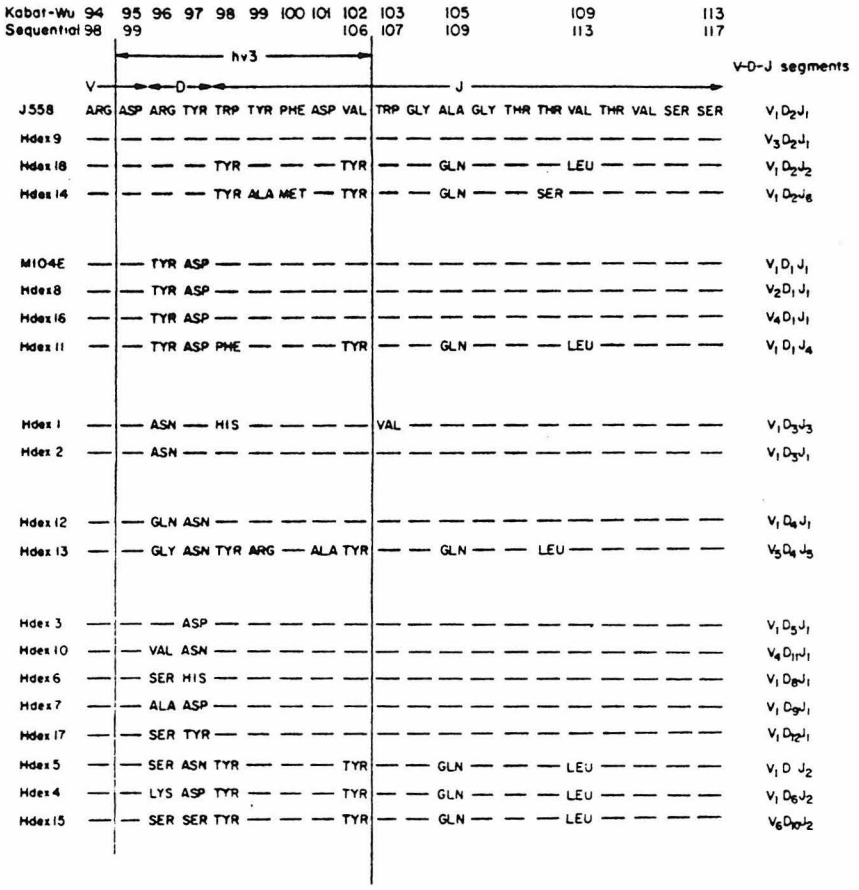

Kabat-Wu 94 95 96 97 98 99 100 101 102 103 105 109 113
Sequential 98 99 106 107 109 113 117
hv3
V-D-J segments
V
D
J
J558 ARG ASP ARG TYR TRP TYR PHE ASP VAL TRP GLY ALA GLY THR THR VAL THR VAL SER SER V1D2J1
Hdex 9 V3D2J1
Hdex 18 TYR TYR TRP GLN LEU V1D2J2
Hdex 14 TYR TYR ALA MET TYR GLN SER V1D2J6
M104E TYR ASP V1D1J1
Hdex 8 TYR ASP V2D1J1
Hdex 16 TYR ASP V4D1J1
Hdex 11 TYR ASP PHE TYR GLN LEU V1D1J4
Hdex 1 ASN HIS VAL V1D3J3
Hdex 2 ASN V1D3J1
Hdex 12 GLN ASN V1D4J1
Hdex 13 GLY ASN TYR ARG ALA TYR GLN LEU V5D4J3
Hdex 3 ASP V1D5J1
Hdex 10 VAL ASN V4D11J1
Hdex 6 SER HIS V1D8J1
Hdex 7 ALA ASP V1D9J1
Hdex 17 SER TYR V1D12J1
Hdex 5 SER ASN TYR TYR GLN LEU V1D J2
Hdex 4 LYS ASP TYR TYR GLN LEU V1D6J2
Hdex 15 SER SER TYR TYR GLN LEU V6D10J2

Figure 2

FIG. 3. A comparison of C-terminal portions of published BALB/c V_H regions. Each has been divided into V_H, D, and J_H segments by homology with dextran heavy chains. The prototype for V_H and J_H segments is M104E and for D segments the first sequence in each group. Those J_H segments which do not extend to amino acid 117 are due to incomplete sequence data and do not represent shorter J_H segments.

Protein	Specificity	V segment	D segment	J segment	References
MOU-E, Hdex16	1,3 Dextran	SSSTAYMQLNSLTSEDSAVYYCARD	YD	WYFDVWGAGTTVTVSS	37
Hdex8	"	---G----------	--	----------	9
J558	"	----------	RY	----------	"
Hdex9	"	--N--F----------	RY	----------	"
Hdex10	"	----------	VN	----------	"
Hdex6	"	----------	SH	----------	"
Hdex3	"	----------	R-	----------	"
Hdex17	"	----------	SY	----------	This paper
Hdex7	"	----------	A-	----------	9
Hdex12	"	----------	GN	----------	"
Hdex2	"	----------	NY	----------	"
Hdex1	"	----------	NY	H----V----------	"
Hdex5	"	----------	SN	Y---Y--Q---L----	"
Hdex4	"	----------	K-	Y---Y--Q---L----	"
Hdex15	"	----------	SS	Y---Y--Q---L----	This paper
Hdex18	"	----------	RY	Y---Y--Q---L----	"
Hdex11	"	----------	--	F---Y--Q---L----	"
Hdex13	"	----------	GN	YR-AY--Q--L---	"
Hdex14	"	----------	RY	YAM-Y--Q--S-----	"
T15,S63,S107,Y5236	Phosphorylcholine	-Q-IL-L-M-A-RA--T-I------	YYGSSY	----------	38
H8	"	-Q-IL-L-M-A-RA--T-I------	---N--	----------	"
W3207	"	-Q-IL-F-M-A-RA--T-I-----N	--XYDL	--V----------	"
M603	"	-Q-IL-L-M-A-RA--T-I-----N	----T	----------	39
M167	"	-Q-VL-L-M-A-RA--T-T---T--	ADYGDSYF	G----------	40
M511	"	-Q-IL-L-M-A-RA--T-I------	GDYG-SY	----------	"
A4	2,1 Levan	-K-SV-L-M-N-RA--TGI---TTG	[ ]	[]-AY--Q--L---	41
E109	"	-K-SVFL-M-N-RA--TGIH--TTG	[ ]	[]-AY--Q--L---	"
U61	"	-K-SV-L-M-N-RA--TGI---TTG	[ ]	[]-AY--Q--L-P-	"
A47N	"	-K-SV-L-M-N-RA--T-I---STG	[ ]	[]-AY--Q--L---	"
M315	DNP	-ENQFFLK-D-V-[ ]T-T----G-	NDH	L---Y--Q---L---	42
M21	Unknown	PKN-LFL-MT--R---T-M-----H	GNYPW	YAM-Y--Q--S-----	43
M173	Unknown	AKN-L-L-MSKVR---T-L-----S	PY	YAM-Y--Q--S-----	44
T601	Galactan	AKN-L-L-MSKVR---T-L-----G	GYY	G----------	45
X44	"	AKN-L-L-MSKVR---T-L-----L	H--	G-AAY--Q--L----A	"
X24	"	AKN-L-L-MSKVR---T-L-----G	---	G---Y--Q---L----	"
J539	"	AKNSL-L-MSKVR---T-L-----L	H--	G-NAY--Q--L----A	"

Figure 3

FIG. 4. J_H segments which may exhibit junctional diversity. See Fig. 3 for references.

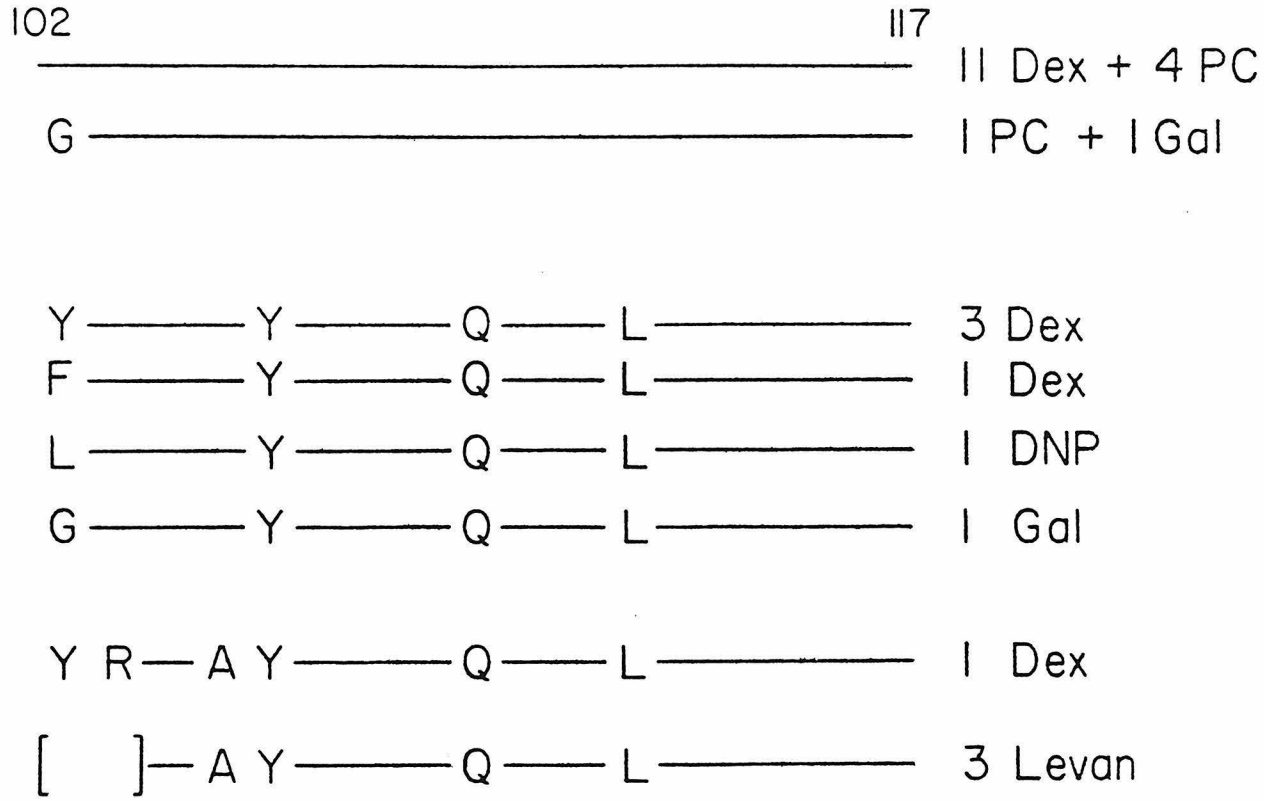
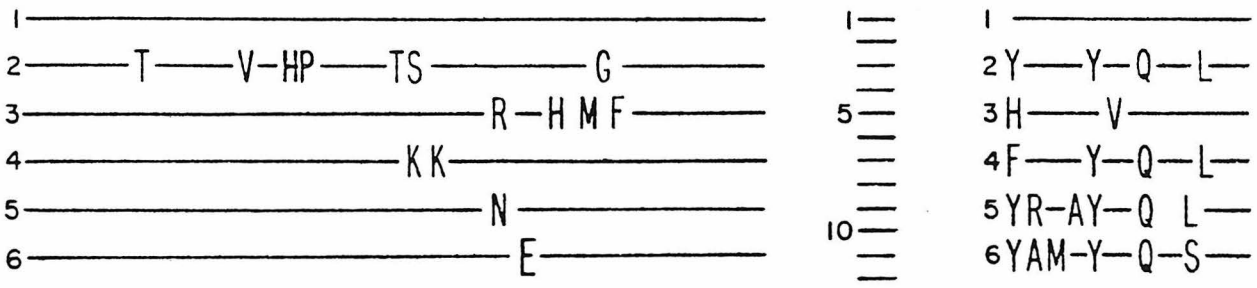

102
117
11 Dex + 4 PC
G
1 PC + 1 Gal
Y Y Q L
3 Dex
F Y Q L
1 Dex
L Y Q L
1 DNP
G Y Q L
1 Gal
Y R A Y Q L
1 Dex
[ ] A Y Q L
3 Levan

Figure 4

FIG. 5. Diversity of dextran heavy chains produced by combinatorial joining of V_H, D, and J_H gene segments.



Diversity of combinatorial joining of 24 gene segments = 6×12×6 = 432

Figure 5

CONCLUSION

During the past five years our understanding of the origin of antibody diversity has advanced substantially through a combination of amino acid and nucleic acid sequence analyses. The single most important feature of immunoglobulin genes discovered is the division of the light chain variable region into two and the heavy chain variable region into three germline gene segments. Multiple germline gene segments of each type exist for both heavy and light chains. The combinatorial joining of multiple separate gene segments clearly makes major contributions to the diversity of $V_{\kappa 21}$ light chains and dextran-binding heavy chains. Further diversification can occur as a consequence of gene segment rearrangement and joining in both systems.

Some V_H, V_κ, J_H and J_κ sequences are not easily explained as germline sequences or by junctional mechanisms of diversification of germline sequences. These variant sequences could have arisen by either somatic mutation or somatic recombination of germline gene segments. Somatic mutation is most attractive to explain variant sequences which differ from germline gene segments by one or two amino acid substitutions, while somatic recombination can best explain variants which differ from germline sequences by several amino acid substitutions. To reach a final conclusion with regard to the contribution of these mechanisms to diversity it will be necessary to know the nucleic acid sequences of all germline gene segments for a particular set of antibodies. These are now known for J_κ (1,2) and J_H (3,4) and are being determined for the V_κ gene segments of the $V_{\kappa 21}$ group by Tonegawa (M. Weigert, personal communication) and for dextran V_H gene segments by the Hood laboratory.

The complete germline J_H gene segment sequences were not known in time for inclusion in Chapter 5 but are presented here in Figure 1,

where they are compared with the J_H sequences of dextran heavy chains. The probable junctional residue in the codon for which recombination is presumed to take place during D-J_H rearrangement is shown at the N terminus of each J_H sequence. It is interesting that for M104 recombination apparently can occur between codons 101 and 102, suggesting a three nucleotide long D segment, while for Hdex 13 recombination can occur between the first and second nucleotides of codon 103, suggesting a D segment 10 nucleotides long. Other protein sequences are consistent with recombinants occurring at intermediate positions and imply D segments of intermediate length. Selection for the ability to bind dextran appears to require a third hypervariable region light amino acid residues long which can originate from a number of different combinations of D and J_H gene segments. A more complete evaluation of these ideas will be possible once the DNA sequences of germline D gene segments are known.

In addition to the DNA sequence analyses discussed above several other problems remain in the dextran heavy chains. One of the most remarkable features of this system is that no repeated sequences have been found in twenty-one variable regions examined. This strongly suggests that we have not yet discovered the full extent of diversity in the system. It would therefore seem reasonable to continue with variable region amino acid sequence analysis. The set of IdX negative molecules in particular has not been adequately characterized. As discussed in Chapter 4 and Appendix 1, we believe the IdX determinant to be located on the V_H segment. Thus the IdX negative population may contain new V_H segment sequences. A second problem which has not been adequately addressed as yet is the diversity of dextran lambda light chains. Preliminary isoelectric focusing data suggest that the

diversity of these chains is limited. Lambda chain variable regions are the best documented system in which somatic mutation seems to occur (5,6). A more detailed sequence determination would be useful to precisely define the diversity of lambda chains from antibodies of known specificity as well as reinforce our conclusions regarding the location of idiotypic determinants.

The dextran antibodies present a set of antibodies which differ one from another in small and defined ways. They present an excellent opportunity for a detailed characterization of antigen binding. It should be possible to examine the roles of V_H, D and J_H segments in antigen binding since there are numerous pairs of dextran antibodies which differ in only one of the three segments. The dextran antibodies themselves can be used as a set of closely related antigens for the production of anti-antibodies (anti-idiotypes). The diversity of antibodies which react with a well defined protein antigenic determinant, such as anti-idiotype antibodies which recognize IdX or IdI determinants, has not yet been carefully studied. Knowledge of the diversity of anti-idiotypes could have important implications for network theories of immune regulation (7).

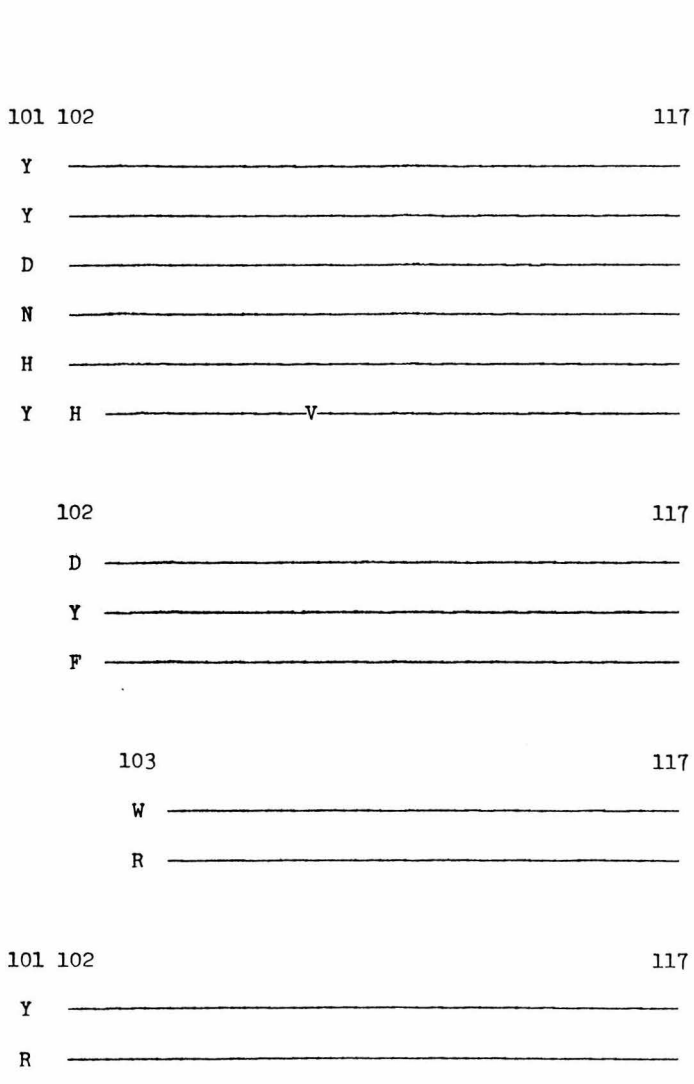
One of the major remaining unanswered questions in the study of immunoglobulin genes is whether there is any order to the rearrangement of these genes during ontogeny. In other words, do specific V_H-D-J_H combinations appear at precise and reproducible stages of development? This question can potentially be approached in the dextran antibody system utilizing anti-IdX and anti-IdI antibodies as probes for the presence of specific V_H and D segments as cell surface immunoglobulin on maturing B cells. Using appropriate combinations of dextran antibodies for immunization and subsequent absorption of serum it also

may be possible to raise J_H segment specific antibodies. One can hope for exciting new features of antibodies to be revealed in the future.

References

1. Max, E. E., J. G. Seidman and P. Leder (1979) Proc. Natl. Acad. Sci. USA 76: 3450.
2. Sakano, H., K. Hüppi, G. Heinrich and S. Tonegawa (1979) Nature 280: 288.
3. Early, P., H. Huang, M. Davis, K. Calame and L. Hood (1980) Cell 19: 981.
4. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder and S. Tonegawa (1980) Nature, submitted.
5. Weigert, M. and R. Riblet (1977) Cold Spring Harb. Symp. Quant. Biol. 41: 837.
6. Bernard, O., N. Hozumi and S. Tonegawa (1978) Cell 15: 1133.
7. Jerne, N. K. (1974) Ann. Immunol. (Inst. Pasteur) 125C: 373.

Figure 1. Comparison of germline DNA (3,4) and amino acid (Chapter 5) sequences for dextran J_H segments. The most probable junctional residue indicated by DNA sequence analysis is on the left of each J_H sequence. The source of each sequence is indicated. The amino acid sequence as identified in Chapter 5 to which each germline DNA sequence most closely corresponds is indicated. The 5'-most site at which recombination during D-J_H joining could occur to generate each amino acid sequence is indicated. Recombination could occur 3' to this site. 100a, 100b, 100c refer to the first, second and third nucleotides, respectively, in codon 100. The length of D segment implied by joining at the indicated sites is shown.



	Most probable recombination site	D segment length
Germline DNA (J_1)		
J558, Hdex 2, 9, 16, 17	100c-101a	3
M104, Hdex 3, 7, 8	101a-101b	4
Hdex 10, 12	101a-101b	4
Hdex 6	101a-101b	4
Hdex 1	102c-103a	9
Germline DNA (J_2)		
Hdex 4, 5, 15, 18	102a-102b	7
Hdex 11	102b-102c	8
Germline DNA (J_5)		
Hdex 13	103a-103b	10
Germline DNA (J_6)	100c-101a	3
Hdex 14		

Figure 1

APPENDIX 1

This paper was published in

The Journal of Experimental Medicine.

STRUCTURAL CORRELATES OF CROSS-REACTIVE AND INDIVIDUAL IDIOTYPIC DETERMINANTS ON MURINE ANTIBODIES TO α-(1 → 3) DEXTRAN*

By BRIAN CLEVINGER, JAMES SCHILLING, LEROY HOOD, and JOSEPH M. DAVIE

From the Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110; and the Division of Biology, California Institute of Technology, Pasadena, California 91125

Since their first description over 20 years ago, idiotypes have been used extensively to study various aspects of the immune system (1, 2). Idiotypes, antigenic determinants associated with immunoglobulin-variable regions, have been found on V_H and V_L regions, or both (3–11), and have been localized to both antigen-binding and framework regions (12–17). Those idiotypes shared by immunoglobulins known to be structurally different are designated IdX,[1] cross-reactive or public idiotypes, whereas those apparently restricted to one or a few closely related types of molecules are designated IdI, individual or private idiotypes.

In spite of these general descriptive associations, the molecular bases for idiotypes have not been precisely defined. This lack of molecular correlation stands in striking contrast to the precise structural correlations for the constant region allotypes of rabbit and human immunoglobulins where allotype-positive and -negative proteins differ by only one or a few amino acid residues. Closely related sets of structurally defined idiotype-positive and -negative proteins have not previously been described.

As part of an examination of the murine repertoire of anti-α-(1 → 3) dextran antibodies, a number of dextran-binding hybridoma proteins have been produced. A comparison of amino acid sequences (18) and idiotype expression of these proteins has allowed us to determine the molecular basis of anti-dextran idiotypes as a result of two features. First, the light chains from two dextran-binding myeloma proteins, MOPC104E and J558, are identical by amino acid sequence analysis. Furthermore, the light chains of these proteins and the 10 hybridoma proteins are identical by isoelectric focusing. If these 12 light chains are identical, idiotypic variability is encoded only in heavy chains. Second, some idiotype-positive and idiotype-negative heavy chains are different by only one or a few adjacent amino acids. Accordingly, precise correlations between IdX and IdI have been made with amino acid sequence variations in certain portions of the V_H regions.

---

* Supported by U. S. Public Health Service grants AI-11635, AI-10781, AI-15353, AI-05599, CA-09118, and GM-07616, and National Science Foundation grant PCM 76-09719.

[1] *Abbreviations and nomenclature used in this paper:* CFA, complete Freund's adjuvant; HGPRT, hypoxanthine guanine phosphoribosyl transferase; I_{50}, 50% inhibition; IdI, individual or private idiotype(s); IdX, cross-reactive or public idiotype(s); IEF, isoelectric focusing.

Materials and Methods

Mice. Male and female retired breeders of the strains BALB/cJ and BALB/c Cum were obtained from The Jackson Laboratory, Bar Harbor, Maine, and The Cumberland View Farms, Clinton, Tenn., respectively. (BAB-14 × BRVR)F_1 mice were bred in our animal facility (Washington University, St. Louis, Mo.). The BAB-14 females were gifts of Dr. L. A. Herzenberg (Stanford University, Palo Alto, Calif.). BRVR male mice were purchased from the Laboratory Animal Facility, State University of New York at Buffalo, Buffalo, N. Y.

Antigens and Immunizations. Primary anti-α-(1 → 3) dextran responses were generated by intraperitoneal injection of 100 μg B1355 dextran (Northern Regional Research Laboratory, Peoria, Ill.) in complete Freund's adjuvant (CFA). Hyperimmune responses were elicited by two monthly injections of 100 μg of B1355 in incomplete Freund's adjuvant followed 1 mo later by three intraperitoneal injections of 2×10^9 *Escherichia coli* B (Calbiochem.-Behring Corp., American Hoechst Corp., San Diego, Calif.) that had been heat killed at 90°C for 30 min before use (19).

Somatic Cell Hybridization. The procedure used was essentially that of Galfre et al. (20). Briefly, ~10^8 spleen cells from dextran-immunized mice were fused by polyethylene glycol (Carbowax 1,500 [1,500 mol wt], Fisher Scientific Co., Pittsburgh, Pa.) to at least 10^7 hypoxanthine guanine phosphoribosyl transferase (HGPRT)-deficient MPC-11 (γ_{2b} k) plasmacytoma cells, line 45.6T61.7 (21) or a nonsecreting variety of MOPC-21 (γ-l,k), line NS1/1-Ag4-1 (22). These cells were generously supplied by Dr. M. Scharff (Albert Einstein College of Medicine, New York) and by Dr. C. Milstein (Medical Research Council, Cambridge, Mass.). Cells were cultured in 24-well culture dishes in HAT selection medium (23). Dextran-binding proteins were detected in culture supernates either by isoelectric focusing or radioimmune assays. Hybrids were cloned in soft agar over 3T3 feeder layers (24) and grown in BALB/c mice as ascites tumors.

Idiotype Assays. The assays for three separate idiotype determinants, IdX, Idl(M104), and Idl(J558), have been described previously (25). Analyses were inhibition-type, solid-phase radioimmunoassays performed in microtiter plates (Cooke Laboratory Products, Div. Dynatech Laboratories, Inc., Dynatech Corp., Alexandria, Va.).

IdX. Goat anti-M104 was absorbed with T183 (μ-K), HOPC-1 ($\gamma_{2a}\lambda$), and normal serum by solid-phase and soluble absorption. The resulting antiserum, called anti-IdX, was used at a 1:20,000 dilution to bind ^{125}I-J558, a related dextran-binding protein with identical λ-chains.

Idl(M104). Rabbit anti-M104 was absorbed with T183, HOPC-1, normal serum, and J558. The resulting antiserum, called anti-Idl(M104), was used at a 1:100 dilution to bind ^{125}I-M104. It was necessary to include 10 μg of T183 and J558 in the assay to achieve the desired specificity.

Idl(J558). Rabbit anti-J558 was absorbed with M315(αλ), HOPC-1, and M104 to generate an antiserum called anti-Idl(J558). This reagent was used at a 1:2,000 dilution to bind ^{125}I-J558.

The relative concentration of these determinants in dextran-binding proteins was measured by calculating the amount of test protein needed for 50% inhibition (I_{50}) of the radiolabeled proband molecules from binding to the plastic-adsorbed anti-idiotype. This value, when divided into the I_{50} value for unlabeled proband, gives a measure of idiotype expression. A value of 1.0 denotes idiotype identity between test protein and the proband; <1.0 indicates nonidentity.

Results

Anti-α-(1 → 3) Dextran Antibodies. Table I summarizes the origins and characteristics of the anti-α-(1 → 3) dextran antibodies used in this study. The 10 hybrid-derived antibodies are all μλ-immunoglobulins and result from fusions of spleens from 9 individual animals. By isoelectric focusing, the λ-light chains of the hybrid proteins are indistinguishable from those of M104 and J558 (data not shown). The hybrid proteins that result from fusions with MPC-11 in particular have variable amounts of plasmacytoma-derived K chains in the IgM molecules, but the μ-K pairs do not bind dextran (26) and probably have little effect on idiotype determinations (see below).

TABLE I

Anti-α-(1 → 3) Dextran Antibodies

Protein	Origin	Hybrid partner*	Immunization‡	Class
M104	BALB/c	—	—	IgMλ
J558	BALB/c	—	—	IgAλ
Hdex 1	(BAB-14 × BRVR)F_1	MPC-11	dextran/*E. coli*	IgMλ
Hdex 2	BALB/c	M21	dextran/*E. coli*	IgMλ
Hdex 3	BALB/c	MPC-11	dextran/*E. coli*	IgMλ
Hdex 4	BALB/c	MPC-11	dextran/*E. coli*	IgMλ
Hdex 5	(BAB-14 × BRVR)F_1	MPC-11	dextran/*E. coli*	IgMλ
Hdex 6	BALB/c	M21	dextran/*E. coli*	IgMλ
Hdex 7	BALB/c	M21	dextran 1°	IgMλ
Hdex 8	BALB/c	M21	dextran 1°	IgMλ
Hdex 9	BALB/c	M21	dextran 1°	IgMλ
Hdex 10	BALB/c	MPC-11	dextran/*E. coli*	IgMλ

* HGPRT-deficient MPC-11 (γ-$_{2b}$K) or nonsecreting variant of MOPC-21 (γ-$_1$K) that synthesizes but does not secrete K chains.

‡ Hybrids Hdex 1 and 5 were derived from a single (BAB-14 × BRVR)F_1 mouse 7 d after hyperimmunization with two monthly injections of 100 μg of dextran B1355 in CFA followed 1 mo later by three intravenous injections of 2×10^9 *E. coli* at 2 d intervals. The other hybrids were derived from individual BALB/c mice: Hdex 2, 3, 4, 6, and 10 were obtained 1-3 d after the dextran/*E. coli* protocol, and Hdex 7, 8, and 9 were obtained 5-6 d after a single injection of 100 μg of dextran in CFA.

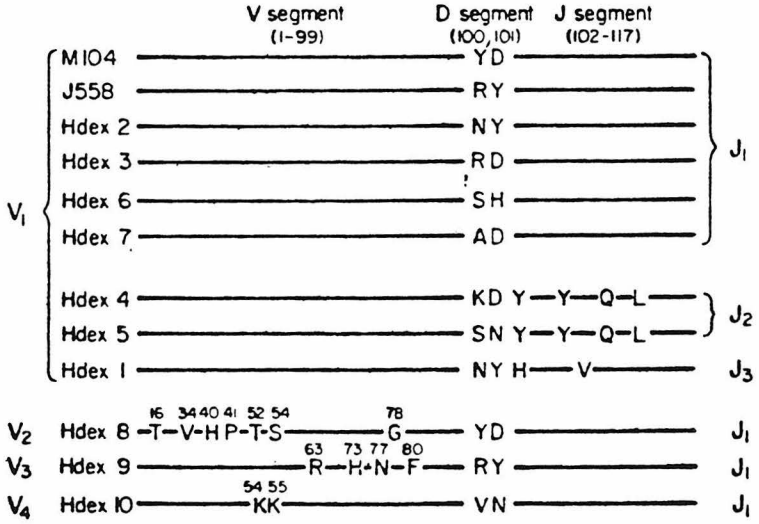


FIG. 1. Diversity patterns of V_H regions of α-(1 → 3) dextran-binding myeloma and hybridoma proteins. The complete V_H amino acid sequences of these 12 proteins are presented elsewhere (18, 31). Protein M104 is used for comparison; solid lines for the V segments (1-99) and J segments (102-117) denote identity with M104. Differences are indicated by the one-letter code of Dayhoff (30). The D-segment sequences (100-101) are given for each protein.

Amino Acid Sequence Patterns of Dextran-binding Proteins. Details of the total amino acid sequences of the V_H regions of M104, J558, and the 10 dextran-binding hybridoma proteins are presented elsewhere (18). Fig. 1 summarizes the major features of these sequences and shows that the diversity patterns divide the V_H regions into

three segments: the V segment (residues 1-99), the D segment (100 and 101), and the J segment (102-117). There are four different V segments that differ from one another at two to seven positions, and three J segments that differ from each other at from two to five positions. There are nine different pairs of amino acids in the D segment with three duplicated sequences. The V and J segments are encoded by separate gene segments analogous to the V- and J-gene segments coding for mouse light-chain-variable regions (27-29).[2]

The genetic basis for the D segment is unknown; however, it must arise from DNA not directly contiguous with either the V- or J-gene segments.[2] The fact that the D segment occurs at the junction of the V- and J-gene segments suggests that the D segment represents yet another gene segment coding for the classical variable region. Diversity in D-gene segments may arise by some process of somatic variation or it may be encoded by germ-line D-gene segments.

Idiotypic Characteristics of Dextran-binding Proteins. The dextran-binding proteins were analyzed for their expression of three idiotypes previously described by Hansburg et al. (25), and the results along with a summary of the sequence data are shown in Fig. 2 and Table II. The IdX, expressed equally by M104 and J558, is found on 11 of the 12 hybridoma proteins. One protein, Hdex 8, shows partial expression of IdX, whereas Hdex 10 is IdX negative. In addition to the IdX, other determinants have been described previously that were found on M104, but not J558 [IdI(M104)] and on J558, but not M104 [IdI(J558)]. These IdI determinants are also found among some of the hybridoma proteins. Two hybridoma proteins, Hdex 8 and Hdex 7, are equivalent to M104 in expressions of IdI(M104). One hybridoma protein, Hdex 9, expresses IdI(J558) fully, whereas two others, Hdex 1 and Hdex 2, show partial expression of this idiotype. This pattern of expression of IdX and IdI determinants among the monoclonal antibodies is consistent with that seen in serum anti-α-(1 $\rightarrow$ 3) dextran antibodies. The bulk of both 7S and 19S serum antibodies react with anti-IdX, whereas only a minority express either IdI(M104) or IdI(J558) (25).

It is unlikely that the existence of plasmacytoma-derived K chains in some of the hybridoma proteins seriously alters our estimates of idiotypic relatedness. Even if it should be found that μ-K pairs do not express a determinant normally expressed by $\mu\lambda$-pairs, the effect is minimized here by normalizing idiotype expression to λ-concentration.

Localization of the Amino Acids Important for IdX Expression. A search was made to correlate amino acid sequence with the IdX determinant. Table II shows that it is unlikely that the D segment is involved directly with IdX expression; the IdX-positive proteins have nine different D segments that contain residues with side chains that differ considerably in size, charge, and bonding abilities. Likewise, the J segments do not contain residues that determine IdX expression because all three J segment prototypes are associated with IdX-positive proteins and the IdX-negative protein contains J_1 as do several IdX-positive proteins. Thus, it is differences in the V segment that most likely determine IdX expression. The four prototype V segments and their IdX expression are summarized in Fig. 3. Comparison of the V_4 (Hdex 10) IdX-negative sequence with the IdX-positive segments, V_1, V_2, and V_3 shows that the residues at positions 54 and 55 correlate with IdX expression. All 10 IdX-positive

[2] Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. An immunoglobulin heavy chain is generated from three segments of DNA: V_H, D, and J_H. *Cell.* In press.

proteins have asparagine at residues 54 and 55 with a carbohydrate side chain at position 55, whereas the IdX-negative V_4 sequence of Hdex 10 lacks the carbohydrate and has lysine residues at positions 54 and 55. Note that differences at positions other than 54 and 55 do not modify expression of IdX (e.g., V_3 of Hdex 9). Interestingly, a serine for asparagine interchange at position 54 results in partial expression of IdX (e.g., V_2 of Hdex 8). This suggests that both positions 54 and 55 are involved in the IdX determinant. The influence of the carbohydrate side chains on IdX expression is as yet unknown.

Localization of the Amino Acids Important for Idl Expression. Table II clearly shows that the Idl(M104) and Idl(J558) determinants are dependent upon D-segment residues. A comparison of the M104 and J558 proteins alone demonstrates this point. These two proteins, whose light chains have been shown to be identical by amino acid sequence analysis (32), differ only in their D segments. Thus, anti-Idl(J558), which discriminates between J558 and M104 by a factor of 1,000, must recognize the influence of arginine and tyrosine at residues 100 and 101 in J558. This is confirmed by the Hdex 9 protein, which fully expresses the Idl(J558) determinant and also has arginine and tyrosine at residues 100 and 101, despite a different V segment (V_3). Interestingly, change of arginine 100 to asparagine (Hdex 1 and 2) results in partial reactivity in the anti-Idl(J558) assays. Thus, the dominant contributor to the idiotype is tyrosine 101, with significant involvement of arginine 100. Similarly, the Idl(M104) determinant involves the D segment, tyrosine-aspartic acid. Both M104 and Hdex 8 express the Idl(M104) determinant equally and exhibit identical D segments. In addition, Hdex 7 with alanine instead of tyrosine at position 100 also shows full expression of Idl(M104). This again suggests that position 101 is the predominant contributor to the determinant. However, two Idl(M104)-negative proteins, Hdex 3 and 4, also have aspartic acid at position 101 but differ from Idl(M104)-positive proteins at position 100. This suggests that some amino acids at position 100 (tyrosine and alanine) allow the expression of Idl(M104), whereas others such as arginine and lysine prohibit it. It may be significant that both arginine and lysine are positively charged amino acids.

Discussion

Antibodies to α-(1 $\rightarrow$ 3) dextran have been thought to be relatively restricted in heterogeneity because of simple and widely shared isoelectric focusing (IEF) patterns (25, 33) and of the existence of an idiotype shared by the majority of anti-dextran antibodies (25, 34). However, by making a correlative study of IEF patterns with a set of four idiotypes developed with the existing dextran-binding myeloma proteins, M104, J558, and UPC102, it became apparent that more heterogeneity existed than was anticipated (33). The anti-idiotype reagents included goat anti-M104, absorbed to remove only isotypic activity, that detected determinants expressed by all three myelomas and thus was an IdX. Three additional reagents prepared by removing anti-IdX activity recognized only the individual myelomas, and thus were labeled anti-Idl(M104), -Idl(J558), or -Idl(UPC102) (25). It was evident that anti-dextran myeloma proteins contained both shared and unique variable-region determinants.

All four of these antisera detected proteins present in the sera of conventionally immunized responder mice. The IdX was found to be associated with 60–80% of 7S and 19S anti-dextran antibodies. This idiotype reagent is probably similar to the anti-

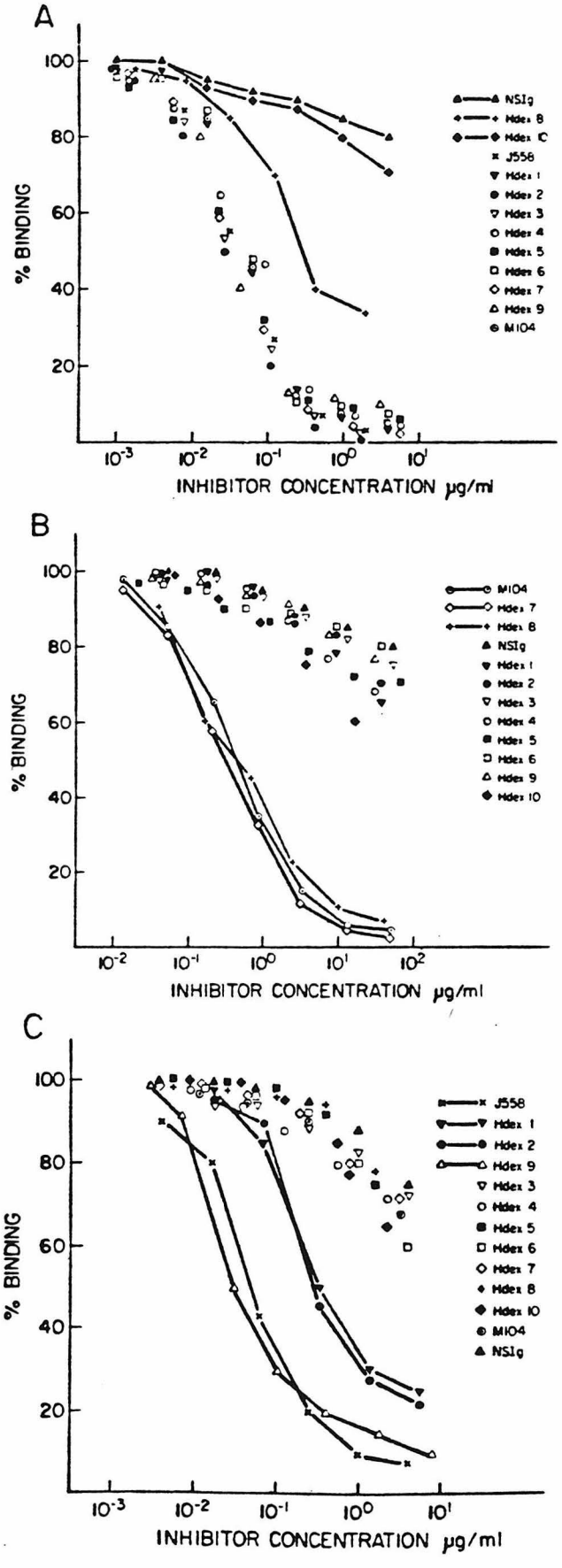

A
100
80
60
40
20
% BINDING
NSIg
Hdex 8
Hdex 10
J558
Hdex 1
Hdex 2
Hdex 3
Hdex 4
Hdex 5
Hdex 6
Hdex 7
Hdex 9
MI04
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
INHIBITOR CONCENTRATION µg/ml
B
100
80
60
40
20
% BINDING
MI04
Hdex 7
Hdex 8
NSIg
Hdex 1
Hdex 2
Hdex 3
Hdex 4
Hdex 5
Hdex 6
Hdex 9
Hdex 10
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
INHIBITOR CONCENTRATION µg/ml
C
100
80
60
40
20
% BINDING
J558
Hdex 1
Hdex 2
Hdex 9
Hdex 3
Hdex 4
Hdex 5
Hdex 6
Hdex 7
Hdex 8
Hdex 10
MI04
NSIg
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
INHIBITOR CONCENTRATION µg/ml

TABLE II

Comparison of Variable-Region Structure with Idiotype Expression

	Variable region*			Idiotype expression‡					
	V	D	J	IdX	(SDF§)	IdI(M104)	(SDF)	IdI(J558)	(SDF)
M104	1	YD	1	1.0	(1.1)	1.0	(1.2)	<0.001	NC‖
Hdex 8	2	YD	1	0.2	(1.3)	0.9	(1.2)	<0.001	NC
Hdex 7	1	AD	1	0.8	(1.3)	1.1	(1.5)	<0.001	NC
J558	1	RY	1	1.2	(1.2)	<0.002	NC	1.0	(1.1)
Hdex 9	3	RY	1	1.4	(1.1)	0.002	(1.5)	1.5	(1.1)
Hdex 1	1	NY	3	1.0	(1.1)	0.005	(2.1)	0.10	(1.2)
Hdex 2	1	NY	1	2.1	(1.2)	0.002	(1.3)	0.11	(1.1)
Hdex 3	1	RD	1	1.6	(1.2)	0.01	(1.2)	<0.001	NC
Hdex 4	1	KD	2	1.6	(1.1)	0.008	(1.4)	<0.001	NC
Hdex 5	1	SN	2	1.5	(1.0)	0.002	(1.2)	0.002	(2.2)
Hdex 6	1	SH	1	1.4	(1.3)	0.002	(1.3)	0.009	(1.5)
Hdex 10	4	VN	1	<0.001	NC	<0.001	NC	0.002	(1.2)

* See Fig. 1 for explanation of variable-region segments. The one-letter amino acid code of Dayhoff (30) is used to define the D segments.

‡ See Materials and Methods.

§ Standard deviation factor. Log SDF = SD of log (idiotype expression).

‖ Not calculable.

J558 idiotype described by Blomberg et al. (34) that inhibited 80% of anti-dextran plaque-forming cells. Also, amino acid sequences of pooled IdX-negative and IdX-positive dextran-binding serum antibodies were identical for at least the first 30–50 residues (35). This demonstrated the limited heterogeneity of V_H segments in anti-dextran antibodies and that the IdX reagent recognized differences in the C-terminal portion of the V_H region of these antibodies. The IdI were expressed to a much smaller degree than the IdX and appeared to be associated with antibodies that also expressed the IdX activity. The frequency of expression for individual idiotypes followed the order M104 > J558 > UPC102, but the sum of the antibodies bearing these IdI determinants accounted for only a minority (~10%) of the anti-dextran antibodies (25). This left the majority of antibodies defined either by an IdX alone or by the absence of any detectable idiotype.

The ability to produce α-(1 → 3) dextran-binding hybridoma proteins has not only enhanced our study of the anti-dextran repertoire, but also has allowed us to study the molecular bases of idiotypes. This is possible primarily for two reasons. First, the two myeloma proteins, M104 and J558, are known to share the identical λ-light chain (32) which by isoelectric focusing appears common to all the hybridoma proteins. Therefore, the light chain should be a constant factor in idiotype expression, thus

FIG. 2. Relative inhibitory patterns of dextran binding proteins in radioimmune assays for (A) IdX, (B) IdI(M104), and (C) IdI(J558). Inhibitory patterns for 12 dextran binding proteins in idiotype assays as described in Materials and Methods. Distinctive patterns are emphasized by solid lines. NSIg represents the inhibitory capacity of an immunoglobulin fraction of normal mouse serum.

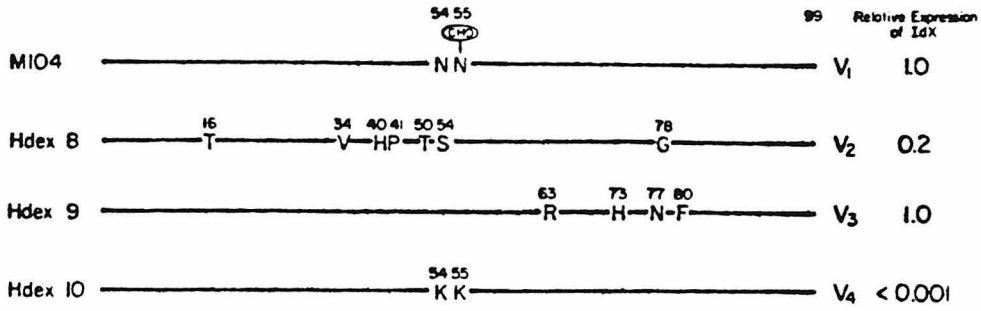


FIG. 3. Localization of amino acids important for IdX expression. The V segments of four dextran binding proteins illustrate the localization of IdX expression to asparagines 54, 55, and/or carbohydrate. M104 is used for comparison; solid lines denote identity with M104. Differences are indicated by the one-letter code of Dayhoff (30). CHO represents the carbohydrate attached to M104, Hdex 8, and Hdex 9, but absent from Hdex 10. The relative expression of IdX refers to idiotype expression in radioimmune assay described in Materials and Methods.

simplifying characterization of sequence correlates for the idiotypes. It should be emphasized that this supposition requires experimental confirmation by direct sequence analysis of the λ-chains. Secondly, the heavy chains share a large degree of sequence identity with only one or two amino acid differences between some idiotype-positive and -negative V_H regions.

Three separate idiotype determinants on dextran-binding immunoglobulins correlate well with hypervariable-region structure. The IdX correlates with two amino acids and/or associated carbohydrate in the second hypervariable region, and the two IdI determinants are dependent upon D-segment structure in the third hypervariable region.

The finding that the dextran idiotypes are dependent upon one or two amino acids is consistent with the probable molecular bases of other serological markers of immunoglobulins, notably, Gm markers of γ-chains, the Inv markers on K chains, and the O_z and Kern markers of λ-chains (36–38). Likewise, the rabbit γ-chain markers a11–12 and a14–15 reflect single amino acid substituents (39, 40). Studies by Vrana et al. (11) on the light-chain idiotypes of inulin-binding myeloma proteins limit the location of an IdI and IdX determinant to a few amino acids each.

Other families of immunoglobulins have been characterized idiotypically. In a study of human IgM cold agglutinins, Williams et al. (41) were the first to show that proteins with similar binding specificities possess both shared and unique antigenic determinants. Furthermore, myeloma and hybridoma proteins that bind inulin, levan, galactan, and poly-(L-glutamic acid^{60}, L-alanine^{30}, L-tyrosine^{10}) (10, 42–44) have been studied extensively and demonstrate both IdX and IdI specificities. Like dextran antibodies, conventionally raised antibodies to these antigens exhibit IdX determinants; however, IdI determinants are rarely detected. This has led to the concepts that IdX determinants reflect germ-line genes, and that IdI determinants reflect random somatic mutation. Our data show how IdI determinants can occur in natural antibody; however, the mechanism by which the D-segment diversity arises is not yet understood.[2] Whether the IdI determinants in other idiotype systems will localize to the D segment awaits further study.

There are several important points to be made from these data concerning the use of idiotypes as structural markers for genetic studies. (*a*) The anti-IdX reagent is

specific for only two adjacent amino acids and/or carbohydrate in the V segment although the antiserum was absorbed only for isotype activity and with normal serum. This suggests that either the two amino acids and/or carbohydrate constitute the major antigenic determinant of the V segment or that normal serum immunoglobulins contain structures similar to the remaining determinants of M104. (*b*) Three of the four dextran V segments express the IdX determinant. Thus, this highly specific idiotypic determinant is shared by more than one germ-line gene if, indeed, the distinct V-segment sequences are encoded by different germ-line V-gene segments. (*c*) Idiotypic determinants have been shown to depend on V and D segments. It is conceivable that some idiotypic determinants may involve J segments. Whereas V- and possibly D-segment idiotypes may indeed be specific for dextran-binding proteins, potential J-segment idiotypes may be shared by proteins binding to a number of antigens. For example, the J_1 sequence has been shown to be present in myeloma proteins binding phosphocholine and galactan (18). Accordingly, it is possible that J segments are shared widely among immunoglobulin groups and that data that demonstrate an unusually high degree of idiotype sharing may be a result of anti-J-segment activity. (*d*) The individual idiotypes in the dextran system are dependent upon the D segment, which may or may not involve a germ-line DNA sequence: one possibility is that the diversity in this segment arises from a somatic variation mechanism. If so, it may be difficult to correlate the IdI determinants with distinct germ-line DNA sequences. Hence, the IdI markers may have uncertain significance in terms of gene mapping, linkage, and expression.

The localization of idiotypic determinants to V, D, and potentially even J segments has implications for studies that have used idiotypes as probes to analyze V-region composition and inheritance (17, 45, 46), to map V-region gene order through the analysis of recombinant animals (14, 45, 47), and to compare the nature of T cell receptors (48, 49) or effector factors (50, 51) to immunoglobulins, to name a few. Clearly, correct interpretations of results that involve idiotypes are dependent upon precise definition of the structure of each determinant and knowledge of the genetic bases for their inheritance.

Summary

For the first time V-region amino acid sequence differences have been correlated with the expression of cross-reactive and individual idiotypes through an analysis of 12 dextran-binding proteins. This correlation has been possible because of the apparent sequence identity of the corresponding lambda chains. Expression of a cross-reactive idiotype was localized to two residues and/or a carbohydrate in the second hypervariable region of the heavy chain. Two individual idiotypes correlate with the two amino acids within the third hypervariable region that comprises the D segment of the dextran-binding proteins. These results demonstrate that idiotype reagents can recognize two amino acid differences within V and D segments of classical variable regions. In anti-dextran antibodies, cross-reactive idiotypes involve V-region determinants, whereas individual idiotype determinants correlate with D-segment variation.

We thank Alexander Shaffer for assistance in producing and maintaining hybridomas and Donna Venturini for assistance with idiotype assays.

Received for publication 14 January 1980.

References

1. Slater, R. J., S. M. Ward, and H. G. Kunkel. 1955. Immunologic relationships among the myeloma proteins. *J. Exp. Med.* **101**:85.
2. Oudin, J. 1966. The genetic control of immunoglobulin synthesis. *Proc. R. Soc. Lond. B Biol. Sci.* **166**:207.
3. Grey, H. M., M. Mannik, and H. G. Kunkel. 1965. Individual antigenic specificity of myeloma proteins. Characteristics and localization to subunits. *J. Exp. Med.* **121**:561.
4. Capra, J. D., and J. M. Kehoe. 1974. Structure of antibodies with shared idiotypy: the complete sequence of the heavy chain variable regions of two IgM anti-gamma globulin. *Proc. Natl. Acad. Sci. U. S. A.* **71**:4032.
5. Grey, H. M., and M. Mannik. 1965. Specificity of recombinants of H and L chains from human γG-myeloma proteins. *J. Exp. Med.* **122**:619.
6. Wells, J. V., H. H. Fudenberg, and D. Givol. 1973. Localization of idiotypic antigenic determinants in the F_v region of murine myeloma protein MOPC 315. *Proc. Natl. Acad. Sci. U. S. A.* **70**:1585.
7. Kunkel, H. G., R. J. Winchester, F. G. Joslin, and J. D. Capra. 1974. Similarities in the light chains of anti-γ-globulins showing cross-idiotypic specificities. *J. Exp. Med.* **139**:128.
8. Ghose, A. C., and F. Karush. 1974. Chain interactions and idiotypic specificities of homogenous rabbit anti-lactose antibodies. *J. Immunol.* **113**:162.
9. Slack, J. H., M. Shapiro, and M. Potter. 1979. Serum expression of VK structure, VK-11, associated with inulin antibodies controlled by gene(s) linked to the mouse IgC_H complex. *J. Immunol.* **122**:230.
10. Lieberman, R., M. Vrana, W. Humphrey, Jr., C. C. Chien, and M. Potter. 1977. Idiotypes of inulin-binding myeloma proteins localized to variable region light and heavy chains. Genetic significance. *J. Exp. Med.* **146**:1294.
11. Vrana, M., S. Rudikoff, and M. Potter. 1979. The structural basis of a hapten-inhibitable K-chain idiotype. *J. Immunol.* **122**:1905.
12. Briant, B. W., and A. Nisonoff, A. 1970. Quantitative investigations of idiotypic antibodies. IV. Inhibition by specific haptens of the reaction of anti-hapten antibody with its anti-idiotypic antibody. *J. Exp. Med.* **132**:951.
13. Briant, B. W., J. Haimovich, and A. Nisonoff. 1971. Reaction of anti-idiotypic antibody with the hapten binding site of a myeloma protein. *Proc. Natl. Acad. Sci. U. S. A.* **68**:3136.
14. Claflin, J. L., and J. M. Davie. 1974. Clonal nature of the immune response to phosphorylcholine. IV. Idiotypic uniformity of binding site associated antigenic determinants among mouse antiphosphorylcholine antibodies. *J. Exp. Med.* **140**:673.
15. Claflin, J. L., and J. M. Davie. 1975. Specific isolation and characterization of antibody directed against the binding site antigenic determinants. *J. Immunol.* **114**:70.
16. Briles, D. E., and R. M. Krause. 1974. Mouse strain specific idiotypes and interstrain idiotypic cross reactions. *J. Immunol.* **113**:522.
17. Carson, D., and M. Weigert. 1973. Immunochemical analysis of cross-reacting idiotypes of mouse myeloma proteins with anti-dextran activity and normal anti-dextran antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **70**:235.
18. Schilling, J., B. Clevinger, J. M. Davie, and L. Hood. 1980. Amino-acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V region gene segments. *Nature (Lond.).* **283**:35.
19. Hansburg, D., D. E. Briles, and J. M. Davie. 1976. Analysis of the diversity of murine antibodies to dextran B1355. I. Generation of a large, pauci-clonal response by a bacterial vaccine. *J. Immunol.* **117**:569.
20. Galfre, G., S. C. Howe, C. Milstein, G. W. Butcher, and J. C. Howard. 1977. Antibodies to major histocompatibility antigen produced by hybrid cell lines. *Nature (Lond.).* **266**:550.
21. Margulies, D. H., N. M. Kuehl, and M. D. Scharff. 1976. Somatic cell hybridization of mouse myeloma cells. *Cell.* **8**:405.

22. Kohler, G., S. C. Howe, and C. Milstein. 1977. Fusion between immunoglobulin-secreting and non-secreting myeloma cell lines. *Eur. J. Immunol.* **6**:231.
23. Littlefield, J. W. 1964. Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. *Science (Wash. D. C.).* **145**:709.
24. Coffino, P., R. Baumal, R. Laskov, and M. D. Scharff. 1972. Cloning of mouse myeloma cells and detection of rare variants. *J. Cell. Physiol.* **79**:429.
25. Hansburg, D., D. E. Briles, and J. M. Davie. 1977. Analysis of the diversity of the murine response to dextran B1355. II. Demonstration of multiple idiotypes with variable expression in several strains. *J. Immunol.* **119**:1406.
26. Clevinger, B., D. Hansburg, and J. Davie. 1978. Murine anti-$\alpha(1 \rightarrow 3)$ dextran antibody production by hybrid cells. *Curr. Top. Microbiol. Immunol.* **81**:110.
27. Seidman, J. G., and P. Leder. 1978. The arrangement and rearrangement of antibody genes. *Nature (Lond.).* **276**:790.
28. Brack, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell.* **15**:1.
29. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature (Lond.).* **276**:785.
30. Dayhoff, M. O. 1976. Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Silver Spring, Md. **5**:189.
31. Kehry, M., C. Sibley, J. Fuhrman, J. Schilling, and L. Hood. 1979. Amino acid sequence of a mouse immunoglobulin μ chain. *Proc. Natl. Acad. Sci. U. S. A.* **76**:2932.
32. Weigert, M. G., I. M. Cesari, S. J. Yonkovich, and M. Cohn. 1970. Variability in the lambda light chain sequences of mouse antibody. *Nature (Lond.).* **228**:1045.
33. Hansburg, D., R. M. Perlmutter, D. E. Briles, and J. M. Davie. 1978. Analysis of the diversity of murine antibodies to dextran B1355. III. Idiotypic and spectrotypic correlations. *Eur. J. Immunol.* **8**:352.
34. Blomberg, B., W. R. Geckler, and M. Weigert. 1972. Genetics of the antibody response to dextran in mice. *Science (Wash. D. C.).* **177**:178.
35. Schilling, J., D. Hansburg, J. M. Davie, and L. Hood. 1979. Analysis of the diversity of murine antibodies to dextran B1355: N-terminal amino acid sequences of heavy chains from serum antibody. *J. Immunol.* **123**:384.
36. Mage, R., R. Lieberman, M. Potter, and W. D. Terry. 1973. Immunoglobulin allotypes. *In* The Antigens. M. Sela, editor. Academic Press, Inc., New York. **1**:299.
37. Hilschmann, N., and L. C. Craig. 1965. Amino acid sequence studies with Bence Jones proteins. *Proc. Natl. Acad. Sci. U. S. A.* **53**:1403.
38. Ein, D., and J. L. Fahey. 1967. Two types of lambda polypeptide chains in human immunoglobulins. *Science (Wash. D. C.).* **156**:947.
39. Prahl, J. W., W. J. Mandy, and C. W. Todd. 1969. The molecular determinants of the A11 and A12 allotypic specificities in rabbit immunoglobulin. *Biochemistry.* **8**:4935.
40. Appella, E., A. Chersi, R. G. Mage, and S. Dubiski. 1971. Structural basis of the A14 and A15 allotypic specificities in rabbit immunoglobulin G. *Proc. Natl. Acad. Sci. U. S. A.* **68**: 1341.
41. Williams, R. C., H. G. Kunkel, and J. D. Capra. 1968. Antigenic specificities related to the cold agglutinin activity of gamma M globulins. *Science (Wash. D. C.).* **161**:379.
42. Lieberman, R., M. Potter, W. Humphrey, E. B. Mushinski, and M. Vrana. 1975. Multiple individual cross-specific idiotypes on 13 levan-binding myeloma proteins of BALB/c mice. *J. Exp. Med.* **142**:106.
43. Mushinski, E. B., and M. Potter. 1977. Idiotypes on galactan binding myeloma proteins and anti-galactan antibodies in mice. *J. Immunol.* **119**:1888.
44. Ju, S.-T., M. Pierres, C. Waltenbaugh, R. N. Germain, B. Benacerraf, and M. E. Dorf. 1979. Idiotypic analysis of monoclonal antibodies to poly($Glu^{60}Ala^{30}Tyr^{10}$). *Proc. Natl. Acad. Sci. U. S. A.* **76**:2942.

45. Lieberman, R., M. Potter, E. B. Mushinski, W. Humphrey, and S. Rudikoff. 1974. Genetics of a new IgV_H (T15 idiotype) marker in the mouse regulating natural antibody to phosphorylcholine. *J. Exp. Med.* 139:983.
46. Eichmann, K., A. Coutinho, and F. Melchers. 1977. Absolute frequencies of lipopolysaccharide-reactive B cells producing A5A idiotype in unprimed, streptococcal A carbohydrate-primed, anti-A5A idiotype-sensitized and anti-A5A idiotype-suppressed A/J mice. *J. Exp. Med.* 146:1436.
47. Eichmann, K. 1972. Idiotypic identity of antibodies to streptococcal carbohydrate in inbred mice. *Eur. J. Immunol.* 2:301.
48. Germain, R. N., S.-T. Ju, T. J. Kipps, B. Benacerraf, and M. E. Dorf. 1979. Shared idiotypic determinants on antibodies and T-cell-derived suppressor factor specific for the random terpolymer L-glutamic acid^{60}-L-alanine^{30}-L-tyrosine^{10}. *J. Exp. Med.* 149:613.
49. Bach, B. A., M. I. Greene, B. Benacerraf, and A. Nisonoff. 1979. Mechanisms of regulation of cell-mediated immunity. IV. Azobenzenearsonate-specific suppressor factor(s) bear cross-reactive idiotypic determinants the expression of which is linked to the heavy-chain allotype linkage group of genes. *J. Exp. Med.* 149:1084.
50. Black, S. J., G. J. Hammerling, C. Berek, K. Rajewsky, and K. Eichmann. 1976. Idiotypic analysis of lymphocytes in vitro. I. Specificity and heterogeneity of B and T lymphocytes reactive with anti-idiotypic antibody. *J. Exp. Med.* 143:846.
51. Julius, M. H., H. Cosenza, and A. A. Augustin. 1977. Parallel expression of new idiotypes on T and B cells. *Nature (Lond.).* 267:437.

APPENDIX 2

This paper was published in Biochemistry.

Sequence of Histone 2B of *Drosophila melanogaster*†

Sarah C. R. Elgin,* James Schilling, and Leroy E. Hood

ABSTRACT: The complete sequence of histone 2B of *Drosophila* has been determined by using an improved Beckman sequenator. Comparing these data with those previously published by other investigators on the histone 2B of calf [Iwai, K., Hayashi, H., & Ishikawa, K. (1972) *J. Biochem.* (*Tokyo*) *72*, 357–367], trout [Koostra, A., & Bailey, G. S. (1978) *Biochemistry 17*, 2504–2510], and *Patella* (a limpet) [van Helden, P. D., Strickland, W. N., Brandt, W. F., & von Holt, C. (1979) *Eur. J. Biochem. 93*, 71–78], it is possible to assess the evolutionary stability of this protein. There is little conservation of sequence in the N-terminal portion of the molecule (residues 1–26 numbering according to calf H2B), while the remainder of the protein, which we designate the C-terminal portion, is highly conserved. In the region of 27–125 residues, there are 9 substitutions in the composite data among the 98 positions, 8 of them conservative. These data indicate that very different selective pressures operate on the two different portions of the H2B molecule, implying the existence of two well-defined regions. Studies on the structure of the nucleosome by others have suggested that the C-terminal portion of H2B is involved in histone–histone interactions while the N-terminal portion is a relatively free "tail" binding to DNA. The sequence data indicate that the function of the C-terminal region of H2B requires considerable sequence specificity while that of the N-terminal region does not.

During the last few years considerable evidence has been obtained which has led to and supported the nucleosome or ν body model of chromatin structure. The chromatin fiber is visualized as a string of beads, each bead made up of eight molecules of the smaller histones, 2A, 2B, 3 and 4, around which the DNA is wrapped. There are ~200 base pairs of DNA associated with each unit. Histone 1 and the nonhistone chromosomal proteins are apparently associated with the DNA on the outside of the core structure. For reviews of the evidence leading to this model and a more detailed discussion, see Elgin & Weintraub (1975), Kornberg (1977), and Felsenfeld (1978).

The pioneering studies of Fambrough, DeLange, and their colleagues (Fambrough & Bonner, 1968; DeLange et al.,

† From the Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138 (S.C.R.E.), and the Division of Biology, California Institute of Technology, Pasadena, California 91125 (J.S. and L.E.H.). *Received April 30, 1979; revised manuscript received September 27, 1979.* Supported by Grant PCM 75-15259 from the National Science Foundation to S.C.R.E. and Grant GM 06945 from the National Institutes of Health to L.E.H. S.C.R.E. is the recipient of a Research Career Development Award from the National Institutes of Health.

1969) demonstrated that the histones are among the most conserved proteins yet analyzed. Similarly, the nucleosome organization of the chromatin fiber is a highly conserved structure, having been observed for essentially all eucaryotes and for all forms of chromatin (interphase, metaphase, polytene) studied to date [e.g., Woodcock et al. (1976)]. It is of interest to determine and compare the primary structures of the histones, the building blocks of nucleosome structure, for several reasons. First, the pattern and degree of sequence conservation should indicate which portions of the molecules are necessary in an invariant form for nucleosome structure and function. Differences in the degree of conservation within a given histone may indicate the positions of different functional domains having different evolutionary constraints. Second, the availability of different histones whose sequences are completely determined should aid in studying the interactions of the histones with each other and with other components of the chromatin complex. Third, the availability of different histones whose structures have been completely determined will allow one to test the specificity of components of the chromatin system, such as histone modification enzymes. It is of particular interest to work with organisms where genetic manipulation is possible and where studies of the histone gene sequences are underway (Lifton et al., 1978). For these reasons, we have determined the complete sequence of histone 2B of *Drosophila*.

Methods

Isolation of Histone 2B. Histones were isolated by extraction with 1.6 N NaCl–0.2 N HCl from crude chromatin of 6–18-h-old *Drosophila melanogaster* Oregon R embryos prepared by the method of Elgin & Hood (1973). The histones were then dialyzed against 0.01 M acetic acid, lyophilized, and stored at −20 °C for future use. Eighty milligrams of histone was redissolved in 0.01 M acetic acid, and histone 1 was removed by extraction with 5% perchloroacetic acid. This step is necessary since histone 1 of *Drosophila* contains cysteine (Alfageme et al., 1974; S. C. R. Elgin, unpublished experiments). The pellet was dissolved in 8 M urea–0.05 M Tris-HCl (pH 8.0)–1% β-mercaptoethanol and dialyzed extensively against 0.1 M Tris-HCl (pH 8)–0.3 M NaCl—1 mM ethylenediaminetetraacetic acid (column buffer, CB). The proteins were applied in this buffer to a 2.5 mm × 12 cm column of preequilibrated activated thio-Sepharose 4B resin (Pharmacia). Histones 2A and 2B were eluted in the runoff peak (flow rate 4.5 mL/h) with some H3 contamination. The bulk of the H3 was covalently bound to the column. H4 remained in association with H3 and could be eluted with CB + 5 M urea, indicating the expected hydrophobic interaction. H3 was eluted with CB + 5 M urea + 20 mM L-cysteine. The column could be regenerated by washing with CB + 1.5 mM dithiouracil. Recovery of purified histone was ~75%. Histone 2B was purified from the runoff peak by repeated extraction with 80% ethanol–0.26 M HCl (Oliver et al., 1972). Electrophoresis on acid–urea gels indicated that the product was 85–90% H2B, the single major contaminant being H2A. No subfractions of histone 2B were observed in this work, nor have any been reported for *Drosophila* in other studies (Oliver & Chalkley, 1972; Alfageme et al., 1974).

Sequencing Strategy. The N-terminal sequence of the H2B protein was determined for residues 1–62 on an improved Beckman sequenator (data shown in Figure 1). A portion of the H2B protein was blocked at its N-terminal α-amino and lysine ϵ-amino groups by succinylation. This "blocked" protein was cleaved with cyanogen bromide at positions 56 and 59, and the resulting mixture of three peptides (1–56, 57–59, and 60–122) was loaded in the sequencer without further purification. The two unblocked peptides (57–59 and 60–122) were sequenced simultaneously for 59 residues (data shown in Figure 2). The H2B protein was contaminated with ~10% H2A protein, and the sequence of a cyanogen bromide fragment from H2A also could be followed. We also carried out a tryptic digestion of succinylated H2B and sequenced the "arginine" peptides 31–69 and 97–122 simultaneously as a mixture (data not shown). The arginine peptides were sequenced to within four residues of the C terminus of the C-terminal peptide.

Succinylation. The H2B was dissolved at 10 mg/mL in 8 M guanidine hydrochloride and reacted on a pH stat at pH 9.0 with a 100-fold excess of succinic anhydride over protein (w/w) at room temperature.

Cyanogen Bromide Cleavage. The H2B protein was dissolved at 10 mg/mL in 70% formic acid and reacted for 18 h at 4 °C with a fivefold excess of cyanogen bromide (w/w).

Tryptic Digestion. The H2B protein was dissolved at 10 mg/mL in 0.2 M ammonium bicarbonate and digested at 37 °C for 2.5 h with 250 μg/mL trypsin.

Preparation of Arginine Fragments. The mixture of peptides resulting from tryptic digestion of succinylated H2B was separated on Sephadex G-15 in 0.2 M ammonia into several peaks. The first peak eluted from the column contained arginine fragments extending from residues 31 to 69 and 97 to 122. This mixture of two peptides was analyzed on the automatic sequenator.

Automated Sequence Analysis. Polypeptides were sequenced on a modified Beckman 120B sequencer as described by Hunkapiller & Hood (1978). The phenylthiohydantoin amino acids were identified by high-pressure liquid chromatography on Du Pont Zorbax ODS columns.

Carboxypeptidase Digestion. The H2B protein was dissolved in 0.2 M ammonium bicarbonate. A mixture of carboxypeptidases A and B (both DFP-treated; Sigma) was added at time 0 at a histone/enzyme ratio of 100:1. Digestion was at room temperature for 5 or 60 min. After being boiled for 2 min to stop the digestion, samples were lyophilized and analyzed on a Durrum D-500 amino acid analyzer.

Results and Discussion

Automated Sequence Analysis. The amino acid sequence for positions 1–62 is given unambiguously in Figure 1. The results have been presented by plotting nanomoles of phenylthiohydantoin amino acids recovered vs. residue position for each amino acid. The samples for steps 39–42 were accidently pooled and analyzed together. The sequence of this region was determined by the sequence analysis of a tryptic (arginine) peptide containing residues 31–69 (data not shown).

The sequence analysis of the cyanogen bromide fragment starting at position 60 is unambiguous out to the threonine at position 119 (Figure 2). Indeed, these data suggest that positions 120 and 121 are Ser-Ser, although the signal-to-noise ratio makes this latter assignment tenuous. Arginine peptides 31–69 and 97–122 have sequences which confirm completely the results given in Figures 1 and 2.

Carboxypeptidase Analysis. A 5-min digestion with carboxypeptidases A and B resulted in the release of lysine, thus identifying this residue as the C-terminal amino acid. A longer digestion time resulted in the release of additional amino acids which are consistent with the sequence assignment of the C-terminal region of histone 2B given in Figure 3.

Summary of Sequence Data. The final sequence of histone H2B of *D. melanogaster* with a summary of the evidence used is given in Figure 3. The assignment of Ser-Ser at positions

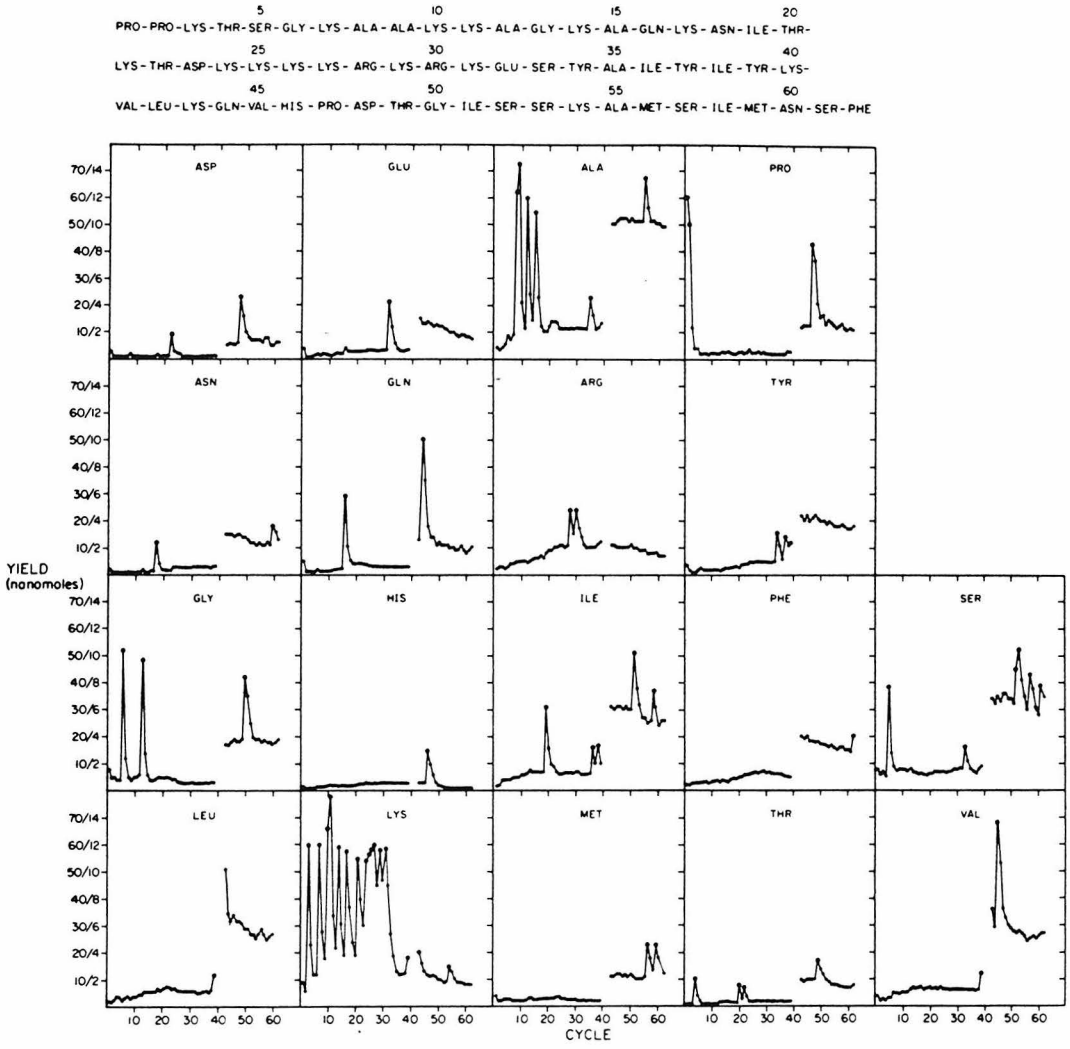


FIGURE 1: Nanomoles of each phenylthiohydantoin amino acid recovered at each step for the H2B N-terminal sequence. The amino acid identified at each step is indicated by a heavier dot. Samples for steps 39–42 were accidentally pooled, and the four amino acids were therefore identified together. The scale is expanded fivefold at cycle 43, except for lysine which is not expanded and histidine and arginine which are plotted throughout on the expanded scale.

120 and 121 remains tentative. The sequence is in complete agreement with the sequence of the H2B gene from the DNA clone cDm 500 from *D. melanogaster* which has been completed for residues 1–117 (M. Goldberg and D. S. Hogness, personal communication). In addition, the data obtained indicate that *Drosophila* H2A, like that of calf, has a blocked N-terminal residue, but unlike calf has a methionine at position 55. The H2A sequence from residues 55–95 (numbering according to calf thymus H2A) can for the most part be detected as the minor peaks in Figure 2.

Conservation of Sequence. A comparison of the sequence of *Drosophila* H2B with complete H2B protein sequences previously determined by others is presented in Figure 4. In preparing Figure 4, an attempt was made (by visual inspection) to align the proteins for maximum homology, using insertions and deletions as necessary. In the following discussion we will refer to amino acid positions numbered according to the calf protein, as in Figure 4. While H2B's of calf and trout are fairly homologous (1 deletion, 7 substitutions), it is readily apparent that H2B of *Drosophila* differs more extensively from its calf counterpart [6 sequence gaps (deletions or insertions), 15 substitutions]. In fact, there is very little sequence homology for the region 1–26 (numbered according to the calf sequence) when the composite data for these three proteins are examined. Within the region 27–125, homology is considerable; there are no sequence gaps and only eight substitutions of which seven are conservative. While this manuscript was in preparation, the complete amino acid sequence of H2B from *Patella granatina*, a limpet, was reported (van Helden et al., 1979). Comparison with calf H2B indicates a minimum of 5 gaps, all before position 27, and 21 substitutions, 15 conservative. As would be expected from the established phylogenetic relationships, *Patella* H2B shows considerably greater homology with *Drosophila* H2B; there are only 2 gaps, both before position 27, and 12 substitutions, 10 conservative, in this comparison.

The results suggest that there are two quite different regions in H2B, possibly indicative of different functions with very

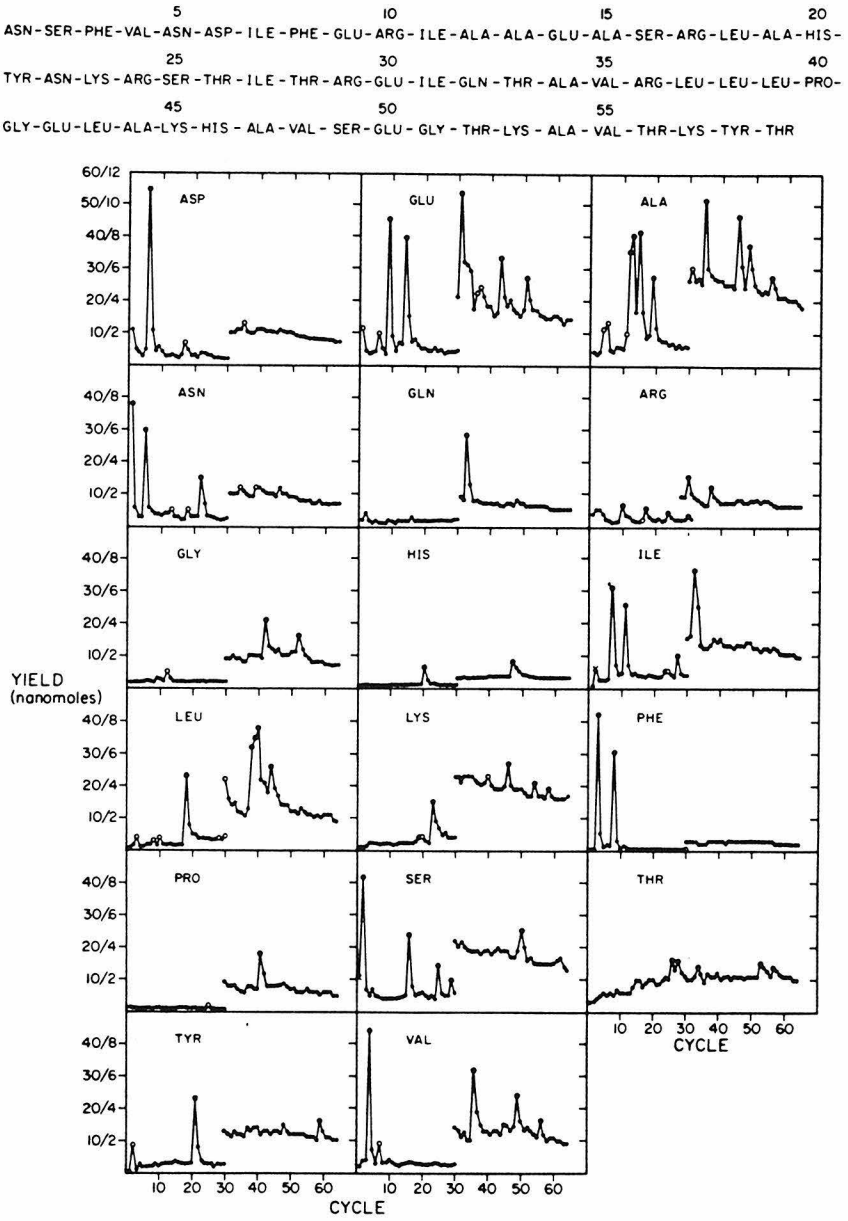


FIGURE 2: Nanomoles of each phenylthiohydantoin amino acid recovered at each step of the sequence of H2B cyanogen bromide peptides. Three sequences can be seen: (●) H2B 60–122; (×) H2B 57–59; (○) contaminating H2A sequences (see Results and Discussion). The N-terminal peptide was blocked. The scale is amplified fivefold at cycle 30 for all amino acids, except threonine which is plotted throughout on the expanded scale.

different requirements for sequence specificity. The N-terminal region, residues 1–26, can be defined as that portion of the molecule which lacks extensive sequence homology, which includes sequence gaps in comparisons of H2B sequences, and which includes most of the clusters of basic amino acids. The boundary of the conserved and nonconserved regions has been assessed by examination of all the currently published data. The sequence of a sea urchin (*Strongylocentrotus purpuratis*) embryo H2B has recently been reported from DNA sequencing of the gene (Sures et al., 1978). (Since the histone genes are repeated, one cannot yet be certain that this gene is expressed.) A comparison of this sequence with that of calf H2B indicates 3 gaps, all occurring before position 27, and 22 substitutions, 15 conservative (see Figure 4). In addition to the protein and DNA sequence data from somatic cell H2B histones used in constructing Figure 4, data on the N-terminal region of the analogous protein from sea urchin sperm may be considered. Five H2B proteins from *Parechinus angulosus* and *Pasammechinus miliaris* sperm have recently been isolated and sequenced (Strickland et al., 1977a,b, 1978). The sea urchin sperm histone 2B's are ~15% larger than those of calf, trout, and *Drosophila*. An excellent alignment, however, can be obtained between these proteins and calf H2B starting at residue 27. The common sequence for these proteins at

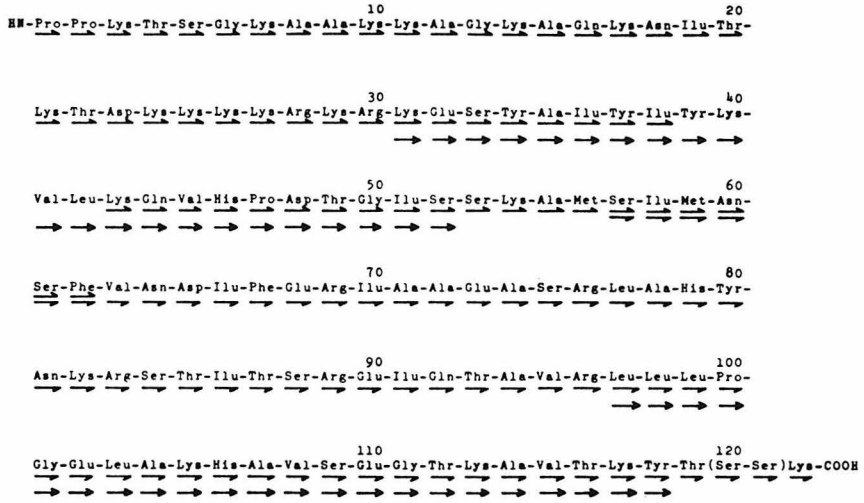


FIGURE 3: Summary of evidence for the sequence of *Drosophila* histone H2B. Residues determined by sequence analysis from the N-terminal residue are indicated by →; those determined by analysis of the cyanogen bromide peptides are indicated by ⟶; those determined by analysis of the arginine peptides are indicated by →. Parentheses indicate tentative assignments. The C-terminal residue was identified as Lys by using carboxypeptidase digestion.

```
                                    10                  20
Calf                HN-P-E-P-A-K-S-A-P---A-P-K-K---G-S-K-K-A-------V-T-K-A-Q-K-K-D-G-
Trout               HN-P-E-P-A-K-S-------A-P-K-K---G-S-K-K-A-------V-T-K-T-A-G-K-G-G-
Sea urchin            HN-A-P-T-A-Q---V-A-K-K---G-S-K-K-A-V---K-G-T-K-T-A-X---G-G-
Drosophila          HN-P-P-K-T-S-G-K-A-A-K-K-A-G-----K-A-Q-K-N-I-T-K-T-D---------
Patella             HN-P-P-K-V-S-S-K-G-A-K-K-A-G-----K-A-----K-A-A-R-S-G-D-------

                         30                  40                  50
Calf                K-K-R-K-R-S-R-K-E-S-Y-S-V-Y-V-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Trout               K-K-R-K-R-S-R-K-E-S-Y-A-I-Y-V-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Sea urchin          K-K-R-N-R-K-R-K-E-S-Y-G-I-Y-I-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-R-A-
Drosophila          K-K-K-R-R-K-R-K-E-S-Y-A-I-Y-I-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Patella             K-K-R-K-R-R-R-K-E-                                       V
Sea urchin sperm    V-K-R-R-R-R-R-E-
Others                    H                                                  V

                    60                  70                  80                  90
Calf                M-G-I-M-N-S-F-V-N-D-I-F-E-R-I-A-G-E-A-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Trout               M-G-I-M-N-S-F-V-N-D-I-F-E-R-I-A-G-E-S-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Sea urchin          M-V-I-M-N-S{F}V{N}D-I-F-E-R-I-A-G-E-S-S-R-L-A-Q-Y-N-K-K{S}T{I}S-
Drosophila          M-S-I-M-N-S-F-V-N-D-I-F-E-R-I-A-A-E-A-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Others                T V                 V           S Q         T T A
                        L                                           S

                                        100                 110                 120
Calf                S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Trout               S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Sea urchin          S-R-E-I-Q-T-A-V-R-L-I-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-T-S-K-OH
Drosophila          S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Others                                                                                A  (V)
```

FIGURE 4: Comparison of amino acid sequence data available for histone 2B. Sequences have been aligned to obtain maximum homology with the calf sequence. Numbering is according to the calf sequence. Complete sequences from the following references are presented: calf H2B, Iwai et al., 1972; trout (*Salmo trutta*) H2B, Koostra & Bailey, 1976, 1978; sea urchin (*S. purpuratis*) early embryo H2B (from DNA sequence), Sures et al., 1978; *D. melanogaster* H2B, this paper. The sequence for *P. granatina* H2B (van Helden et al., 1979) for residues 35–125 is identical with that of *Drosophila* with the exception of valine at position 54. Positions at which substitutions have occurred in the region 27–125 are underlined. The partial sequence of sea urchin sperm H2B is from Strickland et al. (1978) (see text). Data collectively presented as "others" for the region 27–125 amino acids indicate any additional substitutions noted in the sequences from the following: sea urchin (*P. miliaris*) H2B (from DNA sequencing), positions 20–122 (Birnstiel et al., 1977; Schaffner et al., 1978); sea urchin (*P. angulosus*) embryo H2B, positions 27–35 and 59–79 (Brandt & von Holt, 1978); sea urchin (*P. angulosis*) embryo H2B fraction III, positions 59–70, embryo H2B fraction V, positions 59–71, and somatic H2B, positions 59–80 (Brandt et al., 1979); rat chloroleukemia H2B, residues 121–125 (Martinage et al., 1976); mouse H2B, residues 73–79 (Franklin & Zweidler, 1977); calf minor H2B, residues 73–79 (Franklin & Zweidler, 1977). Parentheses indicate ambiguities or uncertainties in the data.

positions 27–35 is given in Figure 4. A comparison of the C-terminal 90 residues of the 5 sperm H2B proteins with calf H2B indicates no gaps and 20 substitutions, 18 of them conservative. The somatic histone 2B's of pea and tobacco have also been reported to be ~20% larger than that of calf (Gigot et al., 1976; Spiker & Isenberg, 1977). In an analysis of the tryptic peptides and BrCN cleavage fragments of pea H2B, Hayashi et al. (1977) were able to align the peptides with the

calf H2B sequence starting at position 35. No significant alignments were obtained in the N-terminal portion. Thus, in both these cases the additional residues must be part of the N-terminal region. Comparison of the C-terminal 90 residues (from positions 35 to 125) tentatively indicates 1 gap and 19 substitutions for pea H2B relative to calf H2B. Most of the substitutions are conservative. If the sequence gap at position 39/40 for pea H2B implied by the peptide data is substantiated in the complete sequence, one would wish to consider moving the boundary position to 40. Otherwise, the data available indicate that the boundary occurs at position 26/27, with the size and position of the block of basic residues starting at position 27 relatively conserved features (see Figure 4).

It is of interest to compare this result to that obtained for the other histones involved in the nucleosome core. Extensive data on H3 and H4 indicate that these proteins are highly conserved over the entire molecule. No sequence gaps and only five positions where substitutions have occurred have been observed in sequence studies of H3 from vertebrates, invertebrates, and plants [see von Holt et al. (1979), Schaffner et al. (1978), and Sures et al. (1978)]; no sequence gaps and only three positions where substitutions have occurred have been seen in these H4's [see von Holt et al. (1979)]. In both cases, no substitutions have been found in the N-terminal 40 amino acids.[1] In particular, we have examined the N-terminal sequence of *Drosophila* H3 to position 32 and found it to be identical with that of calf (S. C. R. Elgin, unpublished experiments). Histone 2A is more variable; 8 sequence gaps and 27 substitutions may be noted in a composite comparison of the sequences from calf, rat, mouse, chick, trout, sea urchin, and *Drosophila* (partial) [see von Holt et al. (1979), Laine et al. (1978), and R. Goldberg and D. S. Hogness (personal communication)]. Insertions and deletions are mostly clustered in the N- and C-terminal regions (residues 1–21 and 121–129); the substitutions appear randomly distributed. A recent study of the sequences of H2A fractions from wheat germ reports the N-terminal region to be variable in both size and sequence (Rodrigues et al., 1979). Thus, while histones H3 and H4 are conserved structures throughout, both H2B and H2A show size and sequence variability in the terminal portions of the molecule with a conserved central core.

It should be noted that while both the size and sequence of the N-terminal region of H2B vary, the overall amino acid composition of the region is relatively constant, with 26–32% basic amino acids for the somatic H2B's (data from calf, trout, *Drosophila*, *Patella*, and sea urchin embryo) and 38–41% basic amino acids for the sea urchin sperm H2B's. There is no conserved, readily identifiable pattern of spacing of either the basic residues or the prolines in the N-terminal region of the somatic 2B histones. These observations suggest that only the very broad structural characteristics predictable from amino acid composition (e.g., net charge) need be conserved for this portion of the protein to function properly. It is known that several of the lysine residues in the N-terminal region can be modified by acetylation. Analysis of the data in the trout system for all the histones has led Dixon et al. (1975) to suggest that sites of the type -X-Lys-X-, where X is glycine, alanine, serine, or threonine, or of the type -Lys-Arg-, -Arg-Lys-, or -Lys-Lys- are preferentially acetylated, although the secondary structure of the protein is no doubt important in limiting acetylation to the N-terminal region. In the composite data of Figure 4 for H2B, 31 of 35 lysines in the N-terminal region occur at such sites. While to some degree this must be a reflection of the amino acid composition of this region, the conservation of this type of lysine site suggests a functional requirement.

Functional Analysis of H2B. Several lines of evidence have suggested the presence of at least two separate functional domains in H2B (and in the other core histones). On the basis of nuclear magnetic resonance and optical–spectroscopic data, Bradbury and his colleagues (Bradbury & Rattle, 1972; Bradbury et al., 1972) have suggested that residues 31–102 constitute a relatively structured part of H2B, while residues 1–30 and 102–125 are primarily involved in DNA binding. It should be noted, however, that these DNA binding studies were carried out in the absence of a full complement of histones. More recent work supports the concept of a H2B structural core formed by residues 55–78 (Lilley et al., 1975; Tancredi et al., 1976). A similar study using proton magnetic resonance and circular dichroism to analyze the H2A–H2B complex indicates that residues 37–114 of H2B are involved in the stable tertiary structure that is formed (Moss et al., 1976). This region is essentially that which is highly conserved in amino acid sequence.

Results of this type have led several investigators to predict that the N-terminal regions of H2B (and those of the other core histones) would project as "tails" from the nucleosome core, available to bind to the associated DNA [e.g., Van Holde et al. (1974)]. Such a model is supported by the observation of Weintraub & Van Lente (1974) that trypsin digestion of chromatin, while degrading H1, H5, and the nonhistone chromosomal proteins completely, removes the N-terminal 20–30 residues from the smaller histones but leaves the core essentially intact. All four of the smaller histones are required to generate the trypsin-resistant complex, which is stable in the absence of DNA in 2 M NaCl (Weintraub et al., 1975).

Studies of histone interactions in solution show strong complex formation between H2B and H2A and H2B and H4 and a relatively weaker interaction between H2B and H3 (D'Anna & Isenberg, 1974; Spiker & Isenberg, 1977). More recently, Spiker & Isenberg (1978) have found that the pattern of pairwise histone–histone interactions is conserved for heterologous complexes of core histones from calf and pea, even though pea H2B is ~20% larger than calf H2B. These interkingdom complexes form with essentially the same binding constants as the regular complexes. Spiker & Isenberg (1978) estimate (given the error limits of their data) that at most one or two residue changes may have occurred at the binding surfaces between two interacting histones. Given the results obtained by protein sequencing, these findings imply that the C-terminal region of H2B is involved in histone–histone interactions, while the N-terminal region is not. Direct studies using chemical cross-linking indicate that the middle one-third of H2B interacts closely with H2A and the C-terminal one-third of H2B interacts closely with H4 (Martinson et al., 1979; DeLange et al., 1979). Similar cross-linking results have been obtained by using chromatin from either plants or animals (Martinson & True, 1979).

Thus, the data suggest that the two very different regions of H2B identified by comparative sequence analysis reflect two very different functional domains of the protein. There is a conserved core of histone 2B (residues 35–125) which is both necessary and sufficient to generate the stable tertiary

[1] Glover & Gorovsky (1979) have recently found that H4 from a protozoa, *Tetrahymena thermophil*, differs substantially from that of calf. In the first 66 residues there are 2 gaps and 13 substitutions. This H4 behaves differently from others in several functional assays, complexing less strongly than calf H4 with both calf and *Tetrahymena* H2B and H3 (Glover & Gorovsky, 1978); consequently, it has not been included in this analysis.

structure of this protein in a complex(es) with histones 2A, 3, and 4. The more basic N-terminal domain 1–26 binds to DNA in a manner requiring considerably less structural specificity. A run of five to eight basic residues occurs at the end of the N-terminal domain; the significance of this cluster is unknown. Why H2B differs from the core histones H3 and H4 in its pattern of conserved structure remains to be resolved. The case of H2A resembles that of H2B. While current evidence suggests that the H3–H4 tetramer is necessary and sufficient for beginning DNA folding (Camerini-Otero et al., 1976; Sollner-Webb et al., 1976; Moss et al., 1977), direct tests using antibodies against H2B indicate that this histone is present in most, if not all, nucleosomes (Bustin et al., 1976; Simpson & Bustin, 1976). Studies on the reactive properties of the histones in chromatin have indicated that the N-terminal region of H2B is relatively exposed [e.g., Malchy (1977)]. A functional analysis of the different constraints placed on the N-terminal domains of the different core proteins must await further data on chromatin structure.

References

Alfageme, C. R., Zweidler, A., Mahowald, A., & Cohen, L. H. (1974) *J. Biol. Chem. 249*, 3729–3736.

Birnstiel, M. L., Schaffner, W., & Smith, H. O. (1977) *Nature (London) 266*, 603–607.

Bradbury, E. M., & Rattle, H. W. E. (1972) *Eur. J. Biochem. 27*, 270–281.

Bradbury, E. M., Carey, P. D., Crane-Robinson, C., Riches, P. L., & Johns, E. W. (1972) *Eur. J. Biochem. 26*, 482–489.

Brandt, W. F., & von Holt, C. (1978) *Biochim. Biophys. Acta 537*, 177–181.

Brandt, W. F., Strickland, W. N., Strickland, M., Carlisle, L., Woods, D., & von Holt, C. (1979) *Eur. J. Biochem. 94*, 1–10.

Bustin, M., Goldblatt, D., & Sperling, R. (1976) *Cell 7*, 297–304.

Camerini-Otero, R. D., Sollner-Webb, B., & Felsenfeld, G. (1976) *Cell 8*, 333–347.

D'Anna, J. A., Jr., & Isenberg, I. (1974) *Biochemistry 13*, 4992–4997.

DeLange, R. J., Fambrough, D. M., Smith, E., & Bonner, J. (1969) *J. Biol. Chem. 244*, 5669–5679.

DeLange, R. J., Williams, L. C., & Martinson, H. G. (1979) *Biochemistry 18*, 1942–1946.

Dixon, G. H., Candido, E. P. M., Honda, B. M., Louie, A. J., Macleod, A. R., & Sung, M. T. (1975) *Ciba Found. Symp. 28*, 229–250.

Elgin, S. C. R., & Hood, L. E. (1973) *Biochemistry 12*, 4984–4991.

Elgin, S. C. R., & Weintraub, H. (1975) *Annu. Rev. Biochem. 44*, 725–774.

Fambrough, D. M., & Bonner, J. (1968) *J. Biol. Chem. 243*, 4434–4439.

Felsenfeld, G. (1978) *Nature (London) 271*, 115–122.

Franklin, S. G., & Zweidler, A. (1977) *Nature (London) 266*, 273–275.

Gigot, C., Phillipps, G., Nicolaieff, A., & Hirth, L. (1976) *Nucleic Acids Res. 3*, 2315–2329.

Glover, C. V. C., & Gorovsky, M. A. (1978) *Biochemistry 17*, 5705–5713.

Glover, C. V. C., & Gorovsky, M. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 585–589.

Hayashi, H., Iwai, K., Johnson, J. D., & Bonner, J. (1977) *J. Biochem. (Tokyo) 82*, 503–510.

Hunkapiller, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124–2133.

Iwai, K., Hayashi, H., & Ishikawa, K. (1972) *J. Biochem. (Tokyo) 72*, 357–367.

Kootstra, A., & Bailey, G. S. (1976) *FEBS Lett. 68*, 76–78.

Kootstra, A., & Bailey, G. S. (1978) *Biochemistry 17*, 2504–2510.

Kornberg, R. D. (1977) *Annu. Rev. Biochem. 46*, 931–954.

Laine, B., Kmiecik, D., Sautiere, P., & Biserte, G. (1978) *Biochimie 60*, 147–150.

Lifton, R. P., Goldberg, M. L., Karp, R. W., & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol. 42*, 1047–1051.

Lilley, D. M. J., Howarth, W. W., Clark, V. M., Pardon, J. F., & Richards, B. M. (1975) *Biochemistry 14*, 4590–4600.

Malchy, B. L. (1977) *Biochemistry 16*, 3922–3927.

Martinage, A., Sautiere, P., Kerckaert, J., & Bisertz, G. (1976) *Biochim. Biophys. Acta 420*, 37–41.

Martinson, H. G., & True, R. J. (1979) *Biochemistry 18*, 1947–1951.

Martinson, H. G., True, R. J., Lau, C. K., & Mehrabian, M. (1979) *Biochemistry 18*, 1075–1082.

Moss, T., Cary, P. D., Abercrombie, B. D., Crane-Robinson, C., & Bradbury, E. M. (1976) *Eur. J. Biochem. 71*, 337–350.

Moss, T., Stephens, R. M., Crane-Robinson, C., & Bradbury, E. M. (1977) *Nucleic Acids Res. 4*, 2477–2485.

Oliver, D., & Chalkley, R. (1972) *Exp. Cell Res. 73*, 295–302.

Oliver, D., Sommer, K. R., Panyim, S., Spiker, S., Chalkley, R. (1972) *Biochem. J. 129*, 349–353.

Rodrigues, J. A., Brandt, W. F., & von Holt, C. (1979) *Biochim. Biophys. Acta 578*, 196–206.

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O., & Birnstiel, M. L. (1978) *Cell 14*, 655–672.

Simpson, R. T., & Bustin, M. (1976) *Biochemistry 15*, 4305–4311.

Sollner-Webb, B., Camerini-Otero, R. D., & Felsenfeld, G. (1976) *Cell 9*, 179–193.

Spiker, S., & Isenberg, I. (1977) *Biochemistry 16*, 1819–1826.

Spiker, S., & Isenberg, I. (1978) *Cold Spring Harbor Symp. Quant. Biol. 42*, 157–164.

Strickland, M., Strickland, W. N., Brandt, W. F., & von Holt, C. (1977a) *Eur. J. Biochem. 77*, 263–275.

Strickland, W. N., Strickland, M., Brandt, W. F., & von Holt, C. (1977b) *Eur. J. Biochem. 77*, 277–286.

Strickland, M., Strickland, W. N., Brandt, W. F., & von Holt, C. (1978) *Biochim. Biophys. Acta 536*, 289–297.

Sures, I., Lowry, J., & Kedes, L. H. (1978) *Cell 15*, 1033–1044.

Tancredi, T., Temussi, P. A., Paolillo, L., Trivellone, E., Crane-Robinson, C., & Bradbury, E. M. (1976) *Eur. J. Biochem. 70*, 403–408.

van Helden, P. D., Strickland, W. N., Brandt, W. F., & von Holt, C. (1979) *Eur. J. Biochem. 93*, 71–78.

Van Holde, K. E., Sahasrabuddhe, C. G., & Shaw, B. R. (1974) *Nucleic Acids Res. 1*, 1579–1586.

von Holt, C., Strickland, W., Brandt, W. F., & Strickland, M. S. (1979) *FEBS Lett. 100*, 201–218.

Weintraub, H., & Van Lente, F. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4249–4253.

Weintraub, H., Palter, K., & Van Lente, F. (1975) *Cell 6*, 85–110.

Woodcock, C. L. F., Safer, J. P., & Stanchfield, J. E. (1976) *Exp. Cell Res. 97*, 101–110.