AN ANALYSIS OF PATTERNS OF DIVERSITY IN ANTIBODIES WITH DEFINED SPECIFICITY

Thesis by

Nelson Daniell Johnson

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted October 3, 1980)

To my parents, for their support,

and to Mariel, for hers.

Acknowledgements

My years as a graduate student at Caltech have been tremendously rewarding. I would especially like to thank the following people for their direct assistance with my research and for their contributions to the improvement of my life:

My advisor Lee Hood, whose enthusiasm and enjoyment of scientific research provided me with a constant source of energy,

Ray Owen, whose concern provided me with an unfailing source of support and good advice,

Jack Richards, who has given me valuable assistance throughout my ten-year stay at Caltech,

Paul Barstad, Elizabeth Blankenhorn, Mike Cecka, Vince Farnsworth, John Frelinger, Jon Fuhrman, Mike Hunkapiller, Tim Hunkapiller, Chin Sook Kim, Mitch Kronenberg, and Elwyn Loh, for their contributions to my life both in and out of the lab,

My friends Jacob Brown, Rosita Hoffman, John Hoffman, John Johnson, Norman Johnson, Jim Loddengaard, Doug Rothnie, Janet Towner, Chris Wood, and Mr. Ohshima and the Caltech Karate Club,

My collaborators Richard Douglas, Barbara Eaton, Pat Gearhart, Judy Greengard, Jeff Hubert, Lee Paul, and Jeff Slankard,

And, in a class by herself, Bertha Jones, who keeps the Hood group running smoothly.

I also want to thank Connie Katz and Bernita Larsh for their cheerful assistance in typing this dissertation.

I am grateful for the support of NIH grants GM 0086, GM 06965, and AI 10781, and for the Jean Weigle Memorial Fund which paid in part the costs of producing this thesis.

Abstract

Antibodies can recognize a large number of molecular determinants (antigens) because of the diversity present in antibody combining sites. This diversity resides in regions of extensive amino acid variability termed variable (V) regions. Variable region diversity is encoded in multiple germline variable region genes and can also arise from somatic modification of these genes. An important class of somatic modifications is the rearrangement of gene segments to form complete variable region genes. In this way complete V_L genes arise from the joining of V_L and J_L gene segments while V_H genes arise from V_H, D, and J_H gene segment joining.

Studies of the V region protein sequences of hybridoma and myeloma immunoglobulins which bind phosphorylcholine show that IgM antibody V regions are considerably less diverse than IgG and IgA V regions. A comparison of protein sequence data with experiments on germline DNA suggests that at least some V segment diversity in IgG and IgA antibodies is the result of somatic mutations. D segments from phosphorylcholine-binding IgM antibodies as well as from IgG and IgA antibodies show extensive amino acid interchanges and size differences. In addition, diversity in the antibody response to phosphorylcholine is generated by associating a single V_H region with at least two different V_L regions.

The complete sequences of the V_L and V_H regions from two antibodies binding $\beta(2\rightarrow 1)$ levan have also been determined. A comparison of these sequences to protein sequence data from other $\beta(2 \rightarrow 1)$ levan-binding proteins and to a germline DNA sequence suggests that the levan-binding proteins may arise from multiple germline genes differing at the protein level by only a few amino acids. Unlike the D segments of the phosphorylcholine binding proteins, the levan-binding immunoglobulin D segments show very little diversity. In addition, the protein sequences of levan-binding immunoglobulins can be compared to published V region idiotype and antigen binding studies. These

comparisons show that idiotypes may focus on certain sections of antibody V regions, and hence be of limited value as a probe of antibody V region fine structure.

Table of Contents

Introduction

Short-lived organisms can avoid many pathogens by virtue of their abbreviated life spans (1). This is possible whenever the time required for pathogenic discovery and invasion is of the same order as or longer than a complete life cycle of the invaded organism. Long-lived organisms, on the other hand, are far more apparent to potential pathogenic invaders. A highly developed immune system is required to preserve the integrity of these organisms in the face of repeated invasion by a wide variety of disease-causing agents.

The vertebrate immune system has several features which are appropriate for long-lived organisms. It can learn to recognize common environmental pathogens and to respond to them in an enhanced, or secondary, fashion (2). This ability to launch a secondary immune response would be important only to an organism likely to live long enough to encounter a pathogen more than once in its lifetime. In addition, the vertebrate immune system has the ability to recognize a virtually unlimited array of molecular determinants (2). An organism which relies on a long life span for reproductive success must be prepared for encounters with pathogens which can rapidly change their physical appearance. It is this property of the vertebrate immune system, the ability to recognize a tremendous array of molecular determinants, which is the primary focus of this dissertation.

Vertebrate immune systems seem to have evolved a set of effector mechanisms based on fundamental properties of invading pathogens. These effector mechanisms include complement fixation, phagocytosis, and various lymphocyte and leukocyte cytotoxicities (2). Because of the fundamental nature of these effector mechanisms, it is apparently very difficult for pathogens to evolve ways to circumvent them. Thus the limited diversity of effector mechanisms is adequate to insure the destruction of most invading organisms. The ability to recognize and attack a large variety

of different molecular determinants, however, relies on a strategy directly contrasting with that used for immune effector functions. Rather than using a few receptors with broad specificity for invading pathogens, the immune recognition system has millions (or more) of recognition units capable of binding to multiple molecular structures (antigens) on virtually any pathogenic invader.

The immune system of the laboratory mouse has been studied more extensively than that of any other animal. The following discussion, while generally true for most vertebrates, refers specifically to observations made on laboratory mice.

Basic antibody structure (2)

Antigen recognition structures are present on the products of both bone marrow-derived (B) cells and thymus-derived (T) cells. The most abundant immune system antigen-recognition structure, the B-cell antibody, has several molecular forms based on a fundamental four-chain antibody unit. The antibody unit has two identical heavy (H) and two identical light (L) chains each containing variable (V) and constant (C) regions. There are two major classes of light chain C regions (λ and κ) and five classes of heavy chain C regions (μ, γ, α, δ, and ε). In the mouse, the γ and λ classes each include several related C regions designated $\gamma 1$, $\gamma 2a$, $\gamma 2b$ and $\gamma 3$, and $\lambda 1$ and $\lambda 2$ (3). Immunoglobulin classes derive their names from their heavy chain C regions; i.e., IgM has a C_μ region, IgG has a C_γ region, etc. The IgG, IgD, IgE, and cell surface IgM molecules are monomeric antibody units, the serum IgM molecule is constructed from five antibody units joined by a J polypeptide, and the IgA molecule can be monomeric or in oligomeric form with a J polypeptide and the polypeptide secretory piece.

Immunoglobulin polypeptides are made up of repeating regions of homology of about 110 residues in length. Heavy chains have four (IgG, IgD, IgA) or five (IgM, IgE) such units and light chains have two. Antibody V regions are the N-terminal-most heavy and light chain homology units and each V_H region interacts with a single

V_L region to form an antigen-binding site. Regions of exceptional variability, termed hypervariable regions, provide the residues which interact with antigen (4-7). Thus the position and amino acid sequence of the V_L and V_H hypervariable regions determine the antigen-binding specificities of antibodies. The ability of the vertebrate immune system to recognize a vast array of antigenic determinants is directly related to its ability to produce a large number of antibody variable regions and, especially, hypervariable regions.

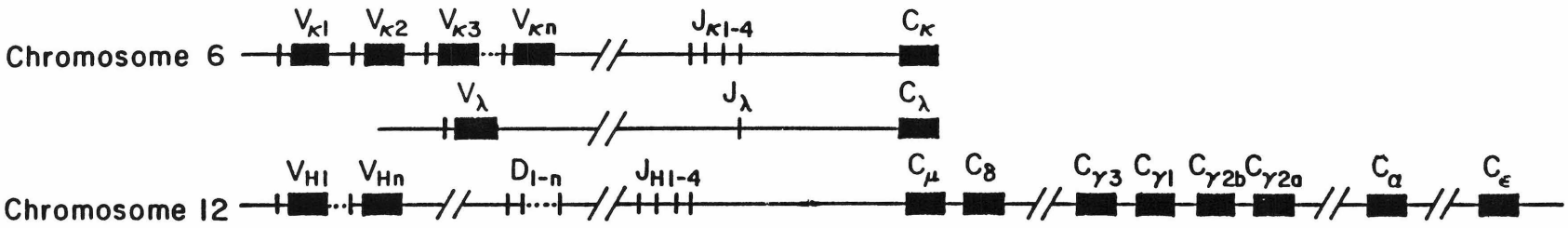
Basic antibody genetics

Antibody heavy and light polypeptides are encoded by three unlinked families of genes: H, λ, and κ. The organization of genes within these families is shown in the figure. The light and heavy chain families each have V and J (joining) gene segments and C region genes. In addition, the heavy chain gene family contains gene segments termed D (diversity) segments. Though D segments are included in the figure, the essential facts of their organization are unknown. V and J gene segments (light chain) or V, D, and J gene segments (heavy chain) are joined together to form complete V genes before transcription occurs (8-10). The rearranged V genes are separated from the C genes by an intervening piece of DNA which is part of the transcription unit. The V and C coding regions are joined by RNA processing to give rise to a completed antibody mRNA (9, 11, 12).

Heavy chain V regions can be expressed with any of the C_H regions by the deletion of intervening DNA (13, 14). Thus in order to express a V_H gene with the $C_{\gamma 1}$ gene, the stretch of DNA containing the C_μ, C_δ, and $C_{\gamma 3}$ genes would have to be deleted (see figure). Each antibody C_H region mediates different immune functions. Monomeric IgM and IgD molecules are cell surface receptors for early B-cells, IgA molecules are the primary antibodies in body secretions, IgG molecules cross the placenta, and IgE molecules cause histamine release from mast cells. In addition, IgG and pentameric IgM are the major serum antibodies and can participate

Figure Legend

The organization of antibody genes in BALB/c mice. H, λ, and κ gene families are each located on different chromosomes with linkage relationships as shown. Gene segments with unknown linkage relationships are separated by slash marks. Intergenic distances are not drawn to scale.


Chromosome 6
$V_{\kappa 1}$ $V_{\kappa 2}$ $V_{\kappa 3}$ $V_{\kappa n}$ $J_{\kappa 1-4}$ C_{κ}
V_{λ} J_{λ} C_{λ}
Chromosome 12
V_{H1} V_{Hn} D_{1-n} J_{H1-4} C_{μ} C_{δ} $C_{\gamma 3}$ $C_{\gamma 1}$ $C_{\gamma 2b}$ $C_{\gamma 2a}$ C_{α} C_{ϵ}

in the classic complement pathway. Thus the immune system can use C_H switching as a way to associate several different effector functions with a single antigen recognition unit (2).

The generation of antibody diversity

The extensive sequence diversity seen in antibody V regions is, as noted above, directly responsible for the ability of the immune system to bind a large number of determinants. Investigations into the molecular aspects of this diversity have given rise to the following ideas about its generation.

Germline diversity

Many antibody gene segments are directly encoded in the germline. There appear to be approximately 200 to 500 light chain V gene segments (15) and possibly the same number of heavy chain V gene segments. There are four J_L (16, 17) and four J_H gene segments (8, 10) and an unknown number of D segments (possibly five to ten) (18).

Combinatorial joining

Since the V_H and V_L gene segments end inside the third hypervariable region, third hypervariable region diversity can be created by combining V_L gene segments with different J_L gene segments or V_H gene segments with different D and J_H gene segments (8, 10, 19, 20). Thus 200 V_L gene segments and 4 J_L gene segments can produce 800 complete V_L genes. In addition, gene segment joining gives rise to diversity at the V_H-D, D-J_H, and V_L-J_L junctions (16, 17, 20-22). This diversity arises in part because gene segments can be joined at any of several nucleotides in their junction regions (21). For example, particular V_L and J_L gene segments can give rise to multiple V_L-J_L junction sequences by varying the V_L-J_L junction point one to three nucleotides towards one or the other gene segment. Junctional diversity is seen as amino acid substitutions, deletions, and insertions (16, 17, 20-22). Because of D segment and D segment junctional diversity, the third hypervariable

region of the heavy chain is far more diverse than any other antibody hypervariable region. There may well be a greater number of third hypervariable region sequences than sequences in all other hypervariable regions combined.

V_H-V_L combinatorial association

This is possibly the most powerful step in antibody diversity amplificaton. If an animal produces 200 (V segment) x 10 (D segment) x 4 (J segment) V_H regions and 200 (V segment) x 4 (J segment) V_L regions, then combinatorial association of heavy and light chains increases the potential diversity to $8000 \times 800 = 6.4 \times 10^6$ antibodies. Probably no other step produces an amplification of this magnitude. If junctional diversity is taken into account in the above calculation (see Summary), the amplificaton due to combinatorial association is even larger.

Somatic mutation

Somatic mutation is, mechanistically, a difficult subject to discuss because it refers to a broad class of mechanisms of immunoglobulin gene alteration. Somatic mutation mechanisms include somatic recombination (23), hypermutation (24-26), and normal somatic mutation (27). In general, however, somatic mutation processes are pictured as changing germline sequences through the alteration, rather than rearrangement, of germline information. Thus processes which randomly introduce base changes, either as a result of normal DNA replication or through special mechanisms, comprise a broad class of postulated somatic mutational processes. Somatic mutation may provide the largest single incremental increase in V region sequences. If an organism makes 10^7 antibodies and each undergoes only one somatic mutation, then somatic mutations contribute an additional 10^7 antibodies to the immune system. An important question about somatic mutation is, how frequent and dependable is it ? If an organism can rely on generating certain sequences or a certain number of sequences somatically, then there will be no selective advantage to the maintenance

of these sequences in the germline. Thus the reliability of somatic mutation will determine the extent to which it is important in the immune system.

V region sequencing experiments

Antibody V region amino acid sequences are relevant to a number of problems in immunology. V region sequencing studies have played important roles in the determination of the three dimensional structure of antibody V regions (7, 28), the molecular basis of antibody-antigen interactions (6, 29), the structure of V region genes (8-10), and the mechanisms of antibody diversity generation (30). Hence the antibodies selected for sequencing studies are determined by the problems being addressed. The following criteria were important in my selection of antibodies for the study of V region diversity. (i) It is most useful to study antibody V regions which are similar in sequence to one another. My study is primarily concerned with the diversity arising from somatic alterations and rearrangements of germline information. Thus it is important to minimize the contribution of germline V segment variability. (ii) Patterns of antibody diversity only become apparent after the comparison of a number of closely related V region sequences. Determining V region sequences similar to sequences already available facilitates the accumulation of large numbers of closely related V region sequences. (iii) A detailed understanding of the origin and function of V region diversity requires an understanding of the function of individual amino acid residues. Studies of antibody three dimensional structure and antigen binding are necessary for such structure-function correlations. (iv) With the advent of recombinant DNA technology, V regions can be studied at the germline as well as at the somatic level. Germline DNA sequencing studies provide information vital to the study of antibody diversity. Thus it is important to select antibodies which are being studied at the DNA level as well as at the protein level.

Using the above criteria, I selected the antibodies binding phosphorylcholine and $\beta(2 \rightarrow 1)$ levan for V region sequence analysis. Levan-binding and phosphorylcholine-

binding V regions have been previously studied by protein sequence analysis (31-36). These V regions have been found to fall into closely related groups which differ from one another by less than six amino acids over their N-terminal 35 residues. Phosphorylcholine-binding and levan-binding antibodies have also been studied by three dimensional structure analysis (7, 37), antigen binding (29, 38), and recombinant DNA techniques (8; S. Crews, unpublished observations). The antibodies binding these two antigens provide an interesting comparison with one another in that the small phosphorylcholine seems to bind to a more heterogeneous collection of antibodies than does the larger levan determinant.

My sequencing studies were done on homogeneous immunoglobulins derived by two procedures: myeloma induction (39) and the hybridoma technique (40). Each, as used in these studies, has its own advantages and faults. Myeloma tumors arise from antibody-secreting B cells in the peritoneal cavity of mice injected intraperitoneally with mineral oil (39). Myeloma sequence data must be examined with the following reservations: (i) Myeloma tumors have been passed through many generations of mice and may have accumulated mutations in their immunoglobulin genes unrelated to normal diversification mechanisms. (ii) Myeloma proteins may contain a different set of phosphorylcholine-specific immunoglobulins than does the normal serum. The C regions of myeloma proteins are predominantly α (37), as compared to the γ and μ found in normal serum. In addition, the set of variable regions expressed in the myeloma population is not identical to that expressed in the serum immunoglobulin population (41). On the other hand, hybridomas are recently derived from normal B cells, and, accordingly, there is less chance that variants arose during hybridoma passaging. Hybridoma immunoglobulins complement those from myeloma tumors in that hybridomas are splenic in origin and produce primarily IgG and IgM immunoglobulins.

The hybridomas which we studied were raised in mice immunized with

phosphorylcholine, while the myelomas came from unimmunized mice. If V region diversity is influenced by the route of antigenic exposure, myeloma proteins may reflect natural V region diversity more accurately than hybridoma proteins. Except as noted above, the diversity in myeloma proteins is probably a result of normal routes of antigenic exposure.

I chose phosphorylcholine-binding antibodies for my sequencing studies partly because of experiments on phosphorylcholine-binding myeloma proteins already under way. These experiments resulted in the six complete myeloma V_H sequences described in chapter one (32). Chapters two and three discuss the data in chapter one in conjunction with sequence data from hybridoma proteins.

The results of a preliminary screening study on the N-terminal 35 residues of 16 phosphorylcholine-binding hybridoma V_L and V_H regions are reported in chapter two (42). It is clear from these sequences that the IgG hybridoma antibodies have significantly more N-terminal diversity than the IgM hybridoma antibodies. The additional diversity seen in IgG antibodies probably results from the selective expansion of both germline and somatically-generated clones initially present at very low levels. This expansion of rare clones may be mediated by antigen and by V region sequence-specific regulation. Such sequence-specific regulation can occur through V region antigenic determinants called idiotypes (43).

Chapter three describes the complete V_H sequences of seven phosphorylcholine-binding hybridoma proteins and the V_H segment sequences of two additional proteins (44). These sequences, along with the sequences from chapter one, define more precisely the differences in diversity between IgM antibodies and IgG and IgA antibodies. The V_H segment diversity of μ chains is seen to be less than that in γ and α chains. even though μ, γ, and α chains have comparably diverse D segments. If these diversity differences between the V regions of different antibody classes

are the result of selection, then it should be possible to identify residues which are involved in such selection. The sequences reported in chapters one and three allow the identification of amino acid substitutions which may reflect this selection process.

Phosphorylcholine-binding antibody D segments are relatively long and contain large numbers of amino acid substitutions, insertions, and deletions. Current studies in our lab of the DNA sequences of germline gene segments will allow an appraisal of the mechanisms necessary to generate this D segment diversity. In addition, a comparison of the sequences of germline phosphorylcholine-binding V_H gene segments to protein V_H sequences will hopefully provide a measure of the contribution of somatic mutation to V_H region diversity.

The second set of antigen-binding proteins, those binding $\beta(2\rightarrow 1)$ levan, is discussed in chapter four (45). These proteins have V_L and V_H regions which differ by only four or less amino acids from prototype sequences, and which have only one amino acid interchange in their D segments. Despite this high degree of sequence homology, however, the V_H and V_L segments may each be encoded by at least two germline genes.

The complete characterization of these levan-binding V_H and V_L regions allows idiotype analysis and structure-function correlations to be carried out using well-defined proteins. Such analyses are basic to an understanding of antibody diversity as they provide information on the roles of individual sequence differences. Knowledge of the function of individual residues is essential if the reasons underlying the selective expansion of germline and somatically generated genes are to be understood. In addition, antibody idiotypes have played a central role in many studies of antibody structure and genetics (se Ref. 46 for reviews). An understanding of the molecular basis of antibody idiotypes is central to the interpretation of such studies.

The appendix contains a technical paper describing the high performance liquid chromatography of phenylthiohydantoin amino acids (47). Phenylthiohydantoin identi-

fication by high performance liquid chromatography is one of the most significant improvements in protein sequence analysis of the past decade. It allows rapid identification and quantitation of all protein amino acids without additional time-consuming and error-prone chemical modifications.

References

1. Feeny, P. In: Coevolution of Animals and Plants. (L. E. Gilbert and P. H. Raven, eds.), pp. 3-19, University of Texas Press, Austin.
2. Hood, L., I. Weissman, and B. Wood. (1978). Immunology. Benjamin-Cummings Publishing Co., Inc., Menlo Park, California.
3. Tonegawa, S., A. Maxam, R. Tizard, O. Bernard, and W. Gilbert. (1978). Proc. Nat. Acad. Sci. U.S.A. **75**, 1485.
4. Wu, T., and E. Kabat. (1970). J. Exp. Med. **132**, 211.
5. Capra, J. D., and J. Kehoe. (1974). Proc. Nat. Acad. Sci., USA **71**, 845.
6. Amzel, L., R. Poljak, F. Saul, J. Varga, and F. Richard. (1974). Proc. Nat. Acad. Sci., USA **71**, 1427.
7. Padlan, E., D. Segal, T. Sparide, D. Davies, S. Rudikoff, and M. Potter. (1973). Nature New Biol. **245**, 165.
8. Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. (1980). Cell **19**, 981.
9. Bernard, O., N. Hozumi, and S. Tonegawa. (1978). Cell **15**, 1133.
10. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. (1980). Nature **286**, 676.
11. Early, P., M. Davis, D. Kaback, N. Davidson, and L. Hood. (1979). Proc. Nat. Acad. Sci., USA **76**, 857.
12. Sakano, H., J. Rogers, K. Hüppi, C. Brack, A. Traunecker, R. Maki, R. Wall, and S. Tonegawa. (1979). Nature **277**, 627.
13. Davis, M., K. Calame, P. Early, D. Livant, R. Joho, I. Weissman, and L. Hood. (1980). Nature **283**, 733.
14. Kataoka, T., T. Kawakami, N. Takahashi, and T. Honjo. (1980). Proc. Nat. Acad. Sci., USA **77**, 919.
15. Seidman, J., A. Leder, M. Nau, B. Norman, and P. Leder. (1978). Science **202**, 11.

16. Max, E., J. Seidman, and P. Leder. (1979). Proc. Nat. Acad. Sci., USA **76**, 3450.

17. Sakano, H., K. Hüppi, G. Henrich, and S. Tonegawa. (1979). Nature (Lond.) **280**, 288.

18. Hood, L., M. Davis, P. Early, K. Calame, S. Kim, S. Crews, and H. Huang. (1980). Cold Spring Harbor Symp. Quant. Biol. **45**, in press.

19. Weigert, M.,L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. (1978). Nature **276**, 785.

20. Schilling, J., B. Clevinger, J. M. Davie, and L. Hood. (1980). Nature (Lond.) **283**, 35.

21. Weigert, M., R. Perry, D. Kelley, T. Hunkapiller, J. Schilling, and L. Hood. (1980). Nature (Lond.) **283**, 497.

22. Johnson, N., P. Gearhart, R. Douglas, and L. Hood. (1980). Manuscript in preparation.

23. Gally, J. A., and G. M. Edelman. (1970). Nature **227**, 341.

24. Jerne, N. K. (1971). Eur. J. Immunol. **1**, 1.

25. Brenner, S., and C. Milstein. (1966). Nature **211**, 242.

26. Baltimore, D. (1974). Nature **248**, 409.

27. Cohn, M., B. Blomberg, W. Geckeler, W. Rashke, R. Riblet, and M. Weigert. (1974). In: The Immune System: Genes, Receptors, Signals. (E. Sercarz et al., eds.), p. 89, Academic Press, New York.

28. Poljak, R., L. Amzel, H. Avey, B. Chen, R. Phizackerley, and F. Saul. (1973). Proc. Nat. Acad. Sci., USA **70**, 3305.

29. Segal, D., E. Padlan, G. Cohen, S. Rudikoff, M. Potter, and D. Davies. (1974). Proc. Nat. Acad. Sci., USA **71**, 4298.

30. Hood, L., E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg, and J. Schilling. (1976). Cold Spring Harbor Symp. Quant. Biol. **41**, 817.

31. Barstad, P., J. Hubert, M. Hunkapillar, A. Goetze, J. Schilling, B. Black, B. Eaton, J. Richards, M. Weigert, and L. Hood. Eur. J. Immunol. **8**, 497.

32. Hubert, J., N. Johnson, J. Greengard, B. Eaton, and L. Hood. (1980). Manuscript in preparation.

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

34. Rudikoff, S., and M. Potter. (1976). Proc. Nat. Acad. Sci., USA. **73**, 2109.

35. Vrana, M., S. Rudikoff, and M. Potter. (1979). J. Immunol. **122**, 1905.

36. Vrana, M., S. Rudikoff, and M. Potter. (1978). Proc. Nat. Acad. Sci., USA **75**, 1957.

37. Potter, M. (1977). Adv. Immunol. **25**, 141.

38. Cisar, J., E. Kabat, J. Liao, and M. Potter. (1974). J. Exp. Med. **139**, 159.

39. Potter, M. (1972). Physiol. Rev. **52**, 631.

40. Kohler, G., and C. Milstein. (1975). Nature (Lond.) **256**, 495.

41. Hood, L., P. Barstad, E. Loh, and C. Nottenburg. (1974). In: The Immune System: Genes, Receptors, Signals. (E. Sercarz et al., eds.), p. 119, Academic Press, New York.

42. Gearhart, P., N. Johnson, R. Douglas, and L. Hood. (1980). Manuscript in preparation.

43. Jerne, N. (1973). Scientific American **229**, 52

44. Johnson, N., P. Gearhart, R. Douglas, and L. Hood. (1980). Manuscript in preparation.

45. Johnson, N., J. Slankard, L. Paul, and L. Hood. (1980). Manuscript in preparation.

46. Möller, G. (ed.) (1977). Idiotypes on T and B cells. Immunol. Rev. **34.**

47. Johnson, N., M. Hunkapillar, and L. Hood. (1979). Anal. Biochem. **100**, 335.

Chapter 1

The complete amino acid sequences of six V_H regions from myeloma proteins binding phosphorylcholine

This manuscript will be submitted to Molecular Immunology

Introduction

The enormous diversity available in the normal antibody repertoire remains one of immunology's most fascinating aspects. The search for the mechanisms responsible for this diversity has been pursued for well over a decade and many important features of antibody variability have been described. Antibody molecules are composed of heavy (H) and light (L) chains, each containing variable (V) and constant (C) regions. V regions are constructed from several gene segments encoded separately in the germline. Complete V_L regions are created by the joining of V_L and J_L (joining) gene segments while V_H, D (diversity) and J_H segments are joined to form complete V_H regions (1, 2).

Antibody V segments are encoded by multiple germline genes (3) as are J segments (1, 4, 5) and, presumably, D segments (1). The V_L-J_L junction and the V_H-D-J_H junctions occur in the V_L and V_H third hypervariable regions. Thus, third hypervariable region diversity can be generated through the combinatorial joining of V_L and J_L gene segments (6, 7) and V_H, D and J_H gene segments (1, 8). Additional diversity may arise through somatic mutations in germline genes (9).

In this paper we describe the complete sequences of six V_H regions from phosphorylcholine-binding myeloma proteins. The diversity seen in these sequences, and in sequences from phosphorylcholine-binding V_H regions reported elsewhere (10-13), provide an assessment of the V_H phenotype of the antibody response to phosphorylcholine. Studies at the DNA level currently under way in our laboratory will determine the germline genotype of the phosphorylcholine-binding myeloma V_H regions. A comparison of protein sequence phenotypic information to germline DNA sequences will provide a measure of the contribution of somatic mutation mechanisms to diversity. In addition, the D regions from phosphorylcholine-binding myeloma V_H regions are quite variable and provide examples of the kinds of diversity which can arise from V_H-D-J_H joining.

Materials and Methods

Myeloma heavy chain isolation

Myeloma tumors were obtained from the Salk Institute and grown in ascites form in BALB/c and CDF1 mice primed one month previously with 0.5 ml mineral oil. Ascites fluid was obtained from myeloma-bearing mice, filtered through Miracloth to remove macroscopic lipid complexes, and precipitated in 40% saturated $(NH_4)_2SO_4$ at 4°C. The precipitate was mildly reduced in 0.2 M Tris buffer, 2 mM EDTA, 10 mM dithiothreitol, pH 8.2, for 1 hr at room temperature and then alkylated for 1 hr at 4°C by the addition of iodoacetamide to 25 mM. The reduced and alkylated protein was applied to a phosphorylcholine-sepharose 4B column (14), eluted with 10^{-3} M phosphorylcholine, desalted over G-25 in 0.2 M NH_3 and lyophilized. These affinity column preparations contained purified myeloma proteins and were loaded on an LKB Ultragel AcA34 column equilibrated in 3 M guanidine, 0.2 M NH_5CO_3. The heavy and light chain peaks were pooled separately and desalted on G-25, 0.2 M NH_3.

Cyanogen bromide digests

Purified heavy chains were digested with 2.5% cyanogen bromide in 70% HCO_2H for 24 hr at 4°C. These digests were then diluted, lyophilized twice and applied to a G-100 column equilibrated in 5 M guanidine, 0.2 M NH_5CO_3. Column peaks were desalted in 0.2 M NH_2OH or 5% HCO_2H. In some cases heavy chains were fully reduced and alkylated in the presence of 5 M guanidine before the CNBr digests. These heavy chains were chromatographed on Ultragel AcA54 in 3 M guanidine, 0.2 M NH_5CO_3.

Automated sequence analysis

Automatic sequence analysis was done using Beckman 890C and modified 890C sequenators (15), and a home-made solid phase sequenator. A Durrum D-500 Automatic

Amino Acid Analyzer, Hewlitt-Packard Gas Chromatograph, and Waters Associates High Performance Liquid Chromatograph (HPLC) were used to analyze sequencer samples (16, 17). Heavy chain variable regions are numbered as in Kabat et al. (18).

Peptide map analyses

Trypsin or chymotrypsin digests were performed at an enzyme:substrate ratio of 1:100 in 1% NH_5CO_3 for 1 hr at 37°C. Two-dimensional peptide paper maps at pH 3.5 were done using the procedure of Katz et al. (19). Analytical maps were dipped in 80:20:8 ethanol:acetic acid:collidine, 0.1% ninhydrin and developed for 12 min at 100°C. Prepatory maps were sprayed lightly with 0.1% ninhydrin in ethanol, developed at 100°C, and the spots cut out and eluted immediately upon visualization. Acidic peptides were eluted in 0.5 M NH_4OH while basic peptides were eluted in 10% HCO_2H. Electrophoresis at pH 6.5 was done in some cases to verify the assignment of amide residues on Eastman chromagram cellulose plates in 10:0.4:90 pyridine:acetic acid:water. After electrophoresis at 1000 volts for 30 min on a Desaga flat plate electrophoresis apparatus, spots were visualized with 0.1% ninhydrin in acetone. Spots staying near the origin contain amide residues while those migrating towards the anode contain acids. Amino acid analysis was done on a Durrum D-500 amino acid analyzer after 18 hr hydrolysis of the peptides at 108°C in 6 N HCl.

Cyanogen bromide fragment precipitation

CN2 (see **Results**) cyanogen bromide fragments were desalted on P2 in 5% HCO_2H, lyophilized, and redissolved in 10% acetic acid. Concentrated NH_4OH was added until a white precipitate formed and the precipitate was spun out and lyophilized. This procedure is from Rudikoff and Potter (10).

Results

Isolation of cyanogen bromide fragments

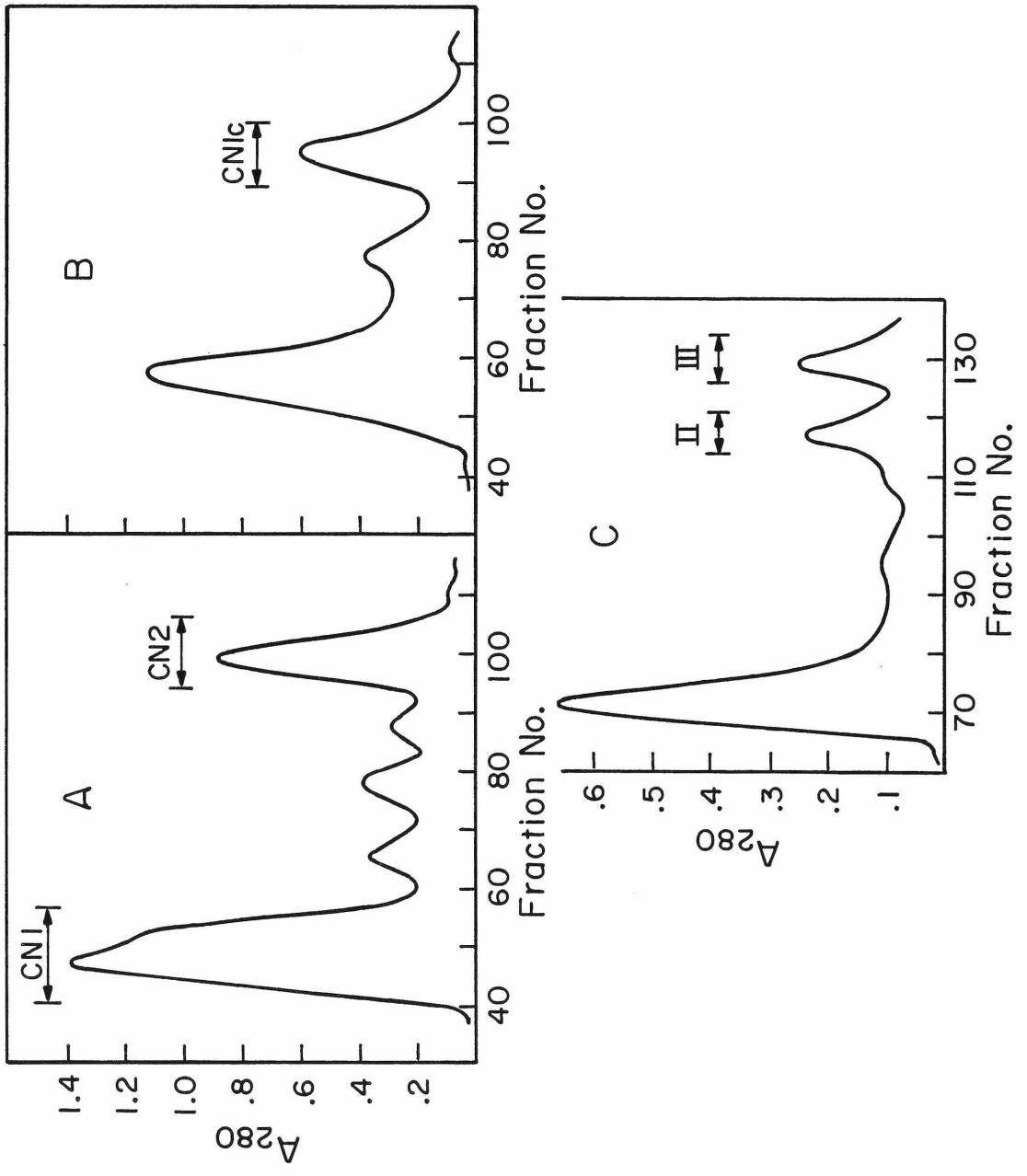
Figure 1a shows a typical G-100 profile of a mildly reduced and alkylated

Figure 1. Gel chromatography of CNBr fragments.

(a) CNBr digest of a mildly reduced and alkylated heavy chain chromatographed on G-100 in 5 M guanidine, 0.2 M NH_5CO_3.

(b) CN1 peak from (a) fully reduced and alkylated and rechromatographed on G-100 in 5 M guanidine, 0.2 M NH_5CO_3.

(c) CNBr digest of a fully reduced and alkylated heavy chain chromatographed on AcA54 in 3 M guanidine, 0.2 M NH_5CO_3.


A
B
C
CN1
CN2
CNIc
II
III
A_{280}
Fraction No.
1.4
1.2
1.0
.8
.6
.4
.2
.6
.5
.4
.3
.2
.1
40
60
80
100
70
90
110
130

heavy chain CNBr digest. The first peak, CN1, contains a mixture of peptides including the peptide from 82a into the C region. This peak was desalted in 0.2 M NH_4OH, lyophilized, and reduced and aklylated in the presence of 5 M guanidine HCl. Figure 1b shows a profile of CN1 fully reduced and alkylated and rerun on G-100. The peak labeled CN1c contains only the peptide from 82a into the C region.

The CN2 peak (Fig. 1a) was desalted in 5% HCO_2H due to its insolubility in dilute base and precipitated as in **Materials and Methods.** The precipitate contained only the fragment from 35 to 82 and will be referred to as CN2p. This CNBr fractionation procedure is the same as that in Rudikoff and Potter (10).

In some cases the intact heavy chains were fully reduced and aklylated prior to CNBr digestion and run on an AcA54 column in 3 M guanidine, 0.2 M NH_5CO_3. A column profile is shown in Fig. 1c. The V region fragment from 82a into the C region can be isolated cleanly as CNII while the fragment from 35 to 82 is contained as a mixture in CNIII.

Sequence determination

The sequencing strategies discussed below are outlined in Fig. 2 for the S107 V_H region.

Purified myeloma heavy chains were degraded 35 cycles by automatic sequenator to determine the N-terminal portion of the molecule. A plot of PTH-amino acid yield vs. cycle number is shown in Fig. 3 for the protein M511H. Ser' and Thr' refer to derivatives of serine and threonine resulting from secondary reactions during sequencing. While data condensed and presented as in Fig. 3 are more difficult to interpret than actual HPLC chromatograms, the M511H sequence is clearly visible. The average repetitive yield calculated using leucines at cycles 4, 11, and 20 is 93%.

The CN2p fragments (positions 32 to 82) were usually sequenced for at least 46 steps with the remainder of the fragment being determined from peptides. Compositions were obtained from CN2p tryptic or chymotryptic peptides and the ultimate

Figure 2. The sequencing strategy for the S107 V_H region. All myeloma V_H regions were sequenced in an identical fashion with the exception of certain regions depicted in Figure 4.

——→ Sequence determined by 890C automatic sequenator

-----> Sequence determined by solid phase sequenator

[ ] Tryptic peptide

<—> Peptide compositional data.

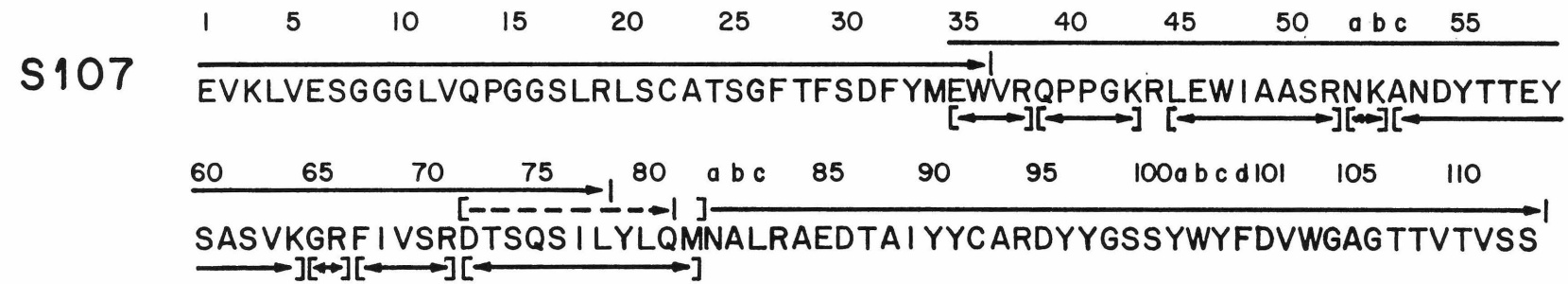
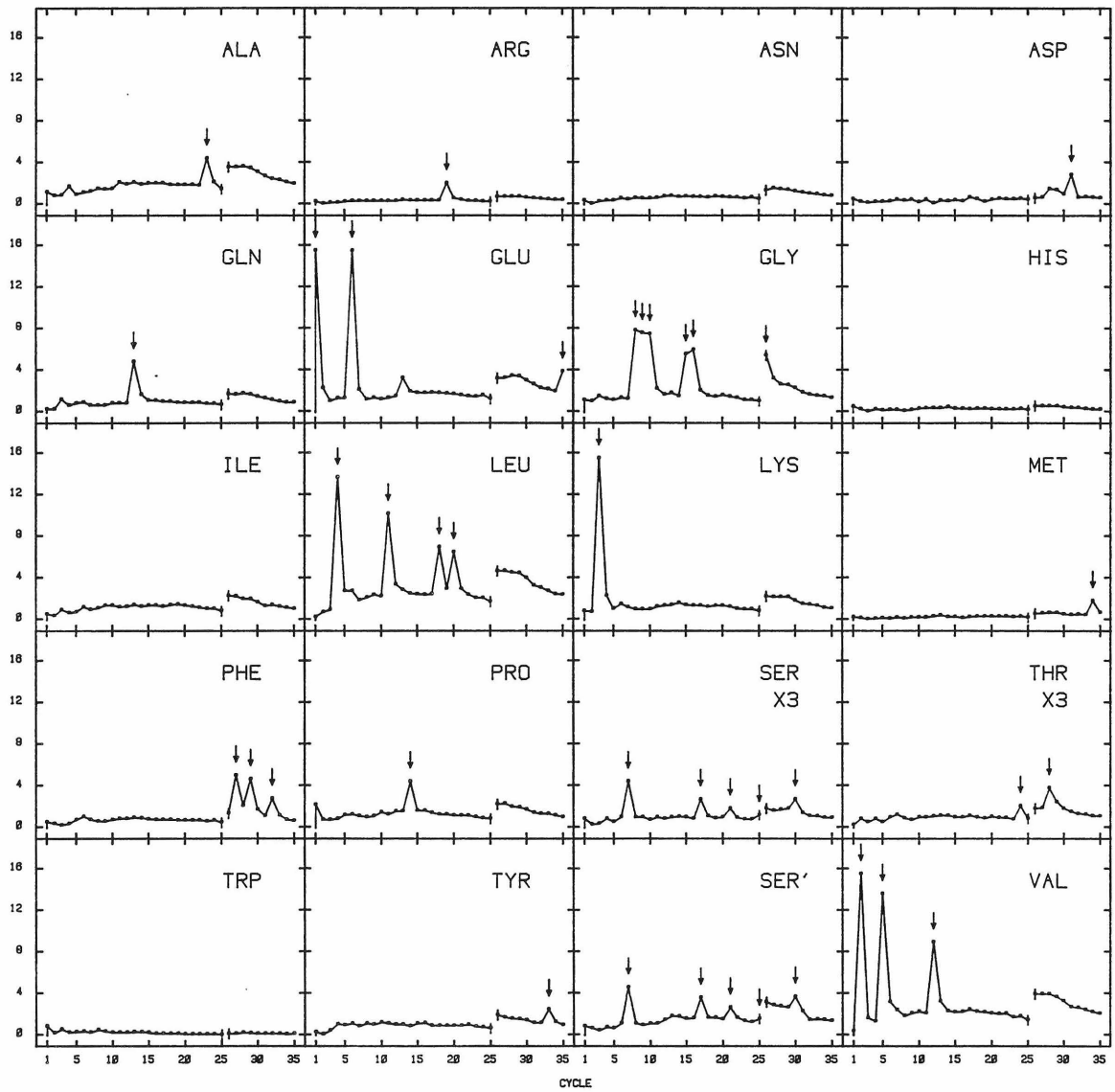

S107
1 5 10 15 20 25 30 35 40 45 50 a b c 55
EVKLVESGGGLVQPGGSLRLSCATSGFTFSDFYMEWVRQPPGKRLEWIAASRNKANDYTTEY
60 65 70 75 80 a b c 85 90 95 100a b c d 101 105 110
SASVKGRFIVSRDTSQSILYLQMNALRAEDTAIYYCARDYYGSSYWYFDVWGAGTTVTVSS

Figure 3. The N-terminal sequence of M511H. Ordinate values are proportional to the absorbance at 254 nanometers of the phenylthiohydantoin-amino acids at each sequenator cycle. Cycle number is plotted on the abscissa. Arrows indicate the M511H N-terminal amino acid sequence. SER' is a secondary reaction product of serine, and serine and threonine ^{254}A values are multiplied X3. The M511 heavy chain used in this sequenator run was mildly reduced and alklyated, and as a result no cystine was seen at position 22. Cystine is not plotted in this figure.


ALA
ARG
ASN
ASP
GLN
GLU
GLY
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
X3
THR
X3
TRP
TYR
SER'
VAL
CYCLE

tryptic peptides were analyzed by automatic sequenator. In most cases the penultimate glutamine was verified by pH 6.5 electrophoresis. A summary of the CN2p sequencing strategies is shown in Fig. 4.

CN1c fragments (residues 82a into the C region) were analyzed for 37 cycles by automatic sequenator. Figure 5 shows a plot of the HPLC data for H8 CN1c. The alanines at residues 5, 14, and 29 give an average repetitive yield of 94%.

Discussion*

Figure 6 presents the complete V_H sequence data from the phosphorylcholine-binding myeloma proteins reported here and elsewhere (10, 11). These sequences have played an important role in the development of ideas about antibody diversity (20). These myeloma V_H sequences exhibit two kinds of diversity: scattered amino acid substitutions in the V segment and extensive substitutions and size changes in the D segment. As germline T15 V_H group gene sequences become available, it will be possible to determine the extent of somatic sequence variation within this set of proteins. In addition, the D segment changes provide important insight into the kinds of DNA rearrangements that must occur to generate D segment diversity. Together with other antiphosphorylcholine V_H sequences (12, 13), the myeloma data presented in Figure 6 provide one of the primary models for the study of antibody diversity.

Summary

Closely related sets of proteins play a crucial role in the study of V region diversity. In this paper, we describe six complete V_H region sequences from myeloma proteins with specificity for phosphorylcholine.

---

*These data are discussed more completely in Chapters 2 and 3 in conjunction with sequences presented there.

Figure 4. The sequencing strategies for the myeloma CNB4 fragments from 35 to 82 (CN2p fragments).

* indicates amide verification by pH 6.5 electrophoresis (see **Materials and Methods**).

( ) Chymotryptic peptide.

Solid lines indicate identity to S63, and amino acid differences are designated by the one-letter code of Dayhoff (22).

All other symbols are as in Figure 2.

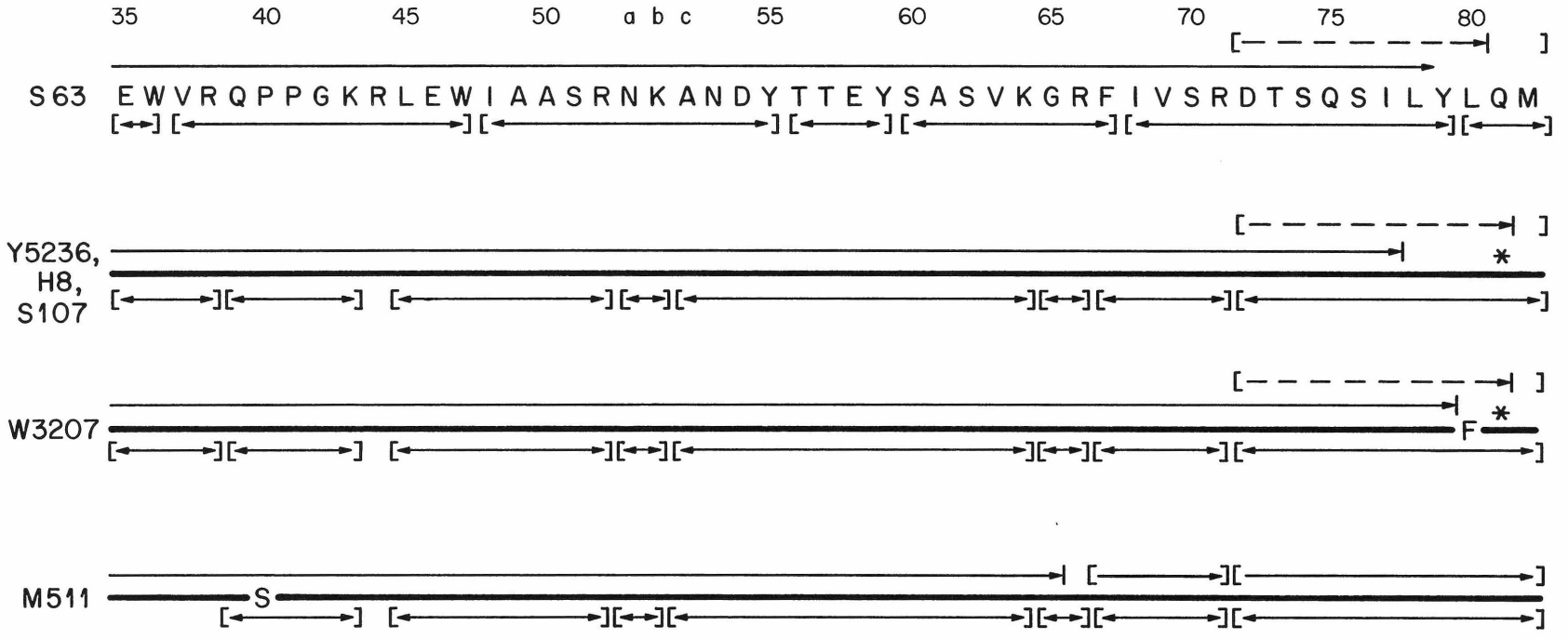

35
40
45
50
a b c
55
60
65
70
75
80
S63 E W V R Q P P G K R L E W I A A S R N K A N D Y T T E Y S A S V K G R F I V S R D T S Q S I L Y L Q M
Y5236,
H8,
S107
*
W3207
F
*
M511
S

Figure 5. The sequence of the N-terminal 37 residues of the H8 CNBr peptide beginning at residue 83. Ordinate, abscissa, and arrows are as in Figure 3. SER' and THR' are secondary reaction products of serine and threonine, and serine, threonine, SER', THR', and tryptophane are all multiplied by the indicated scale factors. AMC stands for amidomethylcystine, the cystine derivative formed by complete reduction and alkylation of disulfide bonds. No methionine is present in this sequence and methionine is not plotted in Figure 5.

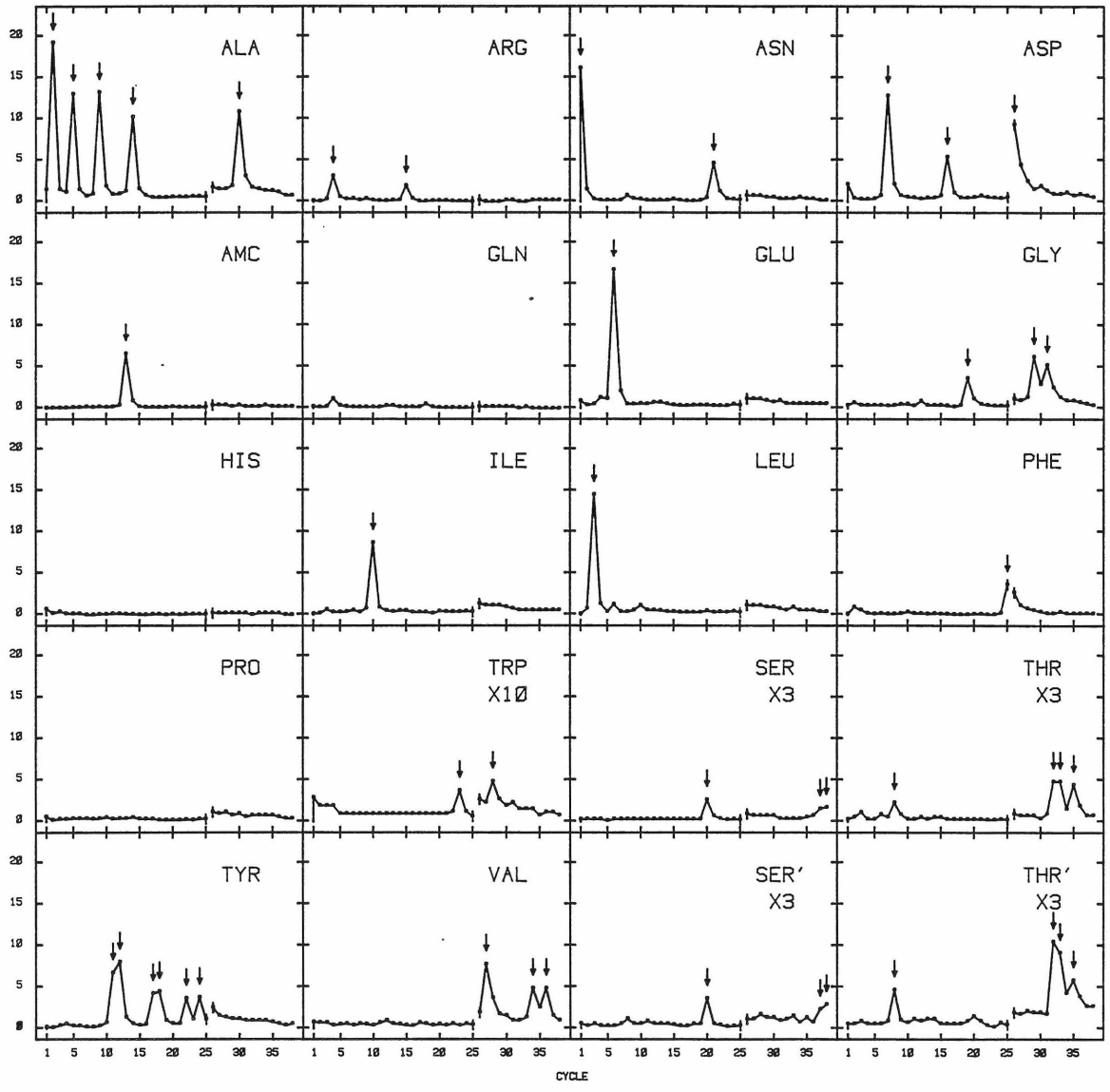
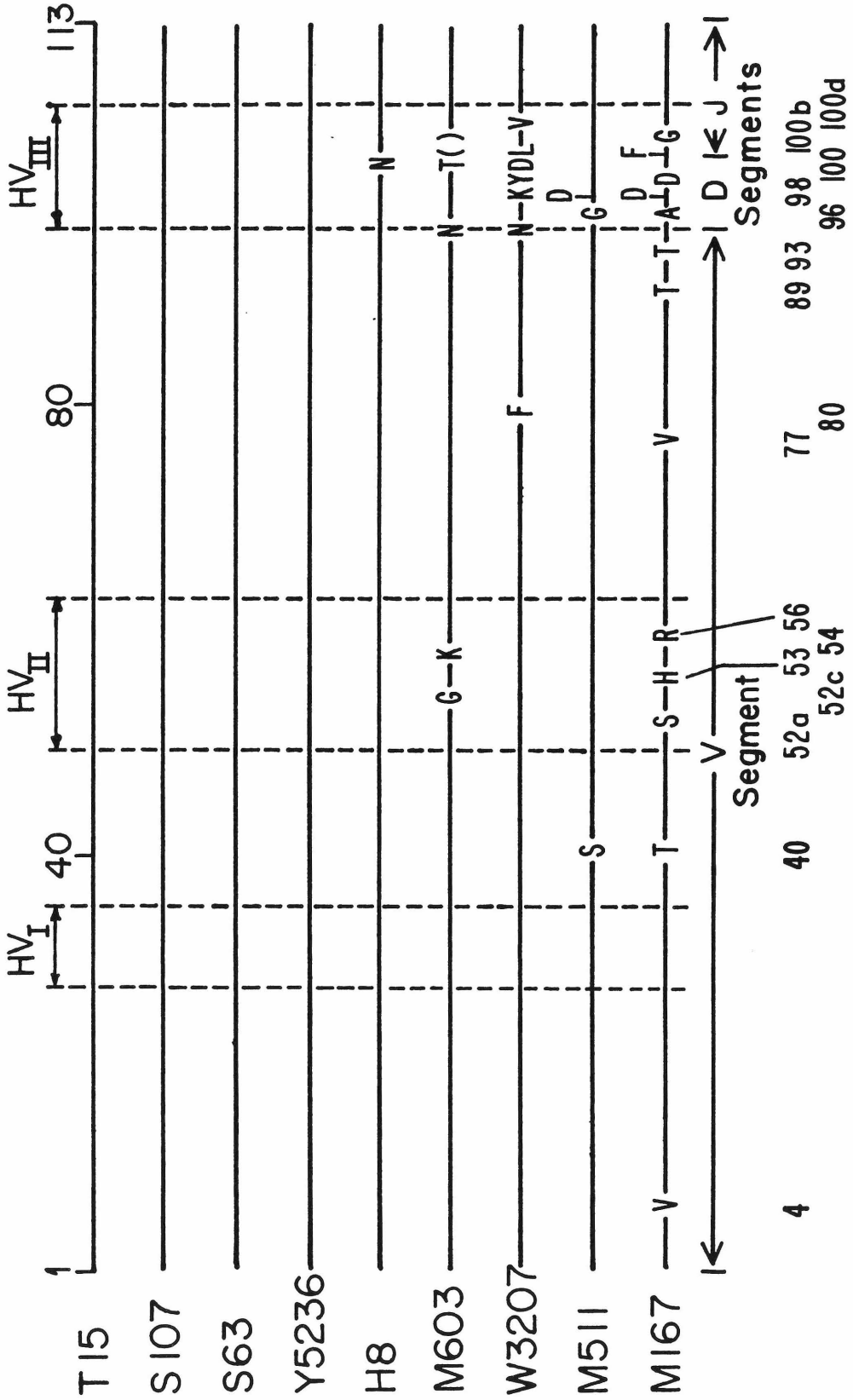

ALA
ARG
ASN
ASP
AMC
GLN
GLU
GLY
HIS
ILE
LEU
PHE
PRO
TRP
X10
SER
X3
THR
X3
TYR
VAL
SER'
X3
THR'
X3
CYCLE

Figure 6. A schematic representation of the complete V_H sequence data from phosphorylcholine-binding myeloma proteins. Straight lines indicate identity to the T15 prototype sequence; a vertical bar indicates an insertion; ( ) indicates a deletion; and substitutions are designated by the one-letter code of Dayhoff (22). The position of each amino acid difference is shown at the bottom of the figure except in the third hypervariable region where every other position is numbered. Hypervariable regions are shown at the top of the figure (HV_I, HV_{II}, HV_{III}) and V, D and J segment boundaries are shown at the bottom. M603 is from ref. 10 and T15 and M167 are from ref. 11.


HV_I
HV_II
HV_III
1
40
80
113
T15
S107
S63
Y5236
H8
M603
W3207
M511
M167
G—K
N—KYDL—V
D
G
D F
T—T—A—D—G
S—H—R
V
T
V
N
T( )
F
S
Segment
4
40
52a 53 56
52c 54
77
80
89 93 98 100b
96 100 100d
Segments

Acknowledgements

We thank Richard Douglas and Tim Hunkapiller for computer graphics. Note added in proof: Since the completion of this work, the V_H region sequence of M511 has been published in a separate study (21). This M511 V_H sequence agrees with the sequence reported here.

References

1. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. Cell **19:** 981.
2. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell **15:** 1133.
3. Seidman, J., Leder, A., Nau, M., Norman, B. and Leder, P. (1978) Antibody diversity. Science **202:** 11.
4. Max, E., Seidman, J. and Leder, P. (1979) Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. Proc. Nat. Acad. Sci. USA **76:** 3450.
5. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature **286:** 676.
6. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. and Hood, L. (1978) Rearrangement of genetic information may produce immunoglobulin diversity. Nature **276:** 785.
7. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling J. and Hood, L. (1980) The joining of V and J gene segments creates antibody diversity. Nature **283:** 497.
8. Schilling, J., Clevinger, B., Davie, J. and Hood, L. (1980) Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. Nature **283:** 35.
9. Cohn, M., Blomberg, B., Geckler, W., Rashke, W., Riblet, R. and Weigert, M. (1974) First order considerations in analyzing the generation of diversity. In The Immune System: Genes, Receptors, Signals. (Edited by E. Sarcarz, A. Williamson and C. Fox). Academic Press, New York. p. 89.

10. Rudikoff, S. and Potter, M. (1974) Variable region sequence of the heavy chain from a phosphocholine-binding myeloma protein. Biochemistry **13:** 4033.

11. Rudikoff, S. and Potter, M. (1976) Size differences among immunoglobulin heavy chains from phosphocholine-binding proteins. Proc. Nat. Acad. Sci. USA **73:** 2109.

12. Johnson, N., Gearhart, P., Douglas, R. and Hood, L. (1980) The amino acid sequences of nine V_H regions from hybridomas binding phosphorylcholine. Manuscript in preparation.

13. Gearhart, P., Johnson, N., Douglas, R. and Hood, L. (1980) IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. Manuscript in preparation.

14. Chesebro, B. and Metzger, H. (1972) Affinity labeling of a phosphorylcholine-binding mouse myeloma protein. Biochemistry **11:** 766.

15. Hunkapiller, M. and Hood, L. (1978) Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier, and high pressure liquid chromatography. Biochemistry **17:** 2124.

16. Johnson, N. D., Hunkapiller, M. and Hood, L. (1979) Analysis of phenylthiohydantoin amino acids by high performance liquid chromatography on DuPont Zorbax cyanopropylsilane columns. Anal. Biochem. **100:** 335.

17. Barstad, P., Farnsworth, V., Weigert, M., Cohn, M. and Hood, L. (1974) Mouse immunoglobulin heavy chains are coded by multiple germline variable region genes. Proc. Nat. Acad. Sci. USA **71:** 4096.

18. Kabat, E., Wu, T. T. and Bilofsky, H. (1979) Sequences of Immunoglobulin Chains. U.S. Department of Health, Education and Welfare. Public Health Service, National Institutes of Health.

19. Katz, A., Dreyer, W. J. and Anfinson, C. B. (1959) Peptide separation by two-dimensional chromatography and electrophoresis. J. Biol. Chem. **234:** 2897.

20. Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P., Fuhrman, J., Johnson, N., Kronenberg, M. and Schilling, J. (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.

21. Robinson, E. and Apella, E. (1980) Complete amino acid sequence of a mouse immunoglobulin α chain (MOPC 511). Proc. Nat. Acad. Sci. USA **77:** 4909.

22. Dayhoff, M. (1976) Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Silver Spring, Maryland, **5:** 189.

Chapter 2

IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts

This manuscript will be submitted to the Journal of Experimental Medicine

IgG ANTIBODIES TO PHOSPHORYLCHOLINE EXHIBIT MORE DIVERSITY THAN THEIR IgM COUNTERPARTS*

By PATRICIA J. GEARHART, NELSON D. JOHNSON, RICHARD DOUGLAS, and LEROY HOOD

(From the Department of Embryology, Carnegie Institution of Washington, 115 W. University Parkway, Baltimore, Maryland 21210; and the Division of Biology, California Institute of Technology, Pasadena, California 91125)

Running Title: Hybridoma Antibodies Binding Phosphorylcholine

Footnotes

*Supported by U.S. Public Health Service grants CA-25507 and AI-10781.

[1] Abbreviations used in this paper: L, light; H, heavy; V, variable; C, constant.

Introduction

One goal in molecular immunology is to understand the mechanisms responsible for antibody gene diversification. The antibody molecule is composed of two light (L)[1] and two heavy (H) polypeptides, both of which have a variable (V) and a constant (C) region. Antibodies binding to the simple hapten phosphorylcholine have served as a model for the cellular and molecular analyses of the immune response for several reasons. First, a closely related set of eleven myeloma proteins that bind phosphorylcholine have been isolated from BALB/c mice (1). Complete amino acid sequence analyses of the V_H regions from nine of the proteins reveal that four V_H regions are identical, and the remainder differ from these by one to 13 amino acid substitutions and occasional sequence insertions or deletions (2, 3). Approximately two-thirds of the substitutions fall within the third hypervariable region. The phosphorylcholine V_H regions and other sets of closely related variable regions have permitted us to place important constraints on mechanisms of antibody diversification (4-7). Second, anti-idiotypic sera have been raised against several of the myeloma proteins binding phosphorylcholine. These antisera have revealed that the majority of anti-phosphorylcholin antibodies from BALB/c mice have the idiotype of the myeloma protein T15 (8-11). A variety of interesting analyses have subsequently been carried out on the developmental expression (12), genetic mapping (13, 14), and regulation (15-18) of this predominant idiotype.

Because the initial amino acid sequence analyses were carried out on myeloma V regions, several concerns arise about the generality of the observed diversity patterns. 1) Myeloma tumors have been passed through many generations of mice and may have accumulated somatic mutations in their antibody genes unrelated to normal diversification mechanisms. 2) Myeloma proteins contain a different subset of phosphorylcholine-specific immunoglobulins than does the normal serum. The constant

regions of myeloma proteins are predominantly α (1), as compared to the γ and μ found in normal serum. In addition, the set of variable regions expressed in the myeloma population is not identical to that expressed in the normal serum immunoglobulin population (19). In order to circumvent these limitations of the myeloma system, we have employed the hybridoma technique of Köhler and Milstein (20) to obtain homogeneous antibodies from normal spleen cells in BALB/c mice immunized against phosphorylcholine. We reasoned that hybridoma cells are recently derived from normal B cells, and, accordingly, there would be less chance for somatic variants to arise during the more limited number of passages of the hybridoma cell lines. In addition, the hybridoma technique appears to accurately reflect the class distribution of splenic B cells. We report here the N-terminal amino acid sequences for the V_H and V_L regions of 16 hybridoma antibodies which bind phosphorylcholine. The most striking observation is that the IgG antibodies exhibit far more diversity than their IgM counterparts. The data also suggest that somatic variation may play a role in V region diversification.

Materials and Methods

Mice and immunizations

BALB/cJ mice were purchased from the Jackson Laboratory, Bar Harbor, Maine. Mice at eight weeks of age were immunized intraperitoneally with 0.1 mg of phosphorylcholine-hemocyanin (11) in complete Freund's adjuvant. One to two months later (four days prior to cell fusion), they were injected intraperitoneally with 0.1 mg of phosphorylcholine-hemocyanin in complete Freund's adjuvant.

Cell fusion

The cells from individual spleens were added to 10-20 x 10^6 SP2/0-Ag14 cells, which do not secrete immunoglobulin (21), and then were fused according to Kennett

et al. (22). Hybrid cells were cloned in agar. Hybrid cell lines producing antibody to phosphorylcholine were produced by the same procedure in two other laboratories. The M8(1F8), G5(1B6), G6(2F5), G8(2B4), G9(1G9), G10(2C5), G14(1B10), G15(24B5), and G16(1D8) hybridomas were obtained from J. Hurwitz (23). The M3(63b4) and M4(64a5) hybridomas were obtained from B. Clevinger and J. Davie, Washington University, St. Louis, Missouri, and were generated from adult mice which were suppressed for expression of the T15 idiotype by treatment with anti-T15 serum at birth. The numbers in parentheses refer to the nomenclature of the donor laboratory.

Radioimmunoassays for C_H region and idiotype

Antibodies against phosphorylcholine from hybridoma culture fluids were bound to phosphorylcholine-bovine serum albumin-coated plastic microtiter plates. The microtiter plates were then rinsed and ^{125}I rabbit anti-murine Fab, anti-μ, anti-γ1, or anti-γ3 antibody was added to determine the class of the phosphorylcholine-specific hybridoma immunoglobulins (24). Antibody with the T15 idiotype was detected by the inhibition of the binding of ^{125}I-T15 to plastic tubes coated with A/He anti-T15 serum (11).

Antibody purification

Hybridoma cells were grown as ascites in BALB/c mice, and the ascites fluids were precipitated at 50% saturation with ammonium sulfate and resuspended in 0.017 M boric acid, 0.03 M sodium borate, 0.15 M sodium chloride, pH 8.0. The antibody was then absorbed to a phosphorylcholine-glycyltyrosine-Sepharose column (25), and eluted with 10^{-2} M phosphorylcholine.

Equilibrium dialysis

Phosphoryl(methyl-^{14}C)choline was purchased from New England Nuclear, Boston, Massachusetts. Antibodies were exhaustively dialyzed against 0.17 M boric

acid, 0.03 M sodium borate, 0.15 M sodium chloride, pH 8.0 with 2% sodium azide to remove phosphorylcholine bound during affinity chromatography. Equilibrium dialysis was performed at 25°C in 0.1 ml compartments at an antibody concentration of 1.5-3.5 mg/ml. Protein molarity was determined by absorbance at 280 nm using $E^{0.1\%}_{1\ cm} = 1.37$, and a molecular weight of 150,000 for IgG and IgA, and 170,000 for IgM. The data were plotted by the method of Scatchard (26) to calculate number of sites and by the Sips distribution function (27) to calculate K_a. A least squares fit was used to obtain the best line.

Protein sequence determination

The N-terminal amino acid sequences of light and heavy chains were determined by two different procedures. 1) Individual antibodies were dialyzed against 0.1 M ammonium chloride and the light and heavy chains were sequenced simultaneously. The residue assignments to light and heavy chains were made based on homology considerations. Because of the high degree of sequence conservation among these immunoglobulins, and because of our ability to quantify precisely the phenylthiohydantoin derivatives (28), we are confident of the data obtained using this rapid approach to N-terminal sequence analysis. Immunoglobulins G8 and G15 were analyzed in this manner. 2) Antibodies were reduced in 8 M guanidine, 50 mM dithiothreitol, and 0.5 M Tris at pH 8.2 at 37°C for 1 h, and alkylated for 1 h at 4°C by the addition of 110 mM recrystallized iodoacetamide. The heavy and light chains were then separated on Ultrogel AcA34 (LKB Instruments, Inc., Rockville, Maryland) in 3 M guanidine-0.2 M ammonium bicarbonate, and the protein fractions were desalted on Sephadex G-25 in 0.2 M ammonium hydroxide. The desalted heavy and light chain pools were lyophilized and individually subjected to automatic amino

acid sequence analysis on a modified 890B Beckman sequencer (29) or, in some cases, only the heavy chain was sequenced. If the heavy chain alone was sequenced, the light chain sequence was obtained by subtraction from the alternatives determined as described in procedure 1). The M1, M2, M3, M4, M5, M6, M8, G3, G6, G9, G10, G14 and G16 immunoglobulins were analyzed in this fashion.

Results

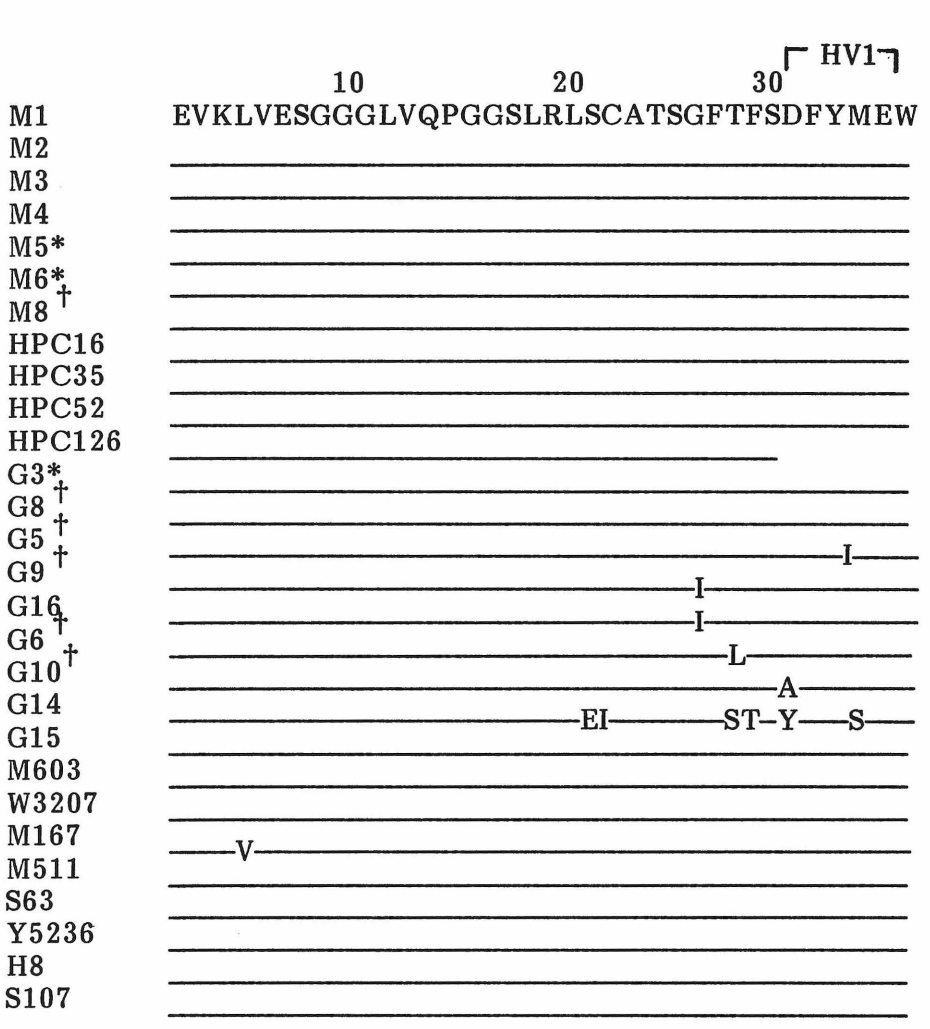
The V_H regions from phosphorylcholine-binding immunoglobulins fall into one V_H group

The N-terminal amino acid sequences of the V_H regions from eleven IgM hybridomas, nine IgG hybridomas and nine IgA myeloma proteins that bind phosphorylcholine are shown in Figure 1. Immunoglobulin V_H regions which differ by two or less amino acids over their N-terminal 27 residues are classified as members of a single group (1). As seen in Figure 1, all 29 V_H regions fall within a single V_H group. Twenty-two of the V_H sequences are identical, and six of the remaining seven V_H regions differ by a single residue from the prototype T15 sequence. The remaining V_H region G14, differs by six residues. The 12 differences from T15 in these V_H regions occur at 10 different positions with two identical substitutions seen at position 28 and two nonidentical substitutions at position 29. Eleven of 12 substitutions can be explained by single nucleotide substitutions, and eight of 12 substitutions lie outside the first hypervariable region as defined by Kabat et al. (30).

The V_L regions from phosphorylcholine-binding immunoglobulins fall into three V_L groups

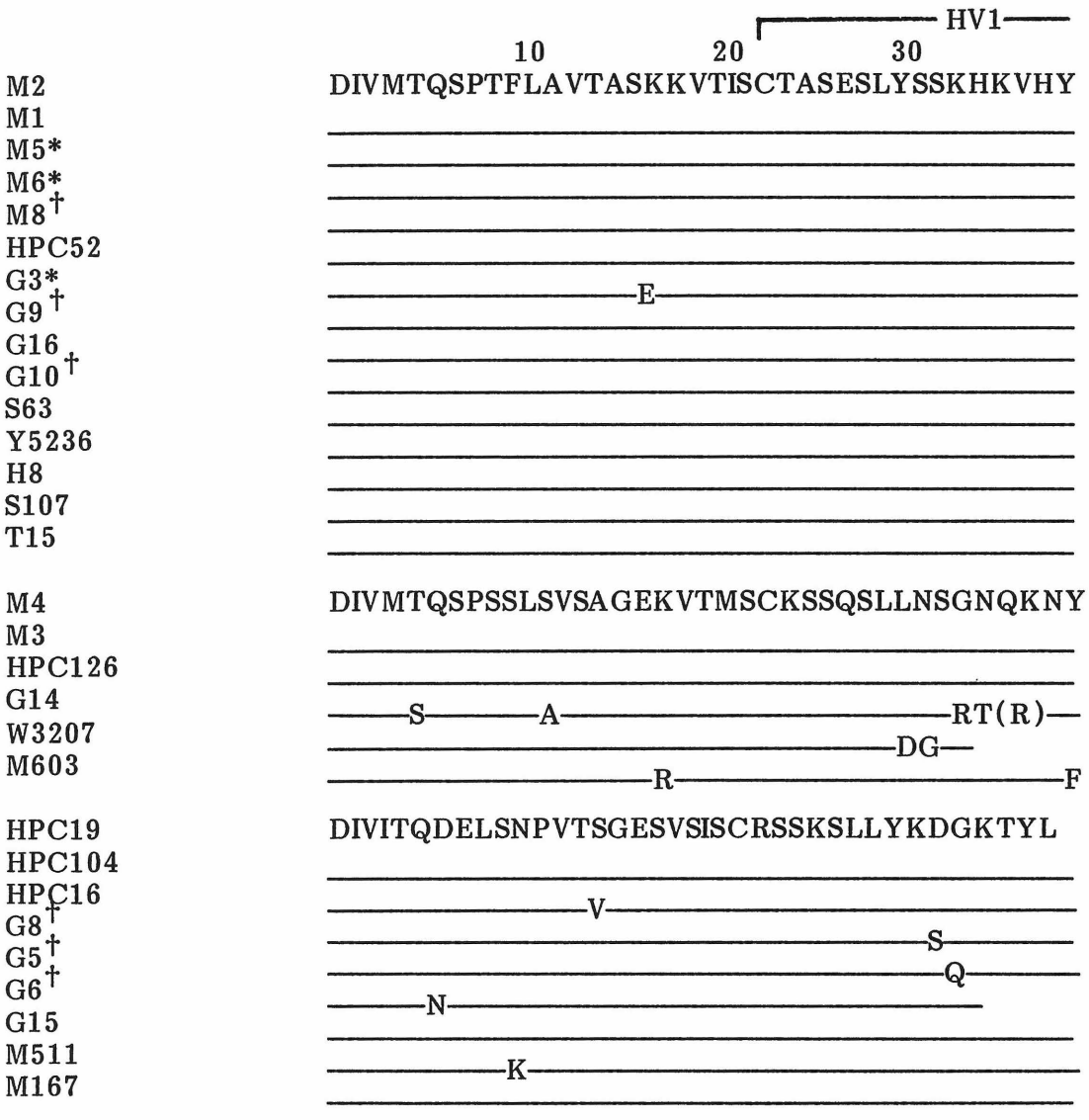
The N-terminal amino acid sequences of 30 V_L regions from immunoglobulins binding phosphorylcholine are given in Figure 2. The data indicate that the 30 V_L regions fall into three groups. These three groups are the same as those initially

Figure 1. V_H regions of phosphorylcholine-binding antibodies. Amino acids are noted by the one-letter code of Dayhoff (46). HV1 designates the first hypervariable region. All * proteins come from one mouse, while the † proteins come from a second mouse. Other proteins each come from a different individual. The sequences of HPC16, HPC35, HPC42, and HPC26 are from ref. 18, M603 is from ref. 45, M167 and W3207 are from ref. 2, and W3207, M511, S63, Y5236, H8, and S107 are from ref. 3.



	Source	Class	T15 Idiotype	V_L Group
M1	hybridoma	IgM	+	T15
M2	"	IgM	+	T15
M3	"	IgM	–	M603
M4	"	IgM	–	M603
M5*	"	IgM	+	T15
M6*†	"	IgM	–	T15
M8†	"	IgM	–	T15
HPC16	"	IgM	–	M167
HPC35	"	IgM	–	M21
HPC52	"	IgM	+	T15
HPC126	"	IgM	–	M603
G3*†	"	IgG3	+	T15
G8†	"	IgG3	–	T15
G5†	"	IgG3	–	M167
G9†	"	IgG3	+	M167
G16†	"	IgG3	+	T15
G6†	"	IgG1	–	M167
G10†	"	IgG1	–	T15
G14	"	IgG1	–	M603
G15	"	IgG1	–	M167
M603	myeloma	IgA	–	M603
W3207	"	IgA	–	M603
M167	"	IgA	–	M167
M511	"	IgA	–	M167
S63	"	IgA	+	T15
Y5236	"	IgA	+	T15
H8	"	IgA	+	T15
S107	"	IgA	+	T15

Figure 2. V_L regions of phosphorylcholine-binding antibodies. V_L sequences are categorized according to T15, M603 and M167 V_L groups with the most typical protein of each group used as the prototype. The sequence of HPC16 is written assuming Asx residues are identical to prototype asparagine or aspartic acid residues. Parentheses ( ) indicate probable assignments. HV1, *, †, as in Figure 1. The sequences of HPC19, HPC104, HPC52, HPC126, and HPC16 are from ref. 18, M167, S63, W3207, and M511 are from ref. 47, and T15, H8, S107, and M603 are from ref. 48.



	Class	Idiotype
M2	IgM	+
M1	IgM	+
M5*	IgM	+
M6*	IgM	–
M8†	IgM	–
HPC52	IgM	+
G3*	IgG3	+
G9†	IgG3	+
G16	IgG3	+
G10†	IgG1	–
S63	IgA	+
Y5236	IgA	+
H8	IgA	+
S107	IgA	+
T15	IgA	+
M4	IgM	–
M3	IgM	–
HPC126	IgM	–
G14	IgG1	–
W3207	IgA	–
M603	IgA	–
HPC19	IgM	–
HPC104	IgM	–
HPC16	IgM	–
G8†	IgG3	–
G5†	IgG3	–
G6†	IgG1	–
G15	IgG1	–
M511	IgA	–
M167	IgA	–

delineated by analysis of light chains from myeloma proteins binding phosphorylcholine, and are known as the T15, M603, and M167 groups (31). Light chains derived from antibodies to phosphorylcholine generated in mouse strains other than BALB/c also fall into these same three V_L groups (32). Fourteen of 15 V_L regions in the T15 group are identical, and the 15th differs by a single residue substitution which can be explained by a single nucleotide change. Three of six V_L regions in the M603 group are identical and the other three differ by two, two and five residues. All of these substitutions can be accounted for by single nucleotide changes. Four of nine V_L regions in the M167 group are identical and the others differ by a single residue. Three of these substitutions require two nucleotide changes (G5, G8 and HPC16). Seven of the 15 substitutions noted in these groups fall outside the first hypervariable region as defined by Kabat et al. (30). There appears to be more diversity in the M603 and M167 V_L groups than in the T15 group.

The T15 idiotype is present on immunoglobulins with sequence heterogeneity in V_H and V_L regions

Antisera recognizing the T15 idiotype have been used as a probe to identify a restricted group of antibodies that bind to phosphorylcholine. The immunoglobulin class and idiotype classification of antibodies are given in Figures 1 and 2 and Table I, and show that antibodies with the T15 idiotype are found among proteins of the IgM, IgG3, and IgA classes. The data in Figures 1 and 2 also show that antibodies with the T15 idiotype can have different amino acid sequences in their V regions. Four hybridoma antibodies (M1, M2, M5, and HPC52) and five myeloma proteins (S63, Y5236, H8, S107 and T15) exhibit the T15 idiotype and have identical N-terminal V_H and V_L sequences. However, hybridomas G3, G9, and G16 express the T15 idiotype but differ from the proteins described above by one residue either in the V_H (G9 and G16) or the V_L (G3) region. These data clearly indicate that antibodies with

TABLE I

Affinity of Antibodies for Phosphorylcholine

Antibodies	Class	T15 idiotype	$K_a \times 10^5\ M^{-1}$
T15-V_H, T15-V_L groups			
M2	IgM	+	3.5
M5	IgM	+	2.4
M8	IgM	-	3.2
G3	IgG3	+	1.3
G10	IgG1	-	4.4
T15	IgA	+	4.3
T15-V_H, M603-V_L groups			
M3	IgM	-	0.34
G14	IgG1	-	0.13
W3207	IgA	-	2.5
T15-V_H, M167-V_L groups			
G5	IgG3	-	0.25
G6	IgG1	-	1.8
G15	IgG1	-	1.0
M167	IgA	-	1.9

*Equilibrium dialysis experiments were performed at 25°C. Binding constants were accurate to $\pm$10%.

The T15 idiotype comprise a family of closely related but distinct V regions, and, in this regard, are similar to immunoglobulins which bind dextran (33).

Affinity for phosphorylcholine does not correlate with heavy chain class

The affinity constants for various phosphorylcholine-binding immunoglobulins are given in Table 1. The Scatchard plots were linear, and the number of binding sites ranged from 1.7 to 2.0 for monomeric forms. Values of α, the heterogeneity index, were calculated by the Sips distribution function and ranged from 0.9 to 1.06, indicating homogeneity of binding sites. The association constants cover a 30-fold range from 0.13 to 4.4 x 10^5 M^{-1}. There is no significant difference in the average affinity of IgM versus IgG antibodies.

Discussion

The hybridoma technique permits an examination of the normal antibody repertoire in the spleen

The myeloma and hybridoma immunoglobulins binding phosphorylcholine differ in one striking feature - their class distribution. Hybridoma antibodies are of the IgM and IgG heavy chain classes, whereas the myeloma proteins are of the IgA class (Fig. 1). This nonoverlapping distribution of classes is probably due to the different origins of hybridoma and myeloma cells. The hybridoma antibodies were obtained from immunized spleen cells, whereas the myeloma proteins were synthesized by B cells in the peritoneal cavity. It has been previously shown that B cells specific for phosphorylcholine from gut-associated lymphoid tissues primarily give rise to IgA-producing clones, whereas clones derived from splenic B cells produce mainly IgM, IgG, and IgA (34). The fact that the gut-associated lymphoid tissues receive different antigenic exposures than the spleen suggests that different repertoires of V regions will be selectively expressed in the corresponding populations of B cells.

Thus the hybridoma process may transform a different population of B cells than the myeloma process.

The hybridoma technique also allows one to analyze many B cells from an individual mouse, whereas each myeloma tumor comes from a different mouse. For example, four of the seven nonprototype V_H sequences and three of the nine nonprototype V_L regions come from hybridomas generated in a single mouse (Figures 1 and 2). This observation underscores the fact that extensive antibody heterogeneity is a property of individual mice and not just of the population of BALB/c mice.

Another important advantage of the hybridoma technique is the possibility for experimental manipulation of the animal's immune response prior to cell fusion. For example, the idiotypic characterization and N-terminal sequence analysis of five hybridoma IgM antibodies to phosphorylcholine from mice suppressed for the T15 idiotype have recently been examined (HPC104, HPC126, HPC19, HPC16, HPC35) (17, 18) in addition to the two reported here (M3, M4). Five V_H regions fall into the T15 group and the remaining two are derived from a new V_H group (HPC19 and HPC104, not shown). Three V_L regions are of the M167 group, three are of the M603 group, and one is derived from a different V_L group (HPC35, not shown). No light chains of the T15 V_L group were noted. The suppression of the T15 idiotype clearly selects against B cells expressing the T15 V_L group.

Does somatic variation occur in the V_H segments derived from immunoglobulins binding phosphorylcholine?

The V_H region is encoded by three distinct gene segments-V(variable), D (diversity), and J (joining) (33, 35). The V_H gene segment encodes approximately residues 1-99. The question arises as to whether V segment (that portion of the V region encoded by the V gene segment) diversity arises entirely from multiple germline genes, or whether some of the diversity must occur by somatic variation. A cloned V_H probe derived from a full length cDNA copy of the heavy chain mRNA derived from the

S107 myeloma tumor (36) has been used to analyze undifferentiated (embryo) DNA by Southern blot analysis (37). [The S107 and T15 V_H regions are identical in sequence (3).] Approximately eight to nine restriction fragments hybridize with this probe. Experiments have demonstrated that the S107 V_H probe cross-reacts with mRNAs from M167 and M603 (36). Because these are among the most different V_H regions from phosphorylcholine-binding immunoglobulins (3), the S107 cDNA probe should detect most, if not all, of the V_H genes in the T15 group. If each fragment hybridizing with the S107 cDNA probe contains one V_H gene segment belonging to the T15 V_H group, then there appear to be eight or nine V_H germline genes of the T15 V_H group. One can next ask, how many different V_H gene segments are present among the phosphorylcholine-specific myeloma and hybridoma proteins sequenced to date? Among the nine myeloma V_H regions that are completely sequenced, four differ in their V segments from the T15 V_H region by one to seven residues (3). Among the 20 hybridoma V_H regions that have been examined over their N-terminus, five differ from the T15 V_H region and from the myeloma V_H variants. Additional sequence data on the V_H region of G3 also demonstrate that it differs from the T15 V_H sequences (38). Thus, in the 29 V_H regions derived from immunoglobulins binding phosphorylcholine, 11 different V_H segments are present. This is more than the expected number of phosphorylcholine-specific V_H gene segments. Additional sequence analysis of the hybridomas now available, as well as the analysis of new examples, will almost certainly increase the number of V_H segment variants well beyond the number of germline V_H gene segments encoding the T15 V_H group. Accordingly, even if nine somewhat underestimates the number of T15 V_H-like gene segments, the observed phenotypic variation suggests that a portion of the V segment diversity arises by somatic variation. It is interesting to note, as judged by the data in Figure 1, that variation occurs as frequently outside as within the hypervariable region (8 or 12 V_H substitutions are outside the first

hypervariable region). Accordingly, either the framework variants arise from distinct germline V_H gene segments or the presumptive mechanism for somatic variation is not confined to the hypervariable segments.

An alternative explanation for the V region diversity that we have noted above is the possibility of genetic polymorphism in V gene segments among mice of the BALB/c strain. Although we cannot rule this possibility out, we feel it is unlikely for several reasons. First, BALB/c mice have been inbred for hundreds of generations and mice obtained from a single source should be greater than 99% homozygous (39). Second, restriction digests of the DNAs from individuals and pools of mice have never shown any polymorphisms in C genes from BALB/c mice. Finally, the polymorphism would presumably have to be far more extensive than that known for other well-studied systems (e.g., globins and haptoglobins) in order to explain the extensive V region diversity that we have noted here. Admittedly, comparisons between multigene families such as the immunoglobulin V genes and the genes coding for C regions, globins, and haptoglobins are not entirely valid. Nonetheless, we feel that the above observations are relevant. Therefore it appears likely that the V region variation arises from multiple germline V gene segments and possibly some mechanism for somatic variation.

IgG antibodies binding phosphorylcholine exhibit more diversity than their IgM counterparts

The V_H regions from 11 IgM hybridomas do not show any diversity in their N-terminal 37 residues (Fig. 1). In contrast, the corresponding portions of the V_H regions from IgG hybridomas vary in six of the nine examples studied. The same pattern is seen in the corresponding V_L regions (Fig. 2). Eleven of 12 IgM light chains are identical in sequence to the prototype sequence of their respective groups while five of nine IgG V_L regions vary. Thus eight out of nine of the IgG antibodies vary

from the V group prototype sequences in either their V_H or V_L regions, whereas only one of the 10 IgM antibodies varies. Only 10 IgM antibodies are used in this comparison because the heavy chains of HPC19 and HPC104 do not belong to the T15 V_H group, and the light chain of HPC35 does not belong to the T15, M603, or M167 V_L group (18).

One possible explanation for these differences between IgG and IgM proteins is that both arise from the same population and the differences are due to sampling errors. We used Fisher's exact method to demonstrate that it is very unlikely that these differences arise by sampling error. The probablility that V_L regions of the IgM molecules and the V_L regions of the IgG molecules were selected from the same population is less than 0.05; likewise, the probability that the V_H regions from the IgM and IgG molecules were drawn from the same pool is 0.01; finally, the probability that both the V_L and V_H regions of these molecules were drawn from common populations is less than 0.005. Thus the IgG polypeptides are significantly more diverse than their IgM counterparts.

Recombinant DNA studies have suggested that IgG- and IgA-producing cells are generally derived from B cells initially producing IgM (40-43). Thus the greater diversity of the IgG hybridomas must arise from infrequent variants, either in IgM- or IgG-producing B cells, which are selectively expanded and expressed in IgG-producing cells. (In this discussion the word variant refers to a nonprototype sequence of either somatic or germline origin.) Two important questions arise. First, how and when are the infrequent variants generated? Second, what selective pressures give rise to the differences between IgM and IgG diversity?

The infrequent variant clones which undergo expansion to create IgG diversity may arise either by somatic variation or may be encoded by germline V gene segments. Our data suggest that both of these sources may contribute such variant clones. For example, the M167 V_L group, while encoded in the germline separately from

M603 V_L and T15 V_L groups, is not seen in the six IgM light chain sequences from unsuppressed mice. In mice suppressed with anti-T15 sera, however, three out of seven IgM antibodies express the M167 V_L group (only one IgM expresses both the T15 V_H and the M167 V_L). Four out of nine IgG antibodies from unsuppressed mice also express the M167 V_L group. Thus it appears that the T15 V_L group is dominant in the IgM pool in normal mice but that suppression allows infrequent IgM M167 V_L group clones to be observed. These IgM M167 V_L group clones increase in frequency in the IgG pool. As discussed earlier, somatic diversification also may account for some of the observed sequence differences in the IgG pool.

The time at which variants are first expressed cannot be determined from these data. Some variants are probably encoded in the germline and arise early in differentiation. Somatic variants may be created as a result of normal somatic mutation or by special mutational mechanisms. Such somatic mutation processes could operate in IgM- or IgG-producing cells, in cells switching from IgM to IgG, or in cells triggered by antigen.

One may envision two types of selection operating on germline or somatic variants which could result in the observed IgM-IgG diversity difference. i) Antigen-driven selection Variant B-cell clones may be expanded as the result of increased affinity for antigen. As IgG-producing cell populations have undergone more antigen-driven selection than IgM populations, IgG antibodies should shown more variability than IgM antibodies (44). On the surface, the data present in Table I would appear to argue against this possibility. The range of affinity constants seen in the IgM and IgG antibodies are not significantly different. However, these data do not exclude the possibility that the in vitro affinity of soluble antibody for free hapten could be significantly different than the affinity of cell surface antibodies for the hapten-carrier conjugate in vivo. Selection for variants in nonbinding site regions may occur whenever framework residues modify the positions of binding-site residues. ii) Idiotype selection. Selection

of infrequent variants may be accomplished by idiotype-specific regulation. For example, if idiotype-specific suppression helps control the level of predominant clones, then mutation away from such suppressed idiotypes would be a major selective advantage. As IgG-producing cells have undergone more somatic generations than IgM-producing cells, selection will have had more opportunity to generate diversity in IgG molecules. Alternatively, IgM and IgG molecules may have unknown qualitative and quantitative differences in their idiotypic regulation.

IgA antibodies, like IgG antibodies, are produced by secondary cells and have presumably undergone many rounds of antigen and idiotype-specific selection. This similarity would suggest that the diversity patterns of IgA antibodies should resemble those of IgG antibodies. However, only four out of nine IgA proteins differ from the prototype sequences shown in Figure 1, and there is only one amino acid difference in nine IgA V_H N-termini. The IgA diversity appears intermediate between that of the IgG and IgM molecules.

An analysis of the complete V_H segments from five μ chains and four γ chains from hybridomas binding phosphorylcholine has been carried out (38). A comparison of IgM V_H segments with their nine IgA myeloma counterparts (2, 3, 45) suggests that the V_H regions from α chains are significantly more diverse than those from μ chains. Thus, the N-terminal 36 residues of the IgA V_H regions give a misleading picture of the diversity in the remainder of the V_H segment. The IgA as well as the IgG V_H regions exhibit significantly more diversity than their IgM counterparts.

Summary

An amino acid sequence analysis of the N-terminal V_H and V_L regions from 20 hybridoma and nine myeloma proteins directed against phosphorylcholine raises several interesting points. First, the V_H segments of these immunoglobulins demonstrate sufficient variability to raise the possibility of somatic diversification. Second,

the IgG antibodies exhibit significantly more variability than their IgM counterparts. These data suggest that the selection of infrequent germline or somatically-generated clones gives rise to greater heterogeneity in the IgG pool than in the IgM pool. Lastly, the T15 idiotype represents a population of closely related but distinct V_H and V_L regions.

We thank Nadine Nivera for excellent technical assistance, Dr. Fred Karush for help in performing equilibrium dialysis experiments, Julia Hurwitz and Joseph Davie for hybridomas, and Doug Fambrough and Mitch Kronenberg for helpful discussions, Tim Hunkapiller for computer graphics, and J. Myron Hood for help with the statistical analyses.

References

1. Potter, M. 1977. Antigen-binding myeloma proteins of mice. Adv. Immunol. 25: 141.
2. Rudikoff, S., and M. Potter. 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. Proc. Natl. Acad. Sci. U.S.A. 73: 2109.
3. Hubert, J., N. Johnson, B. Eaton, J. Greengard, and L. Hood. 1980. The complete V_H sequences of six myeloma proteins with specificity for phosphorylcholine. Manuscript in preparation.
4. Weigert, M., I. Cesari, S. Yonkovich, and M. Cohn. 1970. Variability in the lambda light chain sequences of mouse antibody. Nature (Lond.) 228: 1045.
5. Cohn, M., B. Blomberg, W. Geckeler, W. Rashke, R. Riblet, and M. Weigert. 1974. First order considerations in analyzing the generation of diversity. In The Immune System: Genes, Receptors, Signals. E. Sercarz et al., editors. Academic Press, New York, 89.
6. Weigert, M., and R. Riblet. 1976. Genetic control of antibody variable regions. Cold Spring Harbor Symp. Quant. Biol. 41: 837.
7. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. Nature (Lond.) 276: 785.
8. Sher, A., and M. Cohn. 1972. Inheritance of an idiotype associated with the immune response of inbred mice to phosphorylcholine. Eur. J. Immunol. 2: 319.
9. Cosenza, H., and G. Köhler. 1972. Specific inhibition of plaque formation to phosphorylcholine by antibody against antibody. Science. 176: 1027.

10. Clafin, J. L., R. Lieberman, and J. M. Davie. 1974. Clonal nature of the immune response to phosphorylcholine. I. Specificity, class, and idiotype of phosphorylcholine-binding receptors on lymphoid cells. J. Exp. Med. 139: 58.

11. Gearhart, P. J., N. H. Sigal, and N. R. Klinman. 1975. Heterogeneity of the BALB/c antiphosphorylcholine antibody response at the precursor cell level. J. Exp. Med. 141: 56.

12. Sigal, N. H., P. J. Gearhart, J. L. Press, and N. R. Klinman. 1976. Late acquisition of a germline antibody specificity. Nature (Lond.) 259: 51.

13. 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.

14. Gearhart, P. J., and J. J. Cebra. 1978. Idiotype sharing by murine strains differing in immunoglobulin allotype. Nature (Lond.) 272: 264.

15. Strayer, D. S., H. Cosenza, W. M. F. Lee, D. A. Rowley, and G. Köhler. 1974. Neonatal tolerance induced by antibody against antigen-specific receptor. Science. 186: 640.

16. Cosenza, H., A. A. Augustin, and M. H. Julius. 1976. Idiotypes and anti-idiotypes as probes in analysis of immunoregulation. Cold Spring Harbor Symp. Quant. Biol. 41: 709.

17. Berek, C., M. H. Shreier, C. L. Sidman, J-C. Jaton, H. P. Kocker, and H. Cosenza. 1980. Phosphorylcholine-binding hybridoma proteins of normal and idiotypically suppressed BALB/c mice. I. Characterization and idiotypic analysis. Eur. J. Immunc 10: 258.

18. Kocker, H. P., C. Berek, M. H. Schreier, H. Cosenza, and J-C. Jaton. 1980. Phosphorylcholine-binding hybridoma proteins of normal and idiotypically suppressed BALB/c mice. II. Variable region N-terminal amino acid sequences. Eur. J. Immunol. 10: 264.

19. Hood, L., P. Barstad, E. Loh, and C. Nottenburg. 1974. Antibody diversity: An assessment. In The Immune System: Genes, Receptors, Signals. E. Sercarz et al., editors. Academic Press, New York. 119.

20. Köhler, G., and C. Milstein. 1975. Continuous culture of fused cells secreting antibody of predefined specificity. Nature (Lond.) 256: 495.

21. Shulman, M., C. D. Wilde, and G. Köhler. 1978. A better cell line for making hybridomas secreting specific antibodies. Nature (Lond.) 276: 269.

22. Kennett, R. H., K. A. Denis, A. S. Tung, and N. R. Klinman. 1978. Hybrid spleen fragments, neonatal spleen cells, and human spleen cells. Curr. Top. Microbiol. Immunol. 81: 77.

23. Hurwitz, J., C. Coleclough, and J. Cebra. 1980. C_H gene rearrangements in IgM-bearing B cells and in the normal splenic DNA component of hybridomas making different isotypes of antibody. Cell. In press.

24. Klinman, N. R., A. R. Pickard, N. H. Sigal, P. J. Gearhart, E. S. Metcalf, and S. K. Pierce. 1976. Assessing B cell diversification by antigen receptor and precursor cell analysis. Ann. Immunol. (Inst. Pasteur). 127C: 489.

25. Chesebro, B., and H. Metzger. 1972. Affinity labeling of a phosphorylcholine-binding mouse myeloma protein. Biochemistry. 11: 766.

26. Scatchard, G. 1949. The attractions of proteins for small molecules and ions. Ann. N.Y. Acad. Sci. 51: 660.

27. Karush, F. 1962. Immunologic specificity and molecular structure. Adv. Immunol. 2: 1.

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

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.

30. Kabat, E. A., T. T. Wu, and H. Biolofsky. 1979. Sequences of Immunoglobulin Chains. Public Health Services, National Institutes of Health.

31. Barstad, P., S. Rudikoff, M. Potter, M. Cohn, W. Konigsberg, and L. Hood. 1974. Immunoglobulin structure: Amino terminal sequences of mouse myeloma proteins that bind phosphorylcholine. Science. 183: 962.

32. Claflin, J. L. 1976. Uniformity in the clonal repertoire for the immune response to phosphorylcholine in mice. Eur. J. Immunol. 6: 669.

33. 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.

34. Gearhart, P. J., and J. J. Cebra. 1979. Differentiated B lymphocytes: Potential to express particular antibody variable and constant regions depends on site of lymphoid tissue and antigen load. J. Exp. Med. 149: 216.

35. 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.

36. Early, P. W., M. M. Davis, D. B. Kaback, N. Davidson, and L. Hood. 1979. Immunoglobulin heavy chain gene organization in mice: Analysis of a myeloma genomic clone containing variable and α constant regions. Proc. Natl. Acad. Sci. U.S.A. 76: 857.

37. Davis, M. W., P. W. Early, K. Calame, D. L. Livant, and L. Hood. 1979. The organization and rearrangement of heavy chain immunoglobulin genes in mice. In Eucaryotic Gene Regulation. R. Axel, T. Maniatis, and C. Fox, editors. Academic, Press, Inc., New York. 393.

38. Johnson, N., P. Gearhart, R. Douglas, L. Hood. 1980. The complete variable region sequences of heavy chains from nine hybridoma proteins binding phosphorylcholine. In preparation.

39. Klein, J. 1975. Biology of the Mouse Histocompatibility-2 Complex. Springer-Verlag, New York.

40. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature (Lond.) 286: 676.

41. Takahashi, N., T. Kataoka, and T. Honjo. 1980. Nucleotide sequences around class switch recombination site of the immunoglobulin γ2b chain gene of mouse. Gene. In press.

42. Honjo, T., T. Kataoka, Y. Yaoita, A. Shimizu, N. Takahashi, Y. Yamawaki-Kataoka, T. Nikaido, S. Nakai, M. Obata, T. Kawakami, and Y. Nishida. 1980. Organization and reorganization of immunoglobulin heavy chain genes. Cold Spring Harbor Symp. Quant. Biol. In press.

43. Davis, M., S. Kim, and L. Hood. 1980. DNA sequences mediating class switching in alpha immunoglobulin genes. Science. In press.

44. Rodwell, J. D., and F. Karush. 1980. Restriction in IgM expression. I. The V_H regions of equine anti-lactose antibodies. Molec. Immunol. In press.

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

46. Dayhoff, M. O. 1976. Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Silver Spring, Maryland. 5: 189.

47. Barstad, P., J. Hubert, M. Hunkapiller, A. Goetze, J. Schilling, B. Black, B. Eaton, J. Richards, M. Weigert, and L. Hood. 1978. Immunoglobulins with hapten-binding activity: Structure-function correlations and genetic implications. Eur. J. Immunol. 8: 497.

48. Hood, L., E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg, 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.

Chapter 3

The amino acid sequence of nine V_H regions from hybridomas binding phosphorylcholine

This manuscript will be submitted to the Journal of Experimental Medicine

THE AMINO ACID SEQUENCE OF NINE V_H REGIONS FROM HYBRIDOMAS BINDING PHOSPHORYLCHOLINE

By N. JOHNSON, P. GEARHART*, R. DOUGLAS, and L. HOOD

(From the Division of Biology, California Institute of Technology Pasadena, California 91125)

Running title: 9 V_H Sequences

*Department of Embryology, Carnegie Institution of Washington, 115 W. University Parkway, Baltimore, Maryland 21210.

An understanding of the mechanisms responsible for antibody diversity has been a goal that has intrigued biologists for the past 80 years. In the last few years, as a result of recombinant DNA techniques and protein sequence analyses, immunologists have been able to directly compare sequence information obtained from antibody genes and proteins and thereby have gained important insights into at least some of the mechanisms responsible for antibody diversity (1-4).

Antibody molecules are composed of heavy (H) and light (L) chains which in turn are both divided into variable (V) and constant (C) regions. Comparative analyses of many variable regions demonstrate two important patterns. First, amino acid sequence variability is unevenly distributed throughout the V region. The antigen-binding site is lined with extremely variable sections termed hypervariable regions, while nonbinding site, or framework, regions are more highly conserved (5-8). The existence of hypervariable regions has suggested that somatic mechanisms may play a role in diversity generation (9, 10). Second, V regions fall into sets or groups of closely related sequences (formerly termed subgroups) (11, 12). The minor variation within the V regions of a particular group has afforded special insights into mechanisms of diversity (4, 12, 13).

The antibody molecules are encoded by three unlinked families of genes--two light chain families, lambda (λ) and kappa (κ), and a heavy chain family. The V_L regions are encoded by two distinct gene segments, V_L and J_L (joining), whereas the V_H regions appear to be encoded by three distinct gene segments, V_H, D (diversity), and J_H. The J_L gene segments encode a small portion of the L chain third hypervariable region, whereas the D and J_H gene segments encode most of the H chain third hypervariable region. The C regions for both light and heavy chains are encoded by C genes separate from the V genes (1, 14).

Analyses at the protein and in certain cases the DNA level for several sets of closely related antibody V regions have provided important insights about mechanisms

for antibody diversity. i) There appear to be several hundred germline V_κ and V_H gene segments. Thus germline diversity in these gene families is extensive (15). ii) There are four functional J_H and J_κ gene segments (1, 16, 17). It appears that any J_H gene segment may be joined with any V_H gene segment and likewise any J_κ gene segment with any V_κ gene segment. This combinatorial joining is an important mechanism for amplifying diversity (18). iii) The site-specific recombination mechanism which joins the V_L and J_L or V_H, D, and J_H gene segments may unite particular gene segments at different nucleotides and, accordingly, generate somatic variation at these junctions during the act of joining (16-18). iv) V regions appear to exhibit somatic variation beyond that which can be explained by the germline V gene repertoire and gene segment junctional diversity. In the V_λ system this diversity appears exclusively in the hypervariable regions (10), but in several others it may be present in framework regions as well (19, 20).

The first paper in this series described the N-terminal V_H and V_L sequences for 16 hybridoma antibodies binding the simple hapten phosphorylcholine (21). Two interesting observations were made. First, somatic variation appears to be necessary to account for the diversity noted in the N-termini of the V_H regions in these immunoglobulins. Second, the V_H regions from γ chains are significantly more diverse than their μ counterparts. We report here the complete V_H region sequences from seven hybridomas binding phosphorylcholine and the complete V_H segment sequences of two additional hybridomas. These studies delineate in more detail the nature of diversity in the V_H, D, and J_H gene segments of phosphorylcholine antibodies and support the earlier observation that the V_H regions from γ chains are more diverse than their μ counterparts.

Materials and Methods

Isolation of Hybridoma Proteins. Hybridoma proteins were isolated as previously

described (21).

Chain Separation. Hybridoma proteins were mildly reduced in 0.4 M Tris, pH 8.2, 0.15 M NaCl, 2 mM EDTA, 30 mM DTT at room temperature for 1.5 h and alkylated at 4°C for 2 h in 65 mM recrystallized iodoacetamide. The reduced and alkylated proteins were precipitated by the dropwise addition of an equal volume of saturated $(NH_4)SO_4$ at 4°C for at least 2 h and the precipitate collected by centrifugation. The precipitate was then dissolved in 8 M guanidine HCl and diluted to either 3 M or 5 M guanidine for chain separation. IgM and IgG1 hybridomas were fractionated on LKB Ultrogel AcA34 in 3 M guanidine HCl, 0.2 M NH_5CO_3 while the IgG3 proteins, because of their tendency to aggregate, were fractionated on Pharmacia Sephacryl S-200 equilibrated in 5 M guanidine, 0.2 M NH_5CO_3. Heavy chain peaks were desalted on Sephadex G-25 in 0.2 M NH_3.

Cyanogen Bromide Digests and Treatment of Cyanogen Bromide Fragments. Desalted and lyophilized heavy chains were digested with 2.5% cyanogen bromide in 70% HCO_2H for 24 h at 4°C. The digest was then diluted 12-fold, lyophilized twice, and applied to an LKB Ultrogel AcA54 column equilibrated in 3 M guanidine 0.2 M NH_4CO_3. After chromatography in guanidine, cyanogen bromide fragments were desalted on G-25 either in 0.2 M NH_3 or 0.5 M NH_3. All fragments from V region position 35 to 82 were desalted in 0.5 M NH_3.

Precipitation of the 35 to 82 fragment was adapted from Rudikoff and Potter (22). Briefly, the fragment was dissolved in 10% acetic acid and the pH was raised by the addition of concentrated NH_4OH until a precipitate formed. This precipitate was spun out in the cold and contained only the 35-82 fragment.

Other Chemical Modifications. Complete reduction and alkylation was as above but performed in the presence of 8 M guanidine.

Succinylation was performed in 0.2 M Tris, 8 M guanidine, pH 9.5 using a 20:1

weight ratio of succinic anhydride to protein. The pH was maintained with 2 N NaOH and an automatic titrator. Succinylated fragments were desalted on G-25 in 0.2 M NH_3, lyophilized, and cleaved at tryptophane residues in 35% cyanogen bromide, 1:1 heptafluorobutyric acid:88% formic acid for 18 h at room temperature (23). This cleavage mixture was dried, redissolved in trifluoroacetic acid, and applied to the sequenator.

Amino Acid Sequence Determination. Automated sequence analysis was done on an extensively modified Beckman 890B protein sequenator (24) and the phenylthiohydantoin amino acids determined using high performance liquid chromatography (25).

Trypsin Digests and 2-Dimensional Paper Maps. Trypsin digests were done at a 1:100 w/w enzyme:substrate ratio. The digests were subjected to chromatography and electrophoresis on paper (26) and then lightly sprayed with 0.1% ninhydrin in ethanol. Colors were developed for 6 min at 100°C and the spots cut out and extracted with 1 M NH_3 (acidic peptides) or 5% HCO_2H (basic peptides).

Results

Isolation of Cyanogen Bromide Fragments from the V_H Region. Figs. 1 and 2 show the cyanogen bromide gel filtration patterns of μ, γ1, and γ3 heavy chains. In each case the V region contains two methionines, Met 34 and Met 82. The cyanogen bromide fragment extending from positions 35 to Met 82 is labeled with the numeral I followed by the heavy chain class (Iμ, Iγ1, Iγ3), and the fragment from position 83 into the C region is labeled with the numeral II followed by the heavy chain class (IIμ, IIγ1, IIγ3).

The procedures for the isolation of the I peptides are shown in Fig. 1. The peptides Iγ3 and Iμ can be isolated cleanly by a single dimension of gel filtration as shown in Figs. 1A and 1B. The Iγ1 peptide can be isolated either by precipitation

Figure 1. CNBr profiles of hybridoma heavy chains. Heavy chains were isolated, CNBr digested, and chromatographed on AcA54 in 3 M guanidine, 0.2 M NH_5CO_3. a) IgG3 (G3 heavy chain). b) IgM (M3 heavy chain). c) IgG1 (G6 heavy chain).

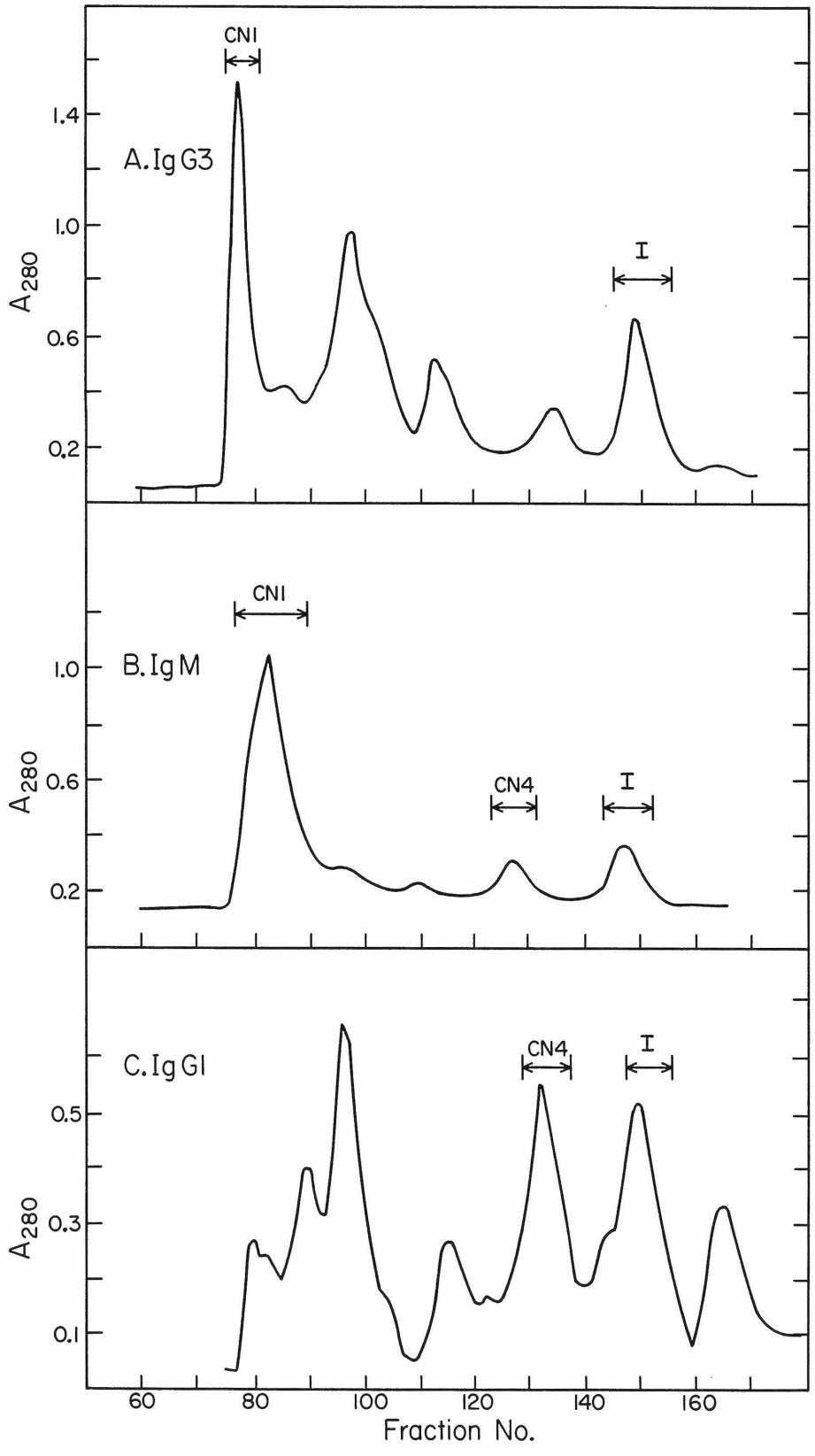
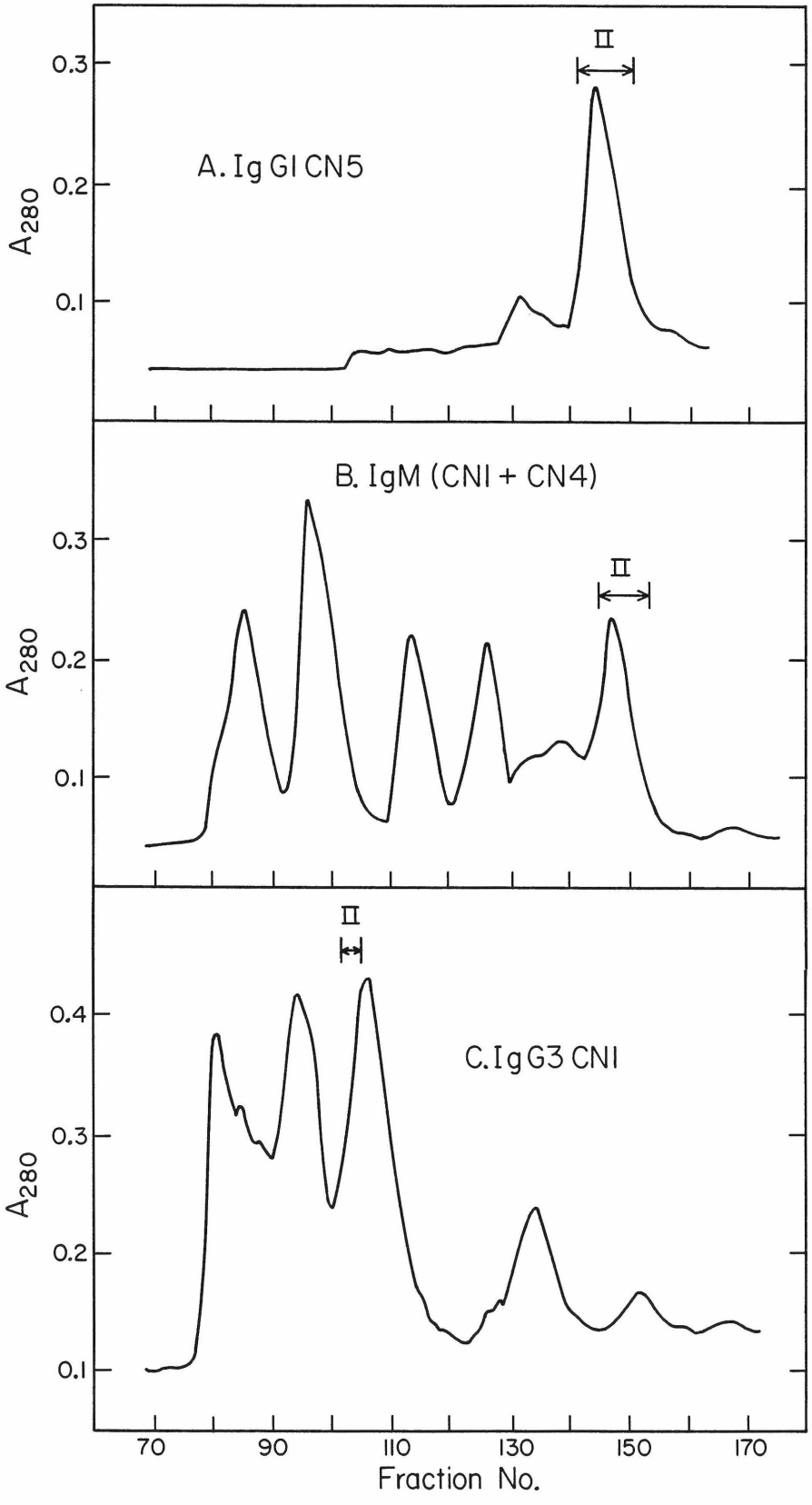

CN1
A. Ig G3
I
A_{280}
1.4
1.0
0.6
0.2
CN1
B. Ig M
CN4
I
A_{280}
1.0
0.6
0.2
CN4
I
C. Ig G1
A_{280}
0.5
0.3
0.1
60
80
100
120
140
160
Fraction No.

Figure 2. Profiles of fully reduced and alkylated heavy chain CNBr fragments. Fragments were obtained as in Figure 1 and were fully reduced and alkylated and rechromatographed on AcA54 (in 3 M guanidine, 0.2 M NH_5CO_3). a) IgG1 CN4, from Figure 1c. b) IgM (CN1 + CN4), from Figure 1b. c) IgG3 CN1, from Figure 1a.


II
A. Ig GI CN5
B. IgM (CNI + CN4)
II
C. Ig G3 CNI
II
A280
0.3
0.2
0.1
0.4
70
90
110
130
150
170
Fraction No.

away from contaminating peptides as described in Materials and Methods or by using only the low molecular weight component of the Iγ1 peak as shown in Fig. 1C.

The II peptides are contained as mixtures in CN1γ3 (Fig. 1A), CN4μ (Fig. 1B), and CN4γ1 (Fig. 1C). Each of these peptide mixtures was fully reduced and alkylated and was refractionated on AcA54. In the case of the μ chains, the CN1 peptide also was included to avoid losses due to disulfide interchange. The profiles of these chromatographic runs are shown in Fig. 2. The IIγ1 and IIμ peptides can be isolated by pooling the designated peaks in Figs. 2A and 2B, while IIγ3 can be isolated cleanly by pooling only the first 40% of the appropriate peak in Fig. 2C.

Heavy Chain Sequence Determination. A summary of the strategies for amino acid sequence analysis are shown schematically in Fig. 3. The N-terminal 38 residues of each heavy chain was determined by a single run on an intact chain.

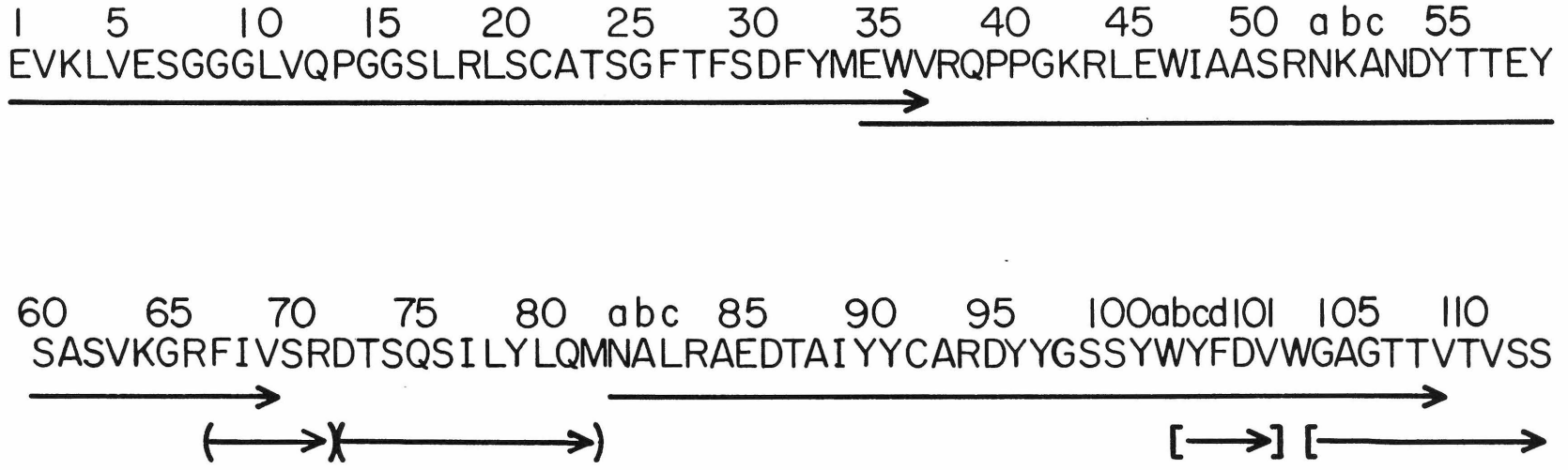
The I peptides were all sequenced for at least 36 cycles. In some cases the entire peptide was sequenced including the C-terminal homoserine, whereas in other cases residues 67 through 82 were determined by the complete sequence analysis of two smaller tryptic peptides isolated from paper (i.e., peptides including positions 67 to 71 and 72 to 82).

The II peptides were generally sequenced for 40 cycles to determine the remainder of the V region sequence. However, in some cases the IIμ peptides washed out of the sequenator cup before the completion of the run. When this occurred, intact II peptides were succinylated, cleaved at the tryptophane residue, and sequenced. The tryptophane cleavage of succinylated IIμ peptides yields two peptides with unblocked N termini: one five residue peptide of known sequence (100C to 103) and one longer peptide of partially unknown sequence (104 to the C terminus of IIμ). A sequenator run on this mixture permitted the unambiguous determination of the remainder of the V region sequence.

Figure 3. V_H region sequencing strategy. All of the hybridoma V_H regions were sequenced in the same general way as outlined for the M3 protein in this figure. In some cases it was necessary to generate peptides from tryptic and tryptophane cleavages. These peptides are indicated in the figure.

——>	amino acids determined by automatic sequenator
( )	tryptic peptide
[ ]	peptide sequenced after succinylation and tryptophane cleavage

Amino acid residues are numbered as in Ref. 30.

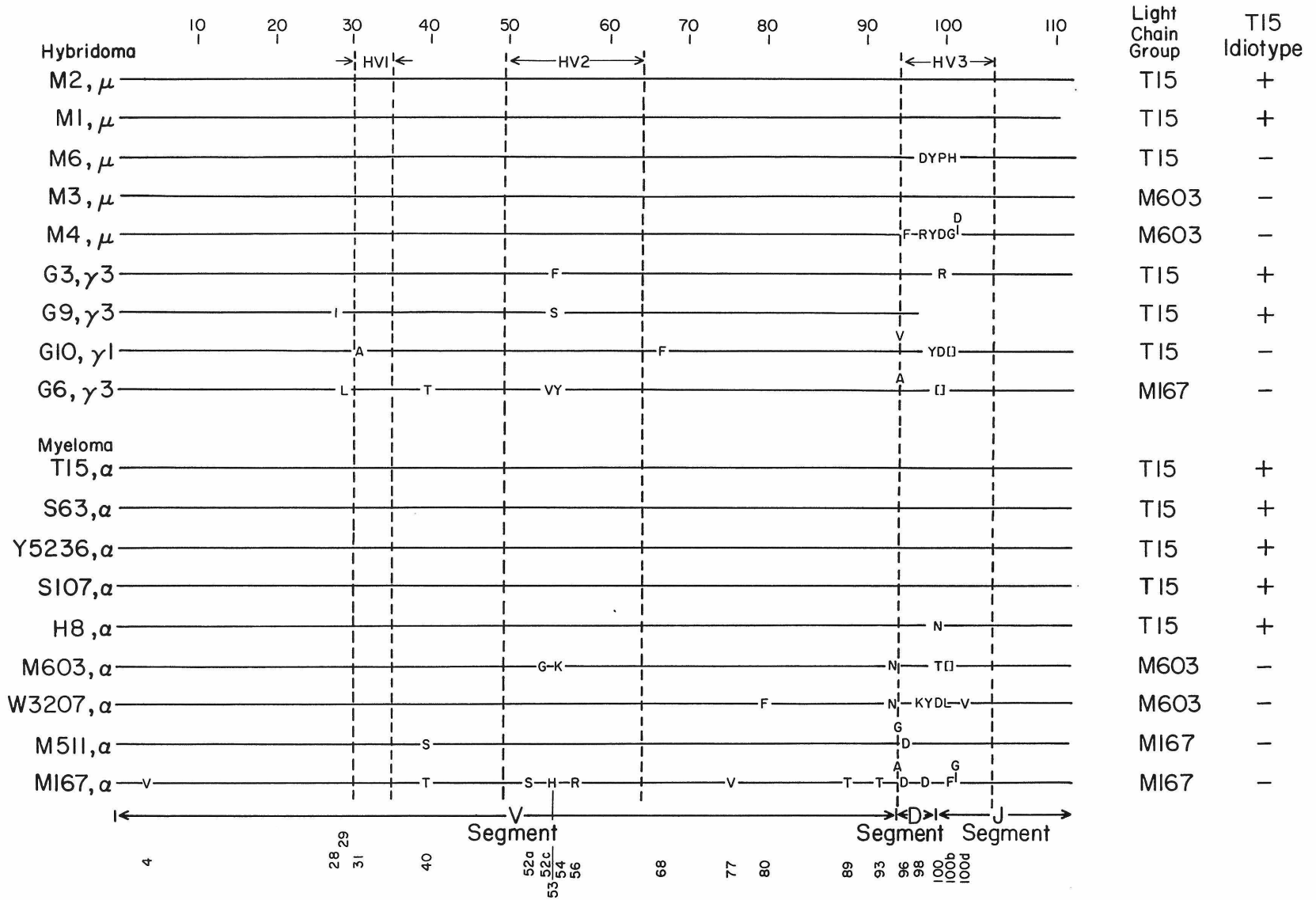

1 5 10 15 20 25 30 35 40 45 50 a b c 55
EVKLVESGGGLVQPGGSLRLSCATSGFTFSDFYMEWVRQPPGKRLEWIAASRNKANDYTTEY
60 65 70 75 80 a b c 85 90 95 100a b c d 101 105 110
SASVKGRFIVSRDTSQSILYLQMNALRAEDTAIYYCARDYYGSSYWYFDVWGAGTTVTVSS

Discussion

The Hybridoma and Myeloma V_H Regions Show Similar Patterns of Diversity. The V_H regions from nine hybridoma antibodies to phosphorylcholine are compared to their nine myeloma counterparts in Fig. 4. M4 is sequenced to the end of the D segment and G9 is sequenced up to the end of the V segment. Three hybridoma and four myeloma V_H regions are identical to the T15 prototype V_H sequence. The variant hybridoma V_H sequences differ by one to six residues from T15 and the variant myeloma V_H sequences differ by one to 11 residues. In both the hybridoma and myeloma V_H regions, insertions and deletions occur only in the third hypervariable region. In the hybridoma V_H regions 16 of 21 (76%) variant residues are found in hypervariable regions, and similarly 22 of 29 (76%) variant residues in the myeloma V_H regions are located in the hypervariable regions. Both hybridoma and myeloma V_H regions are expressed with three V groups of light chains--the T15 V_L group, the M603 V_L group, and the M167 V_L group. Since the basic patterns in the hybridoma and myeloma V_H regions appear similar, we will consider both sets of data together in analyzing the diversity patterns in the V segments (residues 1-95), the D segments (residues 96-100a) and in the J segments (100b-113) which are encoded, respectively, by the V_H, D, and J_H gene segments.

The V_H Segments Demonstrate Diversity in the Framework as well as Hypervariable Regions. Nine V_H segments are identical and 9 others differ by one to eight substitutions. Generally the variant V segments differ by two or more widely scattered substitutions. Twenty-two variant residues are present at 15 different positions. Single base changes can account for 20 of 22 amino acid substitutions, and two substitutions require two base changes. Four positions show variation in more than one V_H region: position 53 has four amino acid substitutions, position 52c has two substitutions, position 28 has an identical substitution in two V_H regions, and two different

Figure 4. The V_H regions from phosphorylcholine binding proteins. Hypervariable regions are shown at the top of the figure (HV1, HV2, and HV3) and V, D, and J segment boundaries are shown at the bottom. A solid line indicates identity to the T15 prototype, insertions and deletions are shown by residues above the solid line and by [ ], and amino acid substitutions are labeled using the one-letter code by Dayhoff (31). Positions of amino acid substitutions are labeled at the bottom of the figure except in HV3 where every other residue is labeled. V_L group, heavy chain class, and presence (+) or absence (-) of the T15 idiotype are shown for each sequence. The T15 and M167 sequences are from Ref. 3, the M603 sequence is from Ref. 22, and the S63, Y5236, S107, H8, W3207, and M511 sequences are from Ref. 32.


10 20 30 40 50 60 70 80 90 100 110
HV1
HV2
HV3
Hybridoma
M2, μ
M1, μ
M6, μ
DYPH
M3, μ
M4, μ
F-RYDG
G3, γ3
F
R
G9, γ3
I
S
G10, γ1
A
F
YD[]
G6, γ3
L
T
VY
[]
Myeloma
T15, α
S63, α
Y5236, α
S107, α
H8, α
N
M603, α
G-K
T[]
W3207, α
F
KYDL-V
M511, α
S
M167, α
V
T
S-H-R
V
T-T-D-D-F
Light Chain Group
T15
T15
T15
M603
M603
T15
T15
T15
M167
T15
T15
T15
T15
T15
M603
M603
M167
M167
T15 Idiotype
+
+
−
−
−
+
+
−
−
+
+
+
+
+
−
−
−
−
V Segment
D Segment
J Segment
4
28 29 31
40
52a 52c 53 54 56
68
77
80
89
93
96 98
100 100b 100d

amino acids occur in the three V_H regions with variability at position 40. It is important to stress that no two of the variant V segments are identical.

As was discussed in the preceding paper, the V_H regions from immunoglobulins binding phosphorylcholine appear to be encoded by perhaps 8-9 V_H gene segments. Accordingly, the V segment diversity noted in these heavy chains might arise from one of several sources: germline V_H gene segments (15), ordinary somatic mutation (10), a special hypermutation mechanism (9), or by somatic recombination, presumably among related V_H gene segments (15).

D Segment Diversity is Extensive. If we count each insertion or deletion as well as amino acid substitutions, the D segment has 23 variant positions over 102 residues analyzed. This D region variation tends to be clustered, since the ten variant proteins each average over two D segment substitutions. Fig. 4 demonstrates that the D segment diversity is of three types: i) isolated single base changes (e.g., position 99 in M167 and 100 in H8 and G3); ii) insertions or deletions at either end of the D segment (e.g., M4, G6, G10, M603, M511, and M167); and iii) blocks of amino acid substitutions (e.g., W3207, M4, M6, and G10). The single base substitutions could arise from the presence of multiple germline D gene segments or from somatic variation. The insertions or deletions of codons associated with either end of the D segment probably arise from the site-specific recombinational mechanism that joins the V_H and D or the D and J_H gene segments. This diversity arises because the actual DNA splice points may occur at different sites on each of these three gene segments, as has been well documented (16-19). The blocks of substitutions may arise from multiple germline D segments, from the splicing together of two germline D gene segments, or from a variety of other possibly somatic diversification mechanisms.

D gene segments have not yet been isolated from germline DNA. Accordingly, we can only guess as to the number and diversity of germline D gene segments.

As a result of D segment variability, the heavy chain third hypervariable region is the most diverse V region segment. This can be clearly seen in the V_H regions of immunoglobulins binding phosphorylcholine as well as those binding dextran (4).

The J_H Segments from Immunoglobulins Binding Phosphorylcholine are Encoded by One Germline J_H Gene Segment. Mouse heavy chains are encoded by four different J_H gene segments. In the 16 V_H regions from hybridoma antibodies binding phosphorylcholine that can be compared, all appear to have J segments derived from the J_{H1} gene segment. Five J segments have N-terminal substitutions and one J_H segment has an additional single base substitution. The N-terminal residue from the J_H segments of M6, G10, M603, and W3207 is different than that of T15. The N-terminal two residues of the J_H segment from M167 also differ from their T15 counterpart. Three of these five J_H segment changes are compatible with two base substitutions. Once, again, we feel this diversity can be explained by variability in the DNA splice point for joining the D and J_H gene segments (16-18). The J_H segment conservation seen in the phosphorylcholine-binding heavy chains provides a counterexample to the J_H segment diversity seen in antibodies which bind dextran. In the 19 V_H regions derived from hybridoma antibodies binding dextran, all four J_H gene segments are expressed (4).

The V_H Segments from IgG Hybridoma and IgA Myeloma Immunoglobulins Binding Phosphorylcholine Show More Diversity than Their IgM Counterparts. An analysis of the N-terminal portions of these V_H regions discussed in the first paper in this series suggested that IgM V_H regions were considerably less diverse than their IgG counterparts (21). For example, ten V_H and V_L N-termini from IgM hybridoma immunoglobulins showed only one one-base substitution when compared to prototype sequences of the corresponding V_H groups. On the other hand, eight of nine IgG antibodies had from one to 11 differences. Using Fisher's exact method, the diversity

differences between IgM and IgG molecules are highly significant ($p<.005$). The extensive sequence data presented in this paper allow a more detailed comparison to be made between IgG, IgA, and IgM V_H segments. Five of five V_H segments from μ chains are identical to the T15 prototype sequences (Fig. 4). Four of four V_H segments from γ chains differ from the T15 prototype. Four of nine V_H segments from α chains differ from the T15 prototype. These data reinforce the early supposition that the γ V_H segments are more diverse than their μ counterparts. Moreover, the α chains also appear to be significantly more diverse than their μ counterparts ($p<0.05$), although they may be somewhat less diverse than their γ counterparts. Possible explanations for these diversity differences have been discussed in the first paper of this series (21).

Selection Imposes Several Constraints on the Immunoglobulins Binding Phosphorylcholine. The contact residues for phosphorylcholine binding are highly conserved. Padlan et al. have proposed that the heavy chain tyrosine 33 and arginine 52 are the principal hapten-contact residues for heavy chains. In addition, there are probably charge interactions with lysine 52b and glutamic acid 35 and van der Waals interactions with residues 98 and 99 (27). While residues 98 and 99 vary somewhat in the sample of proteins analyzed here, the V_H regions in the T15 V_H group from proteins which bind phosphorylcholine show considerable conservation of the other contact residues. In all of these V_H regions sequenced to date, there are no substitutions at tyrosine 33, arginine 52, and lysine 52b and only one glutamic acid-serine interchange at position 35. (The glutamic acid-serine interchange is from Ref. 21.) This conservation is consistent with the importance of these residues in the binding of phosphorylcholine.

IgG and IgA antibodies show special diversity at residues 40 and 53. We have previously postulated that the diversity differences between IgM antibodies and IgG and IgA antibodies are the result of selection. Hence it is important to try to answer the question, is there any pattern to the amino acid variability in IgG

and IgM V_H segments? We feel that Figure 4 includes at least two such patterns of sequence variability, even though the limited number of non-T15 sequences makes these patterns difficult to detect. D segment variability is excluded from this discussion because of its possible special diversification mechanisms.

Ten out of 23 V_H segment substitutions occur at five positions in the second hypervariable region. Four of these substitutions occur at residue 53. We feel that this increased sequence variability at and around residue 53 is significant. It may result from some form of selection or from an unknown site-specific hypermutational mechanism.

A second position with an unusual pattern of sequence variability is position 40. M511, M167, and G6 have a hydroxyamino acid at position 40 while all other V_H regions have proline. Position 40 is five residues outside of the first hypervariable region and is part of a loop that interacts strongly with the light chain (27). M511, M167, and G6 all have M167 L group light chains, while all but one of the other V_H regions associate with T15 or M603 L group light chains. Thus the presence of a hydroxyamino acid at position 40 directly correlates with L chain group. The only exception to this is an IgM protein sequenced elsewhere which has a proline at residue 40 and a M167 L group light chain. Thus the sequence variation at residue 40 appears to be present in IgG and IgA proteins though possibly absent or at low levels in IgM proteins. This difference in the variability of residue 40 between IgM antibodies and IgG and IgA antibodies is consistent with our previous observations about the lack of variability in IgM V segments (21). It suggests that some of the diversity present in IgG and IgA populations may result from selective pressures on amino acids involved in heavy-light chain interactions.

Antibodies Binding Phosphorylcholine Demonstrate Combinatorial Association Between Light and Heavy Chains. Among the IgM immunoglobulins binding

phosphorylcholine, some employ the same heavy chain in conjunction with two very different light chains (e.g., M2 and M3 in Fig. 4). The T15- and M603-like V_L regions differ by more than 50% of their amino acid sequence over their N-terminal 38 residues (28). This is the first formal demonstration of the combinatorial association of light and heavy chains. The assumption has long been held that one heavy chain may pair with all the different light chains in a combinatorial manner (27). Indeed, if 1000 light and 1000 heavy chains can be associated combinatorially, 10^6 different antibodies could be produced. Thus these data support the hypothesis that combinatorial association is a fundamental mechanism for information amplification in the immune system. It is interesting that this first example of combinatorial association generates antigen-binding sites which interact with the same hapten.

The T15 Idiotype is Found on a Closely Related Set of Immunoglobulins. The M1 (μ), M2 (μ), G3 ($\gamma 3$), and G6 ($\gamma 3$) immunoglobulins express the T15 idiotype (21). Claflin and Cubberly have suggested, from idiotypic and isoelectric focusing data, that antibodies bearing the T15 idiotype on IgG1, IgG2 and IgG3 heavy chains in individual BALB/c mice are identical (29). While the V_H regions of M1 and M2 are identical to the V_H region of T15, the V_H regions of the G3 and G9 immunoglobulins are not identical to that of T15. The G3 light chain also has at least one difference from T15 light chain. Our data suggest that among IgG immunoglobulins binding phosphorylcholine, sequence variation will occur within T15 idiotype-positive clones.

Summary

The hybridoma antibodies binding phosphorylcholine have provided important insights into the nature of diversity patterns in the V_H, the D, and the J_H segments. A striking feature of these antibodies is that the diversity in V_H segments derived from IgG and IgA antibodies is significantly greater than that seen in the V_H segments from IgM antibodies. Certain positions in IgG V_H segments have a relatively high

probability of accumulating amino acid substitutions. Finally, these data provide formal proof of the first unequivocal example of the combinatorial association of light and heavy chains and are consistent with the hypothesis that combinatorial association is a fundamental mechanism of information amplification in the antibody system.

References

1. 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.
2. Lenhard-Schuller, R., B. Hohn, C. Brack, M. Hirama, and S. Tonegawa. 1978. DNA clones containing mouse immunoglobulin κ chain genes isolated by in vitro packaging into phage λ coats. Proc. Natl. Acad. Sci. U.S.A. **75:** 4709.
3. Rudikoff, S., and M. Potter. 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. Proc. Natl. Acad. Sci. U.S.A. **73:** 2109.
4. 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.
5. Wu, T., and E. Kabat. 1970. An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma chains and their implications for antibody complementarity. J. Exp. Med. **132:** 211.
6. Capra, J. D., and J. Kehoe. 1974. Variable region sequences of five human immunoglobulin heavy chains of the V_HIII subgroup: Definitive identification of four heavy chain hypervariable regions. Proc. Natl. Acad. Sci. U.S.A. **71:** 845.
7. Amzel, L., R. Poljak, F. Saul, J. Varga, and F. Richard. 1974. The three-dimensional structure of a combining site ligand complex of immunoglobulin NEW at 3.5 Å resolution. Proc. Natl. Acad. Sci. U.S.A. **71:** 1427.
8. Padlan, E., D. Segal, T. Spande, D. Davies, S. Rudikoff, and M. Potter. 1973. Structure at 4.5 Å resolution of a phosphorylcholine-binding Fav. Nature New Biol. **245:** 165.
9. Baltimore, D. 1974. Is terminal deoxynucleotidyl transferase a somatic mutagen in lymphocytes? Nature (Lond.). **248:** 409.

10. Cohn, M., B. Blomberg, W. Geckeler, W. Rashke, R. Riblet, and M. Weigert. 1974. First order considerations in analyzing the generation of diversity. In The Immune System: Genes, Receptors, Signals. E. Sercarz et al., editors. Academic Press, New York. 89.

11. Hood, L., W. Gray, B. Sanders, and W. Dreyer. 1968. Light chain evolution. Cold Spring Harbor Symp. Quant. Biol. **32:** 133.

12. Milstein, C. 1967. Linked groups of residues in immunoglobulin kappa chains. Nature (Lond.). **216:** 330.

13. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. Nature (Lond.). **276:** 785.

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

15. Seidman, J., A. Leder, M. Nau, B. Norman, and P. Leder. 1978. Antibody diversity. Science. **202:** 11.

16. Max, E., J. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. Proc. Natl. Acad. Sci. U.S.A. **76:** 3450.

17. Sakano, H., K. Hüppi, G. Henrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature (Lond.). **280:** 288.

18. 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 (Lond.). **283:** 497.

19. 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 binding particular haptens. Cold Spring Harbor Symp. Quant. Biol. **41:** 817.

20. Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin-binding myeloma proteins. Proc. Natl. Acad. Sci., U.S.A. **75:** 1975.

21. Gearhart, P., N. Johnson, R. Douglas, and L. Hood. 1980. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. J. Exp. Med., this issue.

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

23. Ozols, J., and C. Gerard. 1977. Cleavage of tryptophanyl bonds in cytochrome b_5 by cyanogen bromide. J. Biol. Chem. **252:** 5986.

24. Hunkapiller, M., and L. Hood. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier, and high pressure liquid chromatography. Biochemistry. **17:** 2124.

25. Johnson, N., M. Hunkapiller, and L. Hood. 1979. Analysis of phenylthiohydantoin amino acids by high performance liquid chromatography on duPont Zorbax cyanopropylsilane columns. Anal. Biochem. **100:** 335.

26. Katz, A., W. Dreyer, and C. Anfinsen. 1959. Peptide separation by two-dimensional chromatography and electrophoresis. J. Biol. Chem. **234:** 2897.

27. Segal, D. M., E. A. Padlan, G. Cohen, S. Rudikoff, M. Potter, and D. Davies. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen-binding site. Proc. Natl. Acad. Sci. U.S.A. **71:** 4298.

28. Barstad, P., J. Hubert, M. Hunkapiller, A. Goetze, J. Schilling, B. Black, B. Eaton, J. Richards, M. Weigert, and L. Hood. 1978. Immunoglobulins with hapten-binding activity: structure-function correlations and genetic implications. Eur. J. Immunol. **8:** 497.

29. Claflin, J., and M. Cubberly. 1978. Clonal nature of the immune response to phosphorylcholine. VI. Molecular uniformity of a single idiotype among BALB/c mice. J. Immunol. **121:** 1410.

30. Kabat, E., T. Wu, and H. Bilofsky. 1979. Sequences of Immunoglobulin Chains. U.S. Department of Health, Education, and Welfare. Public Health Service, National Institutes of Health.

31. Dayhoff, M. O. 1976. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Maryland **5:** 189.

32. Hubert, J., N. Johnson, J. Greengard, B. Eaton, and L. Hood. 1980. The complete amino acid sequences of six V_H regions from myeloma proteins binding phosphorylcholine. Manuscript in preparation.

Chapter 4

The complete V domain amino acid sequences of two myeloma levan-binding proteins

This manuscript will be submitted to the Journal of Immunology

The complete V domain amino acid sequences of two myeloma proteins binding levan

N. JOHNSON, J. SLANKARD, L. PAUL and L. HOOD

From the Division of Biology, California Institute of Technology

Pasadena, California 91125

Running title: Levan antibodies

The amino acid sequence analysis of groups of myeloma proteins binding various haptens has provided insights into the structure (1-5), function (6, 7), genetics (8-10), and evolution (11) of antibody molecules. The myeloma proteins binding levan are appropriate in several regards for the study of the properties of antibody V domains. First, the amino acid sequence analysis of the variable regions of light (V_L) and heavy (V_H) chains from myeloma proteins binding levan demonstrate that their diversity is limited in extent. The nearly complete V_L sequences from three such myeloma proteins demonstrate that they differ from one another by four to seven residues (3). Likewise, the V_H regions from four myeloma proteins binding levan differ from one another by one to three residues (4). This limited diversity facilitates structure-function correlations, since critical amino acid interchanges are not obscured by other V region variability. Second, the affinities of the levan-binding myeloma proteins for oligosaccharides of different lengths have been studied (12-14). As a result, amino acid substitutions can be correlated with antigen specificity. Lastly, Lieberman et al. (15) have extensively studied the cross-reacting idiotypes of levan-binding myeloma proteins. The structure of these proteins provides an important basis for a more detailed understanding of the molecular nature of V region idiotypes (3).

In this report, we present the complete V domain sequences (V_L and V_H) of two additional immunoglobulins with specificity for levan. An analysis of these V domain sequences and sequences reported elsewhere reveals interesting patterns of V region diversity, and allows correlations to be drawn between V region structure, cross-reacting idiotypes, and antigen binding.

MATERIALS AND METHODS

Myeloma protein isolation. The BALB/c plasmacytomas W3082 (IgA, κ) and J606 (IgG3, κ) were obtained from the Salk Institute for Biological Sciences. Ascites was collected from these tumors and ammonium sulfate precipitations carried out

at 4°C for 2 hr in 40% saturated ammonium sulfate. Ammonium sulfate-precipitated protein was dissolved in 0.3 M NH_5CO_3 and immunoprecipitated with inulin (Nutritional Biochemicals Corporation). Optimal amounts of inulin were determined separately for each precipitation. Two hundred ml of ascites typically required 100 mg of inulin.

Chain separation. Myeloma protein-inulin precipitates were suspended in 0.5 M Tris-HCl, 2 mM EDTA, pH 8.2, and a 10% inulin solution was added dropwise until the precipitate dissolved. Dithiothreitol was added to a concentration of 50 mM and allowed to react for 2 hr at room temperature. The solution was then cooled to 0°C and iodoacetamide added to 110 mM. After 2 hr, guanidine-HCl was added to a concentration of 3 M or 5 M. The light and heavy chains of the W3082 protein were routinely separated in 3 M guanidine-HCl, 0.2 M NH_5CO_3 on LKB Ultrogel AcA34. Because of the relative insolubility of the IgG3 heavy chain, the J606 immunoglobulin chains were separated in 5 M guanidine-HCl, 0.2 M NH_5CO_3 on Sephadex G-150.

Cyanogen bromide digests. Cyanogen bromide digests were done in 70% formic acid, 2.5% cyanogen bromide at a protein concentration of 10 mg/ml for 24 hr at 4°C. The digests were lyophilized twice and chromatographed on Ultrogel AcA34, AcA44, or AcA54 in 3 M guanidine-HCl, 0.2 M NH_5CO_3. In some cases cyanogen bromide fragments were reduced and alkylated as above, except that guanidine-HCl was added to the reduction-alkylation buffer to 5 M.

Desalting. All column pools were desalted in 0.2 M NH_4OH on Sephadex G-10 with the exceptions of W3082H CN2 and J606HCN4 (see Results for identification of CNBr fragments). These were desalted in 5% HCO_2 on Biogel P-2 due to their insolubility in 0.2 M NH_4OH.

Enzyme digests. Trypsin, chymotrypsin, thermolysin, and α lytic protease digests were done at a 1:100 (w:w) enzyme:substrate concentration in 0.1 M NH_5CO_3 at 37°C with four hourly aliquots of enzyme. Carboxypeptidase A was used at 1:100 enzyme:substrate ratio at room temperature in 0.2 M NH_5CO_3 and aliquots were

removed from the reaction mixture at 2.5 min intervals for the first 10 min followed by ten 5 min and four 15 min intervals. Carboxypeptidase A was inactivated by boiling and all aliquots were analyzed on a Durrum D-500 amino acid analyzer. Norleucine was added to the enzyme digest as an internal standard.

Chromatography and electrophoresis. Two-dimensional paper chromatography and electrophoresis was performed as described previously (16). Analytical maps were dipped in 0.1% ninhydrin, 60:20:8 ethanol:acetic acid:collidine and heated at 100°C for 12 min to develop the colors. Preparatory maps were sprayed with 0.1% ninhydrin in ethanol, heated at 100°C to develop the colors, and the spots cut out and eluted as soon as they were visible. Acidic spots were eluted with 0.5 M ammonia and basic spots with 10% formic acid.

Succinylation. Succinylation reactions were performed in 5 M guanidine, 0.2 M Tris buffer using a 20-fold weight excess of succinic anhydride. The succinic anhydride was added in six aliquots over 30 min, while the pH was regulated at 9.5 using 3 M NaOH and an automatic titrater.

Tryptophan cleavage. Tryptophan cleavages were performed at room temperature in 35% cyanogen bromide, 1:1 88% formic acid:heptafluorobutyric acid as described in ref. 17. Tryptophan cleavage products were fractionated on Sephadex G-50 in 0.2 M NH_5CO_3.

Ion exchange chromatography. Peptides were fractionated on an ion exchange column of carboxymethylglycophase-controlled pore glass (Corning) using a concave gradient from 0.01 M sodium citrate, 8 M urea, pH 6.0 to 0.2 M sodium citrate, 1.5 M NaCl, 8 M urea, pH 6.0.

Automated sequence analysis. Automated sequential degradation was performed on a Beckman 890B sequenator and on a modified 890B (18). Amino acid analysis, gas chromatography (19), and high performance liquid chromatography (18) were

used to analyze the PTH amino acid derivatives. Variable regions are numbered according to Kabat, Wu and Bilofsky (20).

Isolation of heavy chain peptide 35-82. The CN2 fragment of W3082 was dissolved in 5% formic acid and 10% concentrated NH_4OH was added dropwise until a precipitate formed. After standing at 0°C for several hours, the precipitates were spun down and lyophilized. This technique is adapted from ref. 1.

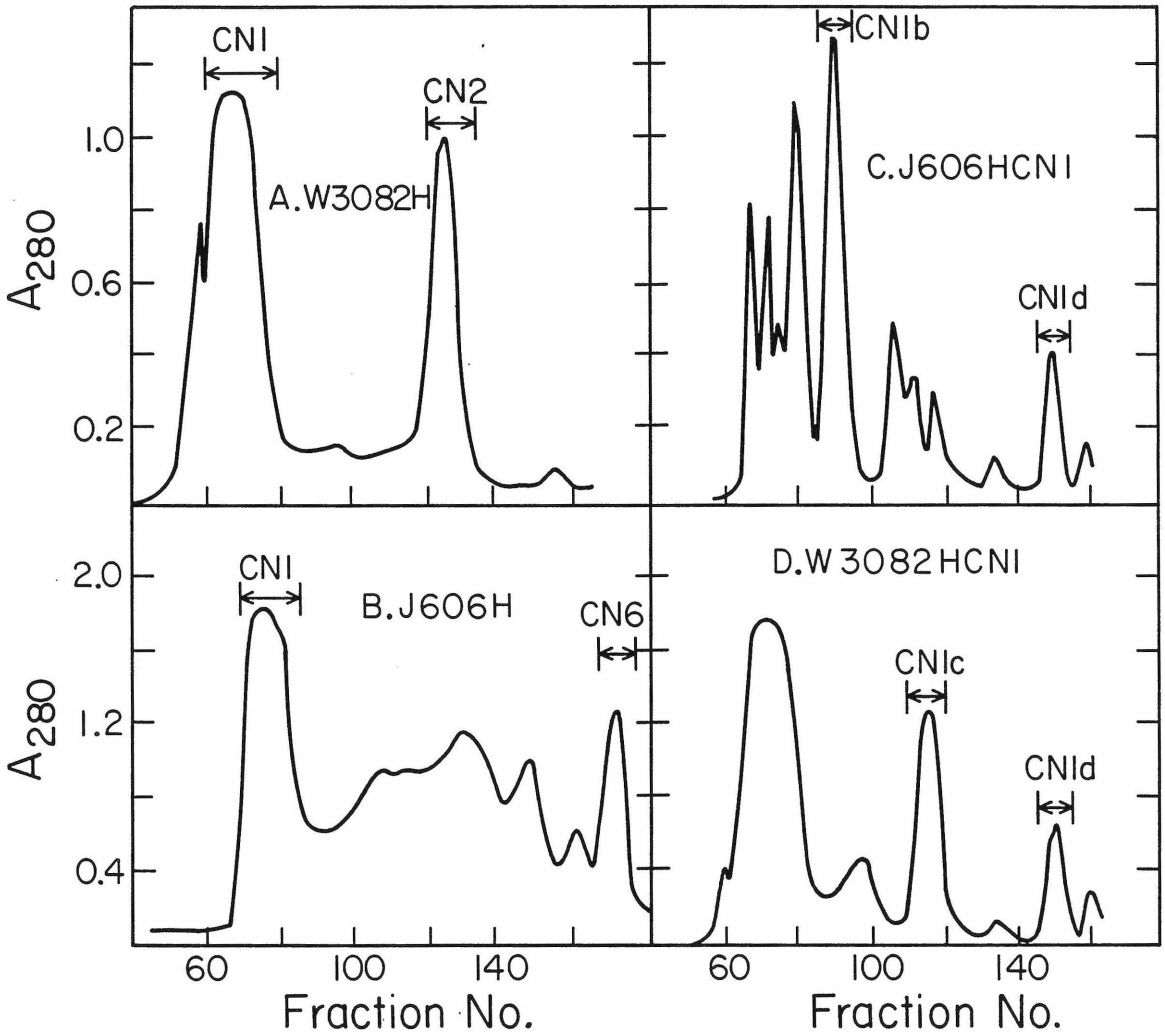
Subtractive Edman degradations. Subtractive Edman degradation was done as in ref. 21.

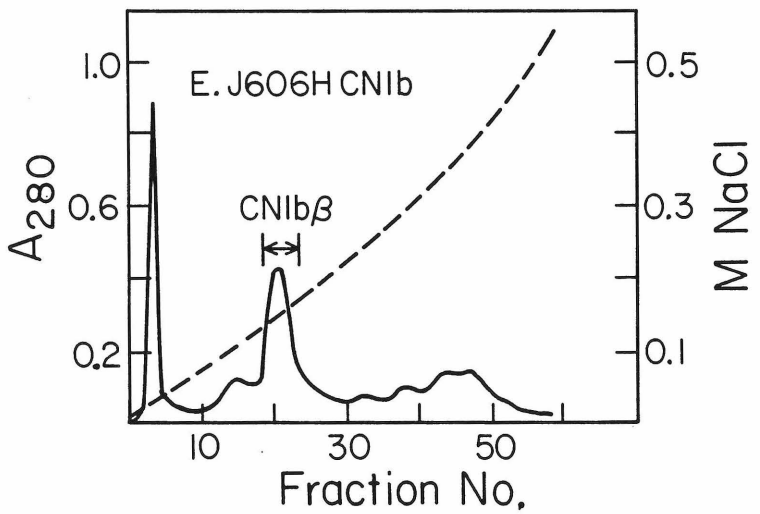
RESULTS

Isolation of heavy chain cyanogen bromide fragments. Purified W3082 and J606 heavy chains were digested with cyanogen bromide and fractionated by column chromatography as shown in Figure 1. The heavy chain from W3082 was run on Utrogel AcA44 and yielded two peaks, W3082H CN1 and W3082H CN2 (Fig. 1A). This fragmentation pattern is similar to that of Vrana et al. (4). The W3082H CN1 fragment contained the V region sequence from positions 19 to 34 and from 82a into the C region. This fragment was desalted, lyophilized, fully reduced and aklylated, and then rechromatographed on AcA44 (Fig. 1C) to yield the 82a to C region fragment as W3082H CN1c and the 19 to 34 peptide as W3082H CN1d. The W3082H CN2 fragment (Fig. 1A), containing the sequence from positions 35 to 82, was further purified by precipitation as described in Materials and Methods.

The J606 heavy chain was digested with cyanogen bromide and run on AcA34 (Fig. 1B). The J606H CN4 fragment contained only the peptide from positions 35 to 82 while the J606H CN1 peak was a mixture of several peptides. The J606H CN1 fragment was fully reduced and alkylated, run on AcA44, and the CN1b and CN1e peaks collected (Fig. 1D). Ion exchange chromatography on CN1b was used to isolate the peptide CN1bβ, which extended from 82a into the C region (Fig. 1E). The CN1e fragment is the peptide extending from positions 19 to 34.

Figure 1. The cyanogen bromide fragmentation patterns of the heavy chains from J606 and W3082. All of the column chromatography was done in 3 M guanidine, 0.2 M NH_5CO_3. A. Mildly reduced and alkylated W3082H digested with cyanogen bromide and chromatographed on AcA44. B. Mildly reduced and alkylated J606H digested with cyanogen bromide and chromatographed on AcA34. C. Fully reduced and alkylated J606H CN1 chromatographed on AcA44. D. Fully reduced and alkylated W3082H CN1 chromatographed on AcA44. E. The J606H CN1b peak from C. run on carboxymethyl-glycophase-controlled pore glass. Peaks were eluted with increasing concentrations of NaCl in citrate-buffered 8 M urea, pH 6.0 (see Materials and Methods).


CN1
CN2
A.W3082H
CN1b
C.J606HCN1
CN1d
CN1
B.J606H
CN6
D.W3082HCN1
CN1c
CN1d
A280
1.0
0.6
0.2
2.0
1.2
0.4
60
100
140
Fraction No.


E.J606H CN1b
CN1bβ
A280
1.0
0.6
0.2
0.5
0.3
0.1
M NaCl
10
30
50
Fraction No.

Heavy chain sequence determination. Sequencing strategies for the V_H regions from W3082 and J606 are summarized in Figure 2. The N-terminal 34 residues of the V_H region from W3082 were determined from sequenator runs on the intact W3082 heavy chain and on the CN1d fragment. The residues between 35 and 82 were identified by a single sequencer run on W3082H CN2p with the exception of certain serines and the C-terminal two amino acids. These missing residues were determined by compositional data and sequence analyses of peptides produced by tryptic, alpha-lytic protease, and simultaneous tryptic plus chymotryptic digestions. A single sequenator run on the CN1c fragment identified the remaining V_H region residues.

The heavy chain from J606 was sequenced in much the same way as the W3082 heavy chain. The N-terminal 34 residues were identified with N-terminal sequence analyses of the intact heavy chain and the CN1h fragment. Residues 35-82 were mostly identified with an N-terminal sequenator run on the CN4 fragment. Trypsin, chymotrypsin, alpha-lytic protease, and carboxypeptidase A were used to identify the C-terminal residues of the CN4 fragment. A single run on the CN1bβ fragment was adequate to determine the remainder of the V_H region sequence.

Isolation of cyanogen bromide fragments from light chains. Purified light chains from W3082 and J606 yield identical patterns when cleaved with cyanogen bromide and chromatographed on Ultrogel AcA54 (Fig. 3A). The CN2 fragment corresponds to a peptide extending from positions 82-214 and the CN4 fragment is a partially separated mixture of peptides extending from positions 5-83 and 84-175. This chromatographic fractionation pattern is similar to that in ref. 3. The CN4 fragment was isolated, succinylated, cleaved at tryptophane with cyanogen bromide and fractionated on Sephadex G-50 (Fig. 3B). The second peak (CN3b) showed a single sequence starting at 35. Intact light chains also were digested with thermolysin to yield the peptide extending from positions 78-86.

Figure 2. Sequencing strategies for the heavy chain from J606 and W3082.

A. W3082H; B. J606H.

⟶	residues determined by automatic sequenator
⟶	residues determined by subtractive Edman degradation
⟵	residues determined by carboxy peptidase A
⟷	residues analyzed using amino acid compositional data
( )	peptide from an α lytic protease digest
[ ]	peptide from a trypsin digest
{ }	peptide from a simultaneous digest with trypsin and chymotrypsin

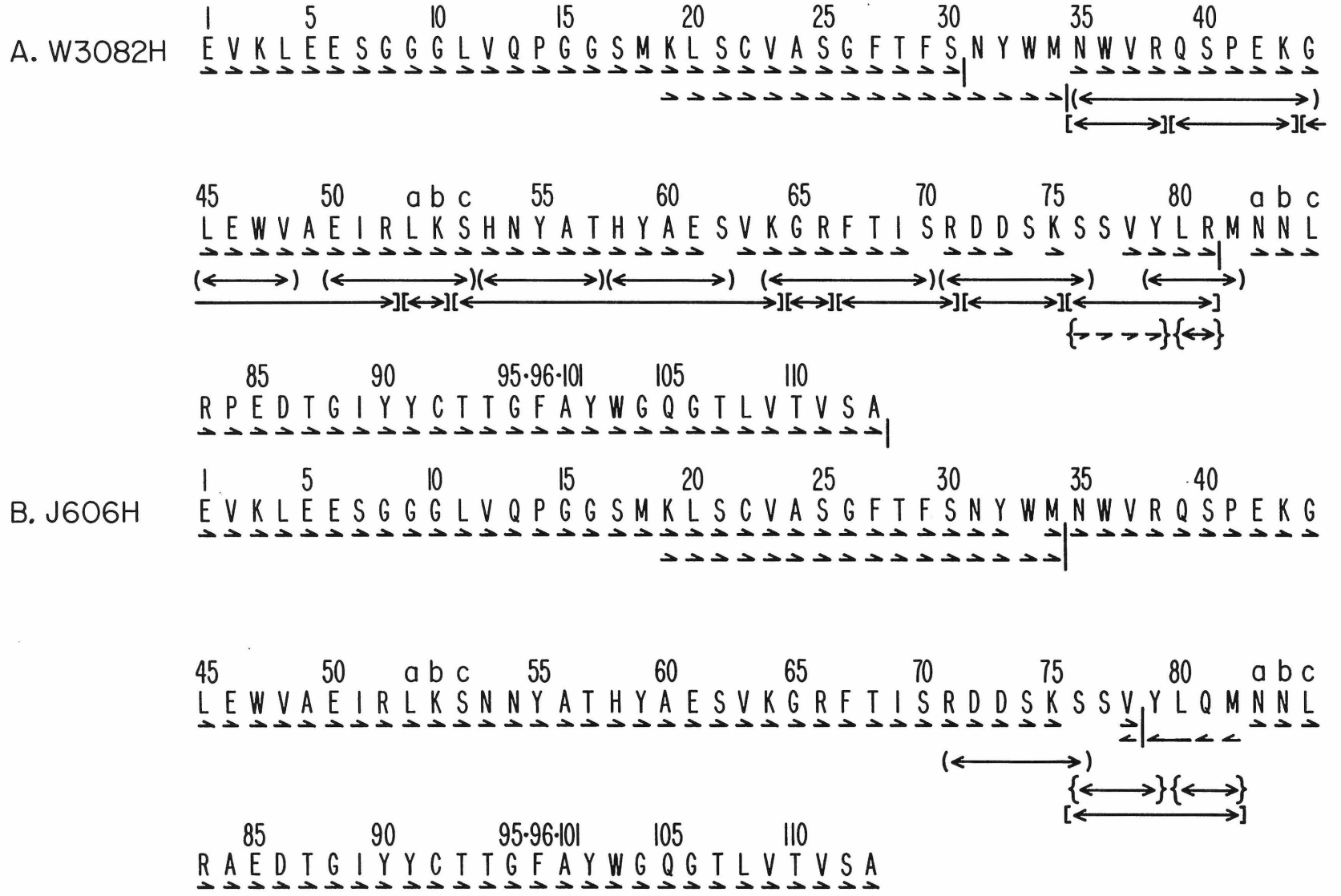
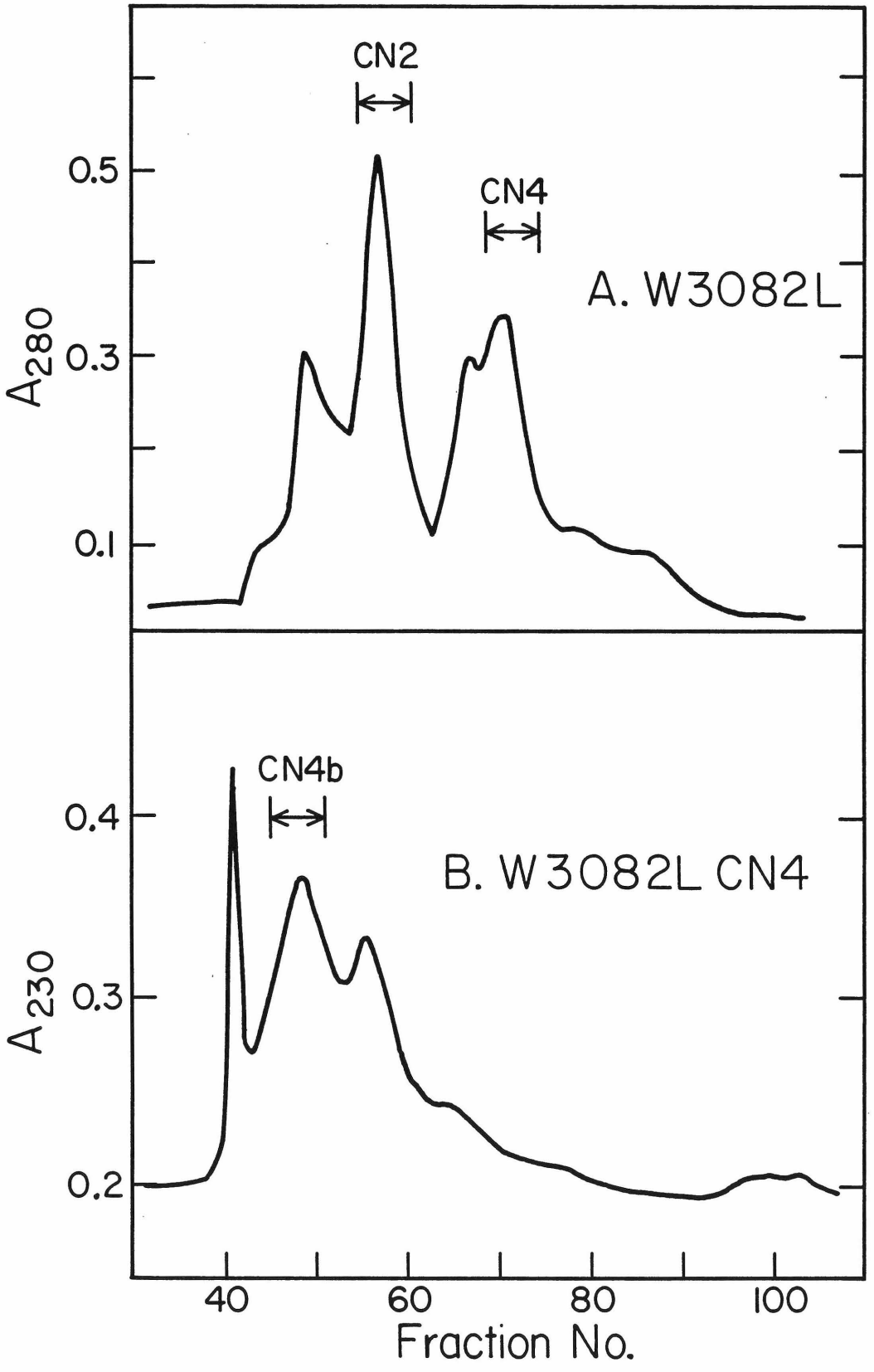

A. W3082H
1 5 10 15 20 25 30 35 40
E V K L E E S G G G L V Q P G G S M K L S C V A S G F T F S N Y W M N W V R Q S P E K G
45 50 a b c 55 60 65 70 75 80 a b c
L E W V A E I R L K S H N Y A T H Y A E S V K G R F T I S R D D S K S S V Y L R M N N L
85 90 95·96·101 105 110
R P E D T G I Y Y C T T G F A Y W G Q G T L V T V S A
B. J606H
1 5 10 15 20 25 30 35 40
E V K L E E S G G G L V Q P G G S M K L S C V A S G F T F S N Y W M N W V R Q S P E K G
45 50 a b c 55 60 65 70 75 80 a b c
L E W V A E I R L K S N N Y A T H Y A E S V K G R F T I S R D D S K S S V Y L Q M N N L
85 90 95·96·101 105 110
R A E D T G I Y Y C T T G F A Y W G Q G T L V T V S A

Figure 3. The cyanogen bromide fragmentation patterns of the light chains from J606 and W3082. Light chains from W3082 and J606 gave identical cyanogen bromide fragmentation patterns. A. A cyanogen bromide digest of fully reduced and alkylated W3082L chromatographed on AcA54 in 3 M guanidine, 0.2 M NH_5CO_3. B. The CN4 peak from A. succinylated, cleaved at tryptophane residues, and chromatographed on Sephadex G-50 in 0.2 M NH_5CO_3.


CN2
CN4
A. W3082L
A_{280}
0.5
0.3
0.1
CN4b
B. W3082L CN4
A_{230}
0.4
0.3
0.2
40
60
80
100
Fraction No.

Light chain sequence determination. The J606 and W3082 light chains were sequenced by the strategies shown in Figure 4 for W3082L. The N-terminal 38 residues were determined by sequenator runs on intact light chains. Sequenator runs on the CN4 peptides (residues 84-214, Fig. 3a) identified the residues from 84 into the C region. The residues from 35-81 were identified by sequenator runs on the CN4b peptides (residues 35-83, Fig. 3b), and residues 82 and 83 were determined by sequence analysis of the thermolysin overlap peptide from 78-86.

DISCUSSION

V_H regions from myeloma immunoglobulins binding levan show variability primarily in framework positions. Figure 5 shows the complete sequences of six heavy chains from myeloma proteins with specificity for levan. These six heavy chain sequences differ from the A4 prototype sequence by one to three substitutions, only one of which occurs in a hypervariable region. Each of the nine amino acid interchanges requires only a one-base substitution and each occurs at a different V region position. This pattern of substitutions in the levan V_H regions differs in a striking manner from those noted for the V_H regions from immunoglobulins binding phosphorylcholine (1, 22-24), dextran (9), and galactan (25). In each of these latter examples, a substantial fraction of the variant residues lie within the hypervariable regions. Sixteen V_H regions from myeloma and hybridoma immunoglobulins binding phosphorylcholine differ in 47 positions, 36 of which are in hypervariable regions. Twenty-one V_H regions from myeloma and hybridoma proteins binding dextran differ by only 24 framework residues out of a total of 78 amino acid substitutions. Four galactan-binding myeloma V_H regions differ by eight hypervariable residues and 12 framework residues. Thus the V_H regions from myeloma proteins binding levan have more conserved hypervariable regions than any other well-studied set of immunoglobulin heavy chains with specific antigen-binding properties.

Figure 4. Sequencing strategies for the light chains from J606 and W3082. The sequencing strategies were the same for both J606L and W3082L. Symbols are as in Figure 2 with the exception of ( ), which indicates a peptide from a thermolysin digest.

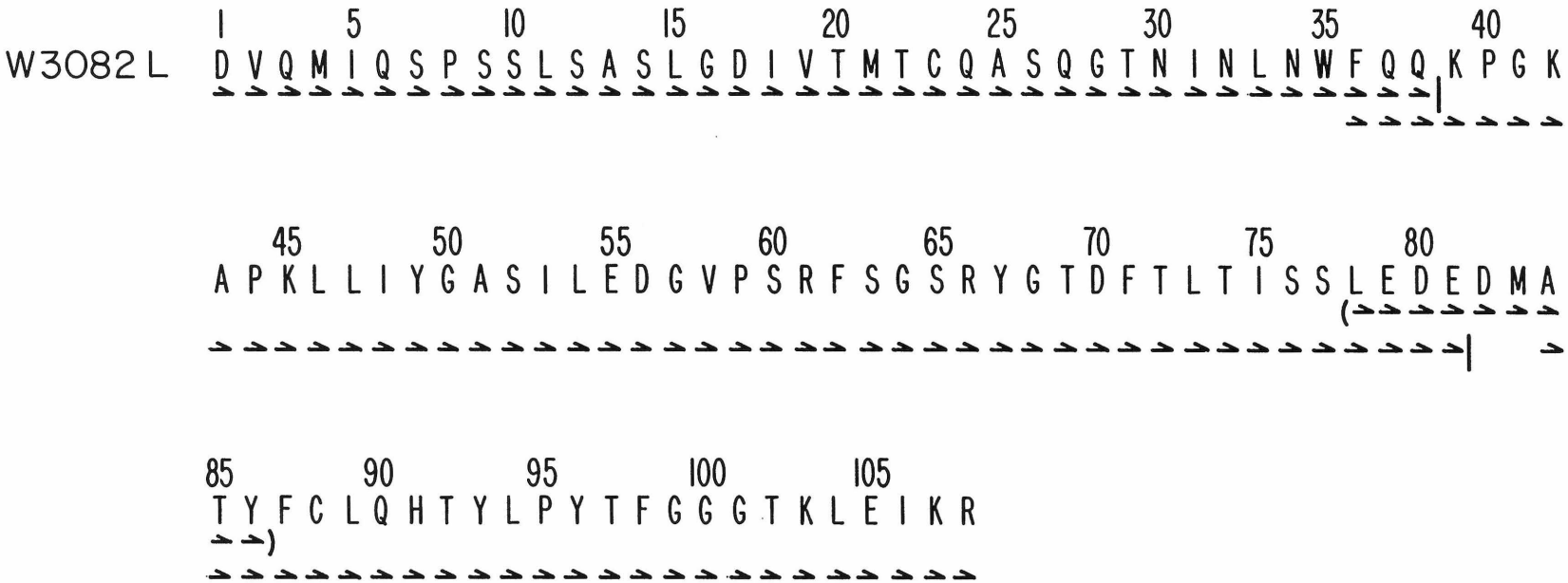
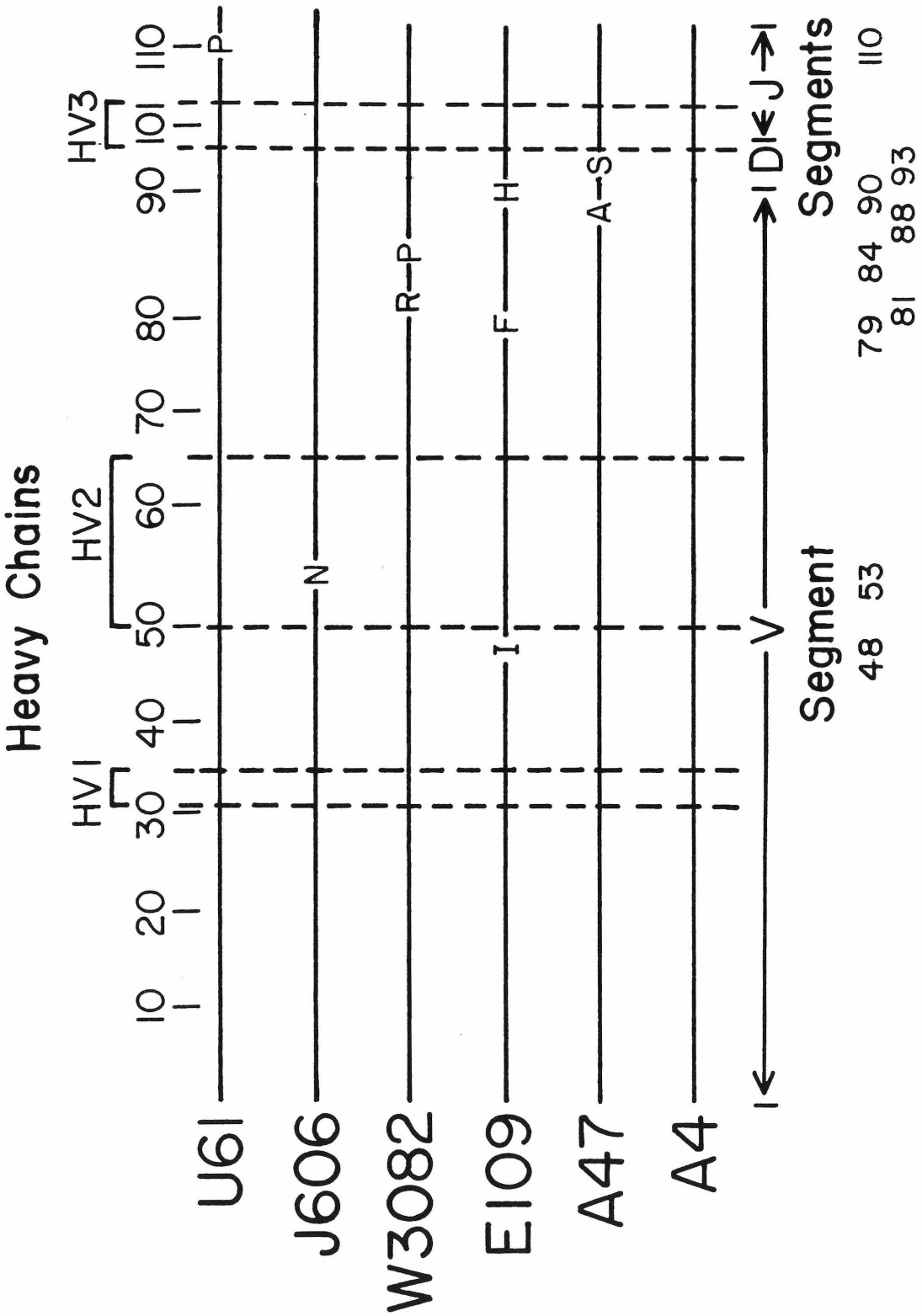

W3082L
1 5 10 15 20 25 30 35 40
D V Q M I Q S P S S L S A S L G D I V T M T C Q A S Q G T N I N L N W F Q Q K P G K
45 50 55 60 65 70 75 80
A P K L L I Y G A S I L E D G V P S R F S G S R Y G T D F T L T I S S L E D E D M A
85 90 95 100 105
T Y F C L Q H T Y L P Y T F G G G T K L E I K R

Figure 5. The complete sequences of six V_H regions from myeloma proteins with specificity for levan. Unbroken lines indicate identity to the prototype sequence from the A4 protein. Substitutions are designated using the one-letter code of Dayhoff (29), and their positions are labeled at the bottom of the figure. Hypervariable regions (HV1, HV2, HV3) and the V, D, and J segments are as shown. The V_H sequences of U61, A4, A47, and E109 are from Ref. 4.


Heavy Chains
HV1
HV2
HV3
10
20
30
40
50
60
70
80
90
101
110
U61
J606
W3082
E109
A47
A4
N
R
P
F
H
I
A
S
P
V
I
D
J
Segment
48 53
Segments
79 84 90
81 88 93
110

The conservation of heavy chain hypervariable regions from myeloma proteins binding levan may reflect fairly strict structural constraints imposed by antigen binding. The framework regions, while more diverse than the corresponding hypervariable regions, are not more diverse than framework regions of other V_H regions with identified antigen-binding specificities.

The V_H segments from myeloma proteins binding levan are probably encoded by at least two closely related germline genes. Antibody V_H regions are encoded by three distinct gene segments, V_H, D, and J_H. DNA analyses demonstrate that the V_H gene segment from myeloma proteins binding levan extends from codons 1-92; the D gene segment codes for positions 93, 94, and the first nucleotide of 95; and the J gene segment extends from the second nucleotide of codon 95 to codon 113 (26). Antibody V gene segments which differ by one or a few single base substitutions (such as those seen in Fig. 5) are frequently thought to arise as somatic variants from a single germline V gene segment (25, 27).

This traditional view would suggest that the V_H segments from myeloma proteins binding levan arise from a single germline V_H gene segment encoding the prototype A4 and E109 sequences. A V_H gene segment that codes for the J606 V_H segment has recently been isolated and sequenced (S. Crews, personal communication). Since this V_H gene segment was isolated from a BALB/c sperm genomic library, the J606 V_H segment must represent a germline sequence. Traditionally, however, the A4 V_H segment would be considered to be the germline sequence for this group of proteins. This paradox can be resolved in one of three ways. i) The J606 germline gene gave rise to the other five V_H segments by somatic mutation. This would require parallel substitutions in all of the V_H segments but J606. ii) The V_H regions from levan binding proteins may be encoded by two extremely closely-related V_H segments, J606 and A4. iii) The J606 and A4 V_H gene segments may represent polymorphic forms of a single genetic locus.

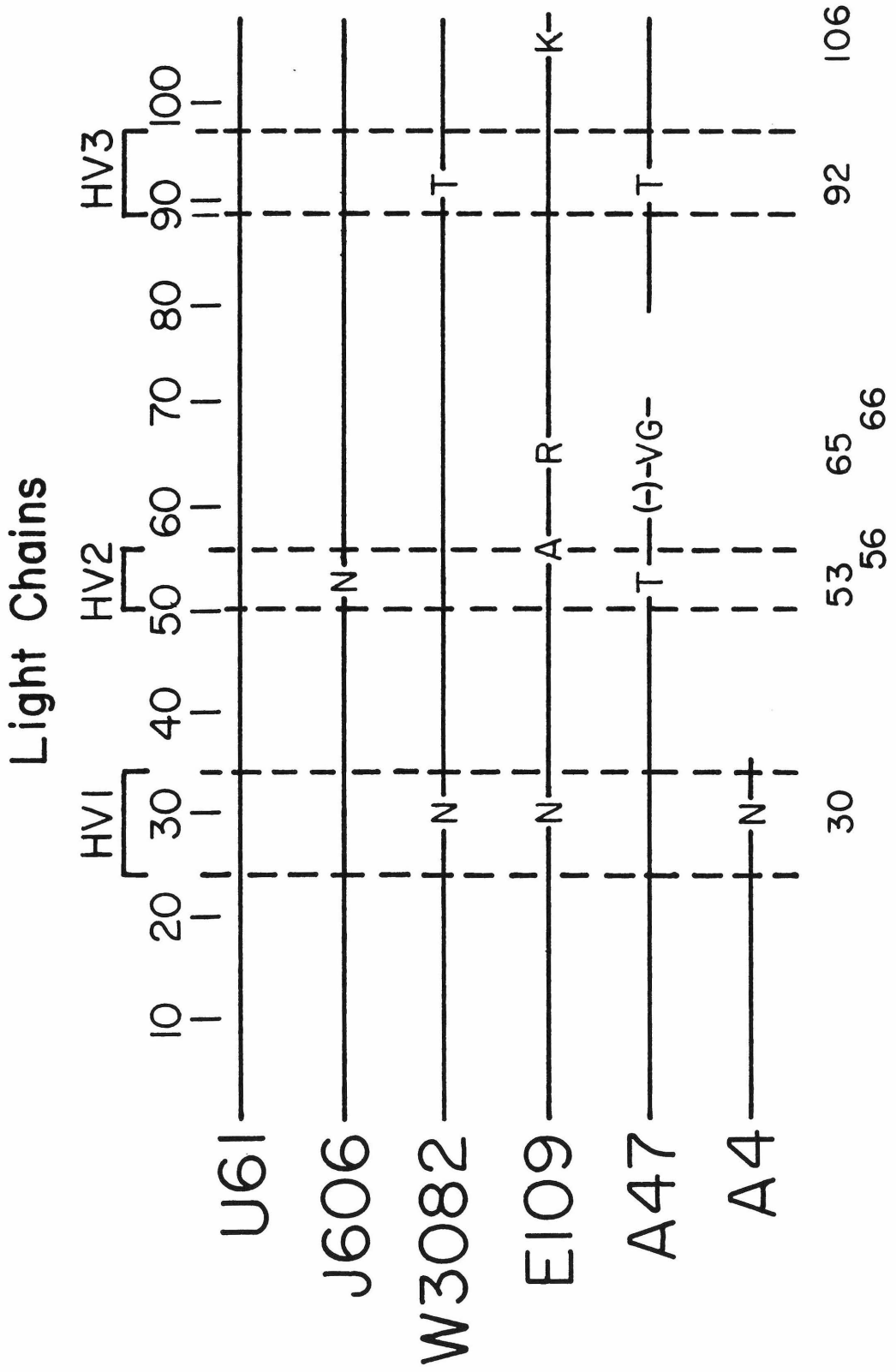
Levan-binding heavy chains have only one D segment substitution. As a result of D segment diversity, the third hypervariable region is the most variable part of an immunoglobulin heavy chain. This is clearly true of the V_H regions from myeloma proteins binding phosphorylcholine (1, 22-24) and dextran (9), where, respectively, 57% and 62% of the total V_H diversity lie in the D segment. In the galactan-binding V_H regions, the D segment is somewhat less diverse, accounting for 35% of the total diversity (25).

In contrast to the D segments from proteins binding phosphorylcholine, dextran, and galactan, there is only one D segment substitution in the heavy chains from proteins binding levan. This lack of D segment variability corresponds to the lack of variability in other V_H hypervariable regions, and may be due to selection by antigen. Alternatively, it may reflect an unknown restriction in the association of certain V_H and D gene segments.

The V_L regions from myeloma proteins binding levan show diversity in the framework as well as in the hypervariable regions. The six V_L regions from myeloma proteins binding levan are each different from one another, varying from the prototype sequence (U61) by one to four residues (Fig. 6). These V_L regions differ by 12 substitutions, 11 of which can be explained by one-base changes and one by a two-base change. Eight of these substitutions are in hypervariable regions. Unlike the V_H regions, several substitutions in the V_L regions occur at the same position. There are three V_L regions with identical substitutions at position 30, two identical substitutions at position 92, and two proteins each with nonidentical substitutions at positions 53 and 65. Thus nine out of 12 substitutions occur at four positions.

Light chains from myeloma proteins binding levan have parallel substitutions which show no correlation with one another. Positions 30 and 92 each have substitutions that appear two times in five examples. The residue alternatives at positions 30 and 92 assort in such a way that all four possible combinations of residues at these

Figure 6. The sequences of five V_L regions from myeloma proteins with specificity for levan. Sequences are diagrammed and labeled as in Figure 5. ( ) indicates a probable residue assignment and the gap in the A47 V_L region is an area of unknown sequence. The V_L sequences U61, A4, A47, and E109 are from Ref. 3.

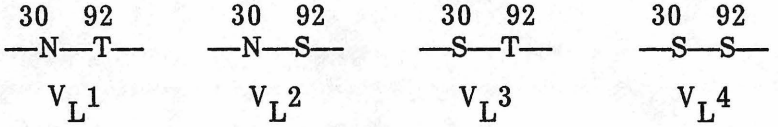

Light Chains
HV1
HV2
HV3
10
20
30
40
50
60
70
80
90
100
U61
J606
W3082
E109
A47
A4
N
N
N
N
T
A
R
K
T
(-)-VG-
T
30
53
56
65
66
92
106

two positions occur. Thus both serine and asparagine at position 30 occur in chains having both threonine and serine at position 92. The variability at positions 30 and 92 could arise in one of three ways as illustrated in Figure 7. i) The four V_L segments may be directly encoded in the germline (Fig. 7A). This possibility is unattractive because of the close similarity of the V_κ gene segments. ii) These variants could arise by the somatic mutation of germline genes. If there are one, two, or three germline V_L segments, then this somatic mutation would require either mutation to a preexisting germline sequence or to a preexisting somatic sequence (parallel mutation). Figure 7B shows a somatic mutation pathway which could generate the observed sequences from one germline V_L gene segment. In the case of the residues at V_L position 30, somatic mutation must create identical substitutions in three different individuals. iii) Somatic recombination between two closely related V_L gene segments could generate the four possibilities (Fig. 7C). The first possibility could be distinguished from the other two possibilities by sequence analysis of germline V_H gene segments. Unless one postulates parallel somatic mutations (Fig. 7B), these theories require at least two very closely related germline V_L gene segments encoding the V_L segments of myeloma proteins with specificity for levan.

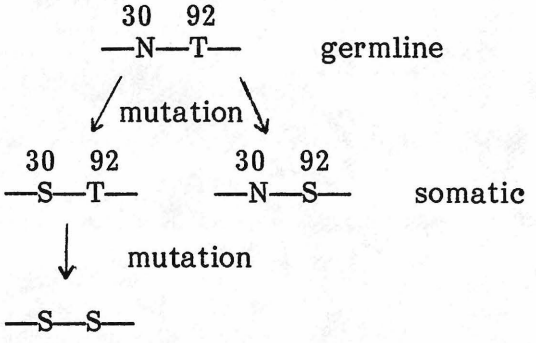
Sequence correlates of antigen binding. The data presented here provide an example of how sequence data and antigen-binding studies can complement each other in making structure-function correlations. The hypervariable region sequence correlates of antigen binding are presented in Table 1 along with the length of oligofructosan bound with maximum affinity by each myeloma proteins (12-14). The heavy chain hypervariable regions without sequence diversity are not shown. The only obvious correlation between the length of oligofructosan bound with maximum affinity and hypervariable-region sequence occurs in the second hypervariable region of the light chain. Here the presence of a residue other than isoleucine at position 53 is associated with maximum affinity for a tetrasaccharide, while isoleucine at position

Figure 7. Mechanisms for the generation of diversity at residues 30 and 92 in the V_L regions of myeloma proteins with specificity for levan. The amino acid residues at positions 30 and 92 encoded by the V_L gene segments are labeled using the one-letter code of Dayhoff. N = asparagine, S = serine, T = threonine (29).

A. Four germline V_L gene segments



B. Somatic mutation



C. Somatic recombination

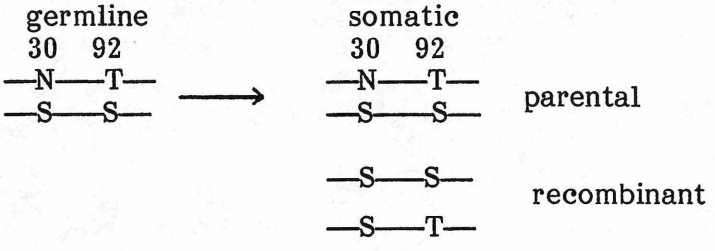


Figure 7

Table 1

Hypervariable region comparisons among levan-binding myeloma proteins. The V_L (HV1, HV2, HV3) and V_H (HV2) hypervariable region differences are depicted for each myeloma protein. Substitution positions are shown above the amino acid interchanges. The oligosaccharide-binding affinity of W3082 is from ref. 12, the affinity of J606 is from ref. 13, and the affinities of A47, A4, U61, and E109 are from ref. 14. Binding affinities are discussed in ref. 5.

		V_L region			V_H region
		HV1	HV2	HV3	HV2
		-30-	-53-56-	-92-	-53-
Proteins with maximum affinity for tetrasaccharide	J606	-Ser-	-Asn-Asp-	-Ser-	-Asn-
	A47	-Ser-	-Thr-Asp-	-Thr-	-His-
	A4	-Asn-	?	?	-His-
Proteins with maximum affinity for pentasaccharide	U61	-Ser-	-Ile-Asp-	-Ser-	-His-
	E109	-Asn-	-Ile-Ala-	-Ser-	-His-
	W3082	-Asn-	-Ile-Asp-	-Thr-	-His-

53 is associated with maximum affinity for pentasaccharide The aspargine-serine interchange at V_L position 30 and the threonine-serine interchange at V_L position 92 do not cause a reduction in the length of oligosaccharide required for maximum affinity. However, it is possible that the asparagine-histidine interchange at V_H position 53 limits the size of the J606 levan-binding site. Thus, antigen-binding data can be used to determine the possible effects of amino acid substitutions. Amino acid substitutions which correlate with antigen binding are possible sites of selection by antigen. Such selection might operate if an antigen stimulated only those clones with certain antigen-binding specificities to divide.

The structures of the V domains from myeloma proteins binding levan provide fundamental insights into the nature of idiotypic determinants. The V domains of the levan-binding myeloma proteins afford a unique opportunity to make correlations between idiotype specificities and primary amino acid sequence because i) the complete V_H sequences are known for six myeloma proteins and the complete V_L sequences are known for four and part of a fifth protein; ii) the idiotypic specificities of these proteins have been extensively analyzed (Table 2) (15); and iii) these proteins generally differ in their V domains by only a few residues from one another (Figs. 5 and 6; Table 2). Accordingly, it is possible to determine some of the molecular properties of the crossreacting idiotypes shown in Table 2.

Lieberman et al. have reported a comprehensive study of the idiotypes of myeloma proteins which bind levan, including the six proteins shown in Figures 5 and 6 (15). Using antisera raised against these levan-binding myeloma proteins, they have defined 10 crossreacting idiotypes (IdX) and determined the hapten inhibitability of each IdX specificity. Their IdX specificity assignments are shown in Table 2. IdX specificities are determined by the ability of myeloma proteins to inhibit the hemagglutination of red blood cells coated with levan-binding myeloma proteins. For example, IdXA is defined by the hemagglutination of A4-coated red blood cells

Table 2

Cross-reactive idiotypes and V region sequences of levan-binding myeloma proteins. Myeloma protein hemagglutination inhibition titers are given under each IdX specificity (see text for explanation of hemagglutination inhibition assay). The hapten inhibitability of each IdX specificity is given at the bottom of the table. V region sequence differences are shown schematically to the right, and the position of each substitution is indicated at the top. IdX specitificies and hapten inhibition data are from ref. 15.

	IdX specificities										V region sequence differences	
Myeloma proteins	A	B	C	D	E	F	G	H	I	J	V_H Substitution positions -48-53-79-81-84-88-90-93-110-	V_L Substitution positions -30-53-56-65-66-92-106-
U61	7	8	8	7	10	6	12	5	7	0	——————————P-	——————————
A4	9	0	5	0	3	0	10	6	2	0	——————————	-N-
J606	0	0	0	2	5	2	4	0	2	0	——N——————————	——N——————
A47	0	0	0	0	0	0	0	0	7	6	——————A——S———	——T——V—G- -T———
W3082	8	9	8	5	10	8	12	0	6	6	————R—P————	-N————————T———
E109	6	0	6	5	11	0	9	0	4	0	-I———F————H———	-N———A—R————K-
Hapten inhibition	+	+	+	+	+	+	+	+	-	-		

with antisera raised against W3082. Levan-binding myeloma proteins which inhibit this hemagglutination are typed as carrying the IdXA determinant. The numbers in Table 2 refer to the $-\log_2$ of the dilution of a 1 mg/ml solution of myeloma protein necessary for the inhibition of hemagglutination of an IdX typing assay. Table 2 also contains a schematic representation of the amino acid interchanges in the levan-binding myeloma proteins. In the following section, we discuss correlations between IdX determinants and V domain amino acid interchanges.

The IdXA and IdXC idiotypes. The IdXA and IdXC specificities have identical distributions and are hapten-inhibitable (Table 2). They are present on the myeloma proteins U61, A4, W3082, and E109. The only residue shared exclusively by these molecules is the isoleucine at position 53 in the V_L region (the residue in A4 is unknown). It is tempting to postulate a structural correlation between V_L residue 53 and IdXA; however, it has been shown that IdXA is dependent only on the heavy chain (28). Since the A47 protein ($IdXA^-$, $IdXC^-$) and the W3082, E109, U61, and A4 proteins ($IdXA^+$, $IdXC^+$) have identical V_H hypervariable regions, the hapten-inhibitable IdXA and IdXC determinants must include framework as well as hypervariable region residues.

The IdXB and IdXF idiotypes. The IdXB and IdXF specificities have identical distributions (Table 2). The IdXB specificity is situated primarily on the light chain (28) and is hapten-inhibitable. These idiotypes correlate best with residues 53 and 56 in the C-terminal part of the second V_L hypervariable region. Our data contradict the conclusions of Vrana et al. who suggested that the IdXB idiotype correlated with serine residues at V_L positions 30 and 92 (3). The W3082 protein has the IdXB idiotype and has asparagine at position 30 and threonine at position 92.

The IdXE and IdXG idiotypes. The IdXE and IdXG specificities are difficult to assign in that they are present on all of the myeloma proteins but the A47 protein, which is not completely sequenced.

The IdXH idiotype. The IDxH idiotype is hapten-inhibitable and present only on the A4 and U61 proteins. The U61 and A4 proteins share their V_H hypervariable regions with all of the levan-binding myeloma proteins except for the second hypervariable region of the J606 protein. Each of the V_L hypervariable regions of the U61 and A4 proteins are shared with at least three other V_L regions (only the first V_L hypervariable region from A4 has been sequenced). The U61 protein does not, however, share all of its V_L hypervariable regions with any one V_L region and thus the IdXH specificity could be dependent on all three hypervariable regions. Alternatively, the IdXH idiotype could be determined by framework as well as hypervariable regions.

The IdXI idiotype. The IdXI specificity is present on all the myelomas. It is impossible to correlate it with particular V region residues.

The IdXJ idiotype. The IdXJ specificity is not hapten-inhibitable even though it is correlated with the threonine residue at position 92 in the V_L third hypervariable region. Hence, either the threonine at position 92 does not participate in antigen binding, or the IdXJ specificity is dependent on residues in other parts of the V domain.

This discussion correlating idiotypic specificities with V domain structure for the myeloma proteins binding levan is admittedly incomplete. It does, however, emphasize some general ideas about idiotypy which are important in using idiotypes as an assay for antibody V region structure. These ideas are the following: i) Distinct idiotypes can arise from the same amino acid interchange. The J606 and U61 proteins differ by four hapten-inhibitable idiotypes and at the same time differ by only three amino acid substitutions of which only two are in hypervariable regions. Thus, single amino acid residues can be involved in determining multiple idiotypic sites. One possible mechanism for this is that certain amino acid interchanges may cause major conformational alterations in the V domain, thus generating new idiotypic sites. Alternatively, idiotypic sites may overlap so that several idiotypes encompass the same amino acid residue. ii) Hapten-inhibitable idiotypes may depend on residues

both within and outside the hypervariable regions. For example, the IdXA idiotype is present on the heavy chains of the W3082, U61, A4, and E109 proteins. These heavy chains have identical V_H hypervariable regions to the A47 protein, which lacks the IdXA specificity. Thus, the hapten-inhibitable IdXA idiotype must be dependent on residues outside as well as residues inside the antigen-binding site. iii) Anti-idiotypic antisera may concentrate on certain portions of the antibody V domain. It is possible to make correlations between idiotype and structure for seven of the ten IdX determinants (all but IdXE, IdXG, and IdXI). Three of the seven idiotypes correlate with the C-terminal portion of the V_L second hypervariable region. Hence, several IdX determinants may be concentrated in a particular region of the V domain and fail to identify variation in other areas. For example, the J606 and E109 proteins share more idiotypes than the J606 and U61 proteins. The J606 and U61 proteins differ by four idiotypes while the J606 and E109 proteins differ by only three. The J606 protein, however, differs by only three amino acid substitutions from the U61 protein and by nine substitutions from the E109 protein. Idiotypic similarity does not always accurately predict sequence similarity.

SUMMARY

The myeloma proteins binding levan afford a unique opportunity to study the V region patterns of variation because of the similarity in their V_L as well as their V_H regions. The diversity patterns in both the V_L and V_H regions suggest that these proteins are encoded by multiple, very similar V gene segments or that somatic mutation to preexisting sequences can occur. Because of the close similarity in the V domains of these proteins and the extensive idiotypic analyses that have been previously carried out, several interesting conclusions can be drawn about the nature of idiotypic determinants. First, distinct idiotypes can arise from the same amino acid interchange. Second, hapten-inhibitable idiotypes may depend on residues within and outside the hypervariable regions. Third, idiotypic similarity does not always predict a corresponding sequence similarity. In addition, antigen binding studies allow the roles of amino acid interchanges in antibody-antigen interactions to be investigated.

REFERENCES

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

2. Padlan, E., D. Segal, T. Spande, D. Davie, S. Rudikoff, and M. Potter. 1973. Structure at 4.5 Å resolution of a phosphorylcholine-binding Fab. Nature New Biol. **245:** 165.

3. Vrana, M., S. Rudikoff, and M. Potter. 1979. The structural basis of a hapten-inhibitable κ chain idiotype. J. Immunol. **122:** 1905.

4. Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin-binding myeloma proteins. Proc. Natl. Acad. Sci., U.S.A. **75:** 1957.

5. Potter, M. 1977. Antigen-binding myeloma proteins of mice. Adv. Immunol. **25:** 141.

6. Amzel, L., R. Poljak, F. Saul, J. Varga, and F. Richard. 1974. The three-dimensional structure of a combining site ligand complex of immunoglobulin NEW at 3.5 Å resolution. Proc. Natl. Acad. Sci., U.S.A. **71:** 1427.

7. Segal, D., E. Padlan, G. Cohen, S. Rudikoff, M. Potter, and D. Davies. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen-binding site. Proc. Natl. Acad. Sci., U.S.A. **71:** 4298.

8. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. Nature (Lond.) **276:** 785.

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 (Lond.) **283:** 35.

10. 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..

11. Riesen, W., D. Braun, and J.-C. Jaton. 1976. Human and murine phosphorylcholine-binding immunoglobulins: conserved subgroup and first hypervariable region of heavy chains. Proc. Natl. Acad. Sci., U.S.A. **73:** 2096.

12. Cisar, J., E. Kabat, J. Liao, and M. Potter. 1974. Immunochemical studies on mouse myeloma proteins reactive with dextrans or with fructosans and on human antilevans. J. Exp. Med. **139:** 159.

13. Lundblad, A., R. Steller, E. Kabat, J. Hirst, M. Weigert, and M. Cohn. 1972. Immunochemical studies on mouse myeloma proteins with specificity for dextran or for levan. Immunochemistry **9:** 535.

14. Streefkerk, D., and C. Glaudemans. 1977. Binding studies on antifructofuranan mouse myeloma immunoglobulins A47N, A4, U61, and E109. Biochemistry **16:** 3760.

15. Lieberman, R., M. Potter, W. Humphrey, E. Mushinski, and M. Vrana. 1975. Multiple individual and cross-specific idiotypes on 13 levan-binding myeloma proteins of BALB/c mice. J. Exp. Med. **142:** 106.

16. Katz, A., W. Dreyer, and C. Antinsen. 1959. Peptide separation by two-dimensional chromatography and electrophoresis. J. Biol. Chem. **234:** 2897.

17. Ozols, J., and C. Gerard. 1977. Cleavage of tryptophanyl bonds in cytochrome b_5 by cyanogen bromide. J. Biol. Chem. **252,** 5986.

18. Hunkapiller, M., and L. Hood. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a non-protein carrier, and high pressure liquid chromatography. Biochemistry **17:** 2124.

19. Barstad, P., V. Farnsworth, M. Weigert, M. Cohn, and L. Hood. 1974. Mouse immunoglobulin heavy chains are coded by multiple germline variable region genes. Proc. Natl. Acad. Sci., U.S.A. **71**: 4096.

20. Kabat, E., T. Wu, and H. Bilofsky. 1979. Sequences of Immunoglobulin Chains. U.S. Dept. of Health, Education, and Welfare, Public Health Service, National Institutes of Health.

21. Konigsberg, W. 1972. Subtractive Edman degradation. In: Methods of Enzymology, Vol. XXV, Enzyme Structure Part B, pp. 326-332, Academic Press, New York.

22. Johnson, N., P. Gearhart, R. Douglas, and L. Hood. 1980. The amino acid sequences of nine V_H regions from hybridomas binding phosphorylcholine. Manuscript in preparation.

23. Hubert, J., N. Johnson, J. Greengard, B. Eaton, S. Rudikoff, and L. Hood. 1980. The complete V_H sequences of six myeloma proteins with specificity for phosphorylcholine. Manuscript in preparation.

24. Rudikoff, S., and M. Potter. 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. Proc. Natl. Acad. Sci., U.S.A. **73**: 2109.

25. Rao, D., S. Rudikoff, H. Krutzch, and M. Potter. 1979. Structural evidence for independent joining region gene in immunoglobulin heavy chains from anti-galactan myeloma proteins and its potential role in generating diversity in complementarity-determining regions. Proc. Natl. Acad. Sci., U.S.A. **76**: 2890.

26. Hood, L., M. Davis, P. Early, K. Calame, S. Kim, S. Crews, and H. Huang. 1980. Two types of DNA rearrangements in immunoglobulin genes. Cold Spring Harbor Symp. Quant. Biol. **45**, in press.

27. Cohn, M., B. Blomberg, W. Geckeler, W. Raschke, R. Riblet, and M. Weigert. 1974. First order considerations in analyzing the generation of diversity. In: The Immune System: Genes, Receptors, Signals (E. Sercarz et al., eds.), p. 89, Academic Press, New York.

28. Lieberman, R., M. Vrana, W. Humphrey, C. Chien, and M. Pptter. 1977. Idiotypes of inulin binding myeloma proteins localized to variable region light and heavy chains: genetic implications. J. Exp. Med. **146:** 1294.

29. Dayhoff, M. O. 1972. Atlas of Protein Sequence and Structure. The National Biomedical Research Foundation, Silver Spring, Maryland.

Summary

This dissertation is primarily a discussion of antibody diversity as seen in sets of immunoglobulin V regions with similar amino acid sequences. As a summary, I would like to describe my personal view of antibody diversity. This description will combine experimental observations with a large dose of speculation, and will be concluded with several ideas about future directions of research.

The diversity of IgM antibodies is less than that of IgG and IgA antibodies (1-3). Because of this, and because IgM antibodies precede IgG and IgA antibodies in ontogeny (4), I will discuss IgM diversity first. The experiments reported in this dissertation were designed to allow the analysis of patterns of antibody V region diversity. The kinds of V region diversity which are observed are necessarily restricted by the relatively small numbers of closely related V regions available for analysis (partial sequences from 29 phosphorylcholine-binding antibodies are included in Chapter 2).

An initial consideration in a discussion of antibody diversity is, what is the nature of the diversity in newly formed B cells in the bone marrow? I feel that this diversity has the following properties: (i) first, heavy chain-light chain combinatorial association leads to nearly complete mixing of V_L and V_H regions. (ii) In addition, gene segment joining mechanisms will lead to the assortment of all V_L gene segments with all J_L gene segments and all V_H gene segments with all D and J_H gene segments (5-8). Thus the immature cells arising from the bone marrow will express all possible V_L and J_L gene segment combinations with all possible V_H, D, and J_H gene segment combinations. (iii) Finally, somatic mutation other than the junctional diversity discussed below does not generate significant V gene variability in new B cells.

It has been estimated that the number of germline V_L gene segments may be around 200 (9), and V_H gene segments may number about the same. There seem to be four functional J_L gene segments (10, 11) and four J_H gene segments (7, 12),

and I will guess that the number of D gene segments is five. V_L-J_L junctions seem to give rise to six junctional sequences as a result of DNA joining mechanisms (8), but no estimate of comparable quality is possible for V_H-D-J_H junctions. In the phosphorylcholine-binding antibody system there are five V_H-D junctional sequences and perhaps five D-J_H junctional sequences which seem to arise as strictly "junctional" diversity (not as a result of D segment combinatorial joining) (1-3). Using the above speculations about gene segment combinatorial joining and numbers of gene segments, one can estimate the total germline antibody V_L region repertoire to be: 200 (V_L) x 6 (V_L-J_L junction) x 4 (J_L) = 4.8×10^3 V_L regions. Likewise, there would be 200 (V_H) x 5 (V_H-D junction) x 5 (D) x 5 (D-J_H junction) x 4 (J_H) = 10^5 V_H regions. Heavy-light chain combinatorial joining would then create 4.8×10^3 (V_L) x 10^5 (V_H) = 4.8×10^8 antibodies.

The B cells that leave the bone marrow circulate through the body and, after elimination or paralysis (13, 14) of self-reactive clones, mature into IgM-secreting cells (4). These IgM-secreting cells can then further differentiate to express different antibody constant regions (4, 15). What is the nature of the diversity in this pool of IgM producers? There are four important considerations. 1) A substantial amount of diversity is presumably eliminated because of self-reactivity, though it is impossible to guess how much. 2) During ontogeny, some clones will leave the bone marrow earlier than others. These clones will have a chance to undergo early clonal expansion and, in the absence of balancing influences, may retain a numerical advantage over later clones. 3) Antigen can selectively expand antigen-reactive clones. Cells leaving the bone marrow early will acquire an additional numerical advantage if they encounter antigen before other clones with similar specificity appear. 4) Antibody production has been postulated to be controlled by idiotype-antiidiotype networks (16-18). Antiidiotypic antibodies or T cells may routinely enhance or depress antibody responses. Thus I envision the IgM-producing B cell

pool as a collection of clones present at widely differing clonal frequencies. Clones which left the bone marrow early and were quickly triggered by antigen will be present at high frequencies, while unstimulated clones which arose relatively late will be present at very low frequencies (perhaps one or a few cells per mouse). If antibody V_L and V_H regions are expressed in a programmed order, cells producing specific heavy-light chain combinations may be present at similar frequencies in different individuals (19, 20).

IgA- and IgG-producing cells show greater V region diversity than IgM-producing cells (1). There are two major considerations in the generation of this increased IgG and IgA diversity. First, when and how is the additional variability in IgG and IgA antibodies created? And second, what is the mechanism by which this additional diversity reaches a relatively high frequency in IgG and IgA antibodies? These questions are discussed in Chapter 2, and so I will use this section to just emphasize those points which I feel are important. I feel that a great deal of the additional IgG V segment variability in the antiphosphorylcholine response is due to somatic mutation. Most importantly, I feel that the V segment variability arising repeatedly at V_H positions 40 and 53 in phosphorylcholine binding proteins (3) is due to somatic mutation. Thus somatic mutation seems to give rise to certain V_H segment variants in a reproducible manner. Additionally, I feel that somatic mutations are created initially in a single cell. Their detection in my study of phosphorylcholine-binding V_H regions suggests that such somatic mutants have been expanded, either clonally or as a population, to a level where they play a significant role in immunity. This expansion has presumably occurred in the face of a strong, germline response to phosphorylcholine. Therefore it is conceivable that even somatic mutants at low frequencies could expand in response to antigen at a fast enough rate to increase the effective available antibody diversity.

The kinds of selection which may operate to generate IgG and IgA diversity

are also discussed in Chapter 2 (1). I will add that I feel that both antigen binding and antiidiotypic network regulation of antibody are important selective forces. Selection by antigen is probably on the basis of antibody affinity, though this assumption is difficult to test experimentally. Antibody affinity measurements performed with free antibody and hapten in vitro may bear little relationship to the affinity of cell surface antibodies for hapten-protein antigens in vivo. Artificially induced selection by antiidiotypic antibodies is well documented in experimental animals (17, 18, 21). A particularly simple selection model postulates that antiidioype inhibits the growth of high frequency clones. Thus infrequent clones would always be at an advantage, tending to maintain antibody heterogeneity. A high level of antibody heterogeneity would in turn give an organism versatility in responding to new antigens. IgG and IgA V regions would be more diverse than IgM V regions either because they have undergone more generations of selection, or because idiotypic regulation is specific for IgG and IgA producing cells.

This, then, is an overview of my thoughts on antibody diversity. In the following section I will deal specifically with each major category of antibody diversity generation mechanisms.

The germline V segment contribution to diversity.

I feel that the germline contribution to antibody diversity perhaps plays the most critical role of all antibody diversity generation mechanisms. Using the numbers mentioned previously, loss of all but one V_L or V_H gene segment would reduce the antibody diversity by a factor of 200. V segments can be divided into groups differing by no more than three residues over their N-terminal 30 amino acids (22). These groups may differ by over 50% among each other. If there are approximately 30 V_L groups and 30 V_H groups (9), then there are only 900 V group sets of germline antibody molecules. The differences among these 900 V group sets will be much greater than those within a set, and the inter-V group set differences will extend

throughout the V segment. Thus germline V segment diversity and the diversity arising from somatic gene segment joining are not equivalent. An increase in D segment diversity probably would not compensate for a loss of V segment diversity.

V_L and V_H third hypervariable region diversity.

V_H-D-J_H and V_L-J_L junctional diversity and combinatorial joining creates a considerable amount of heavy and light chain third hypervariable region variability (2, 3, 5-8, 11). In fact, there is probably well over ten times the number of V_L plus V_H third hypervariable regions than all other hypervariable regions combined. Without V_L-J_L and V_H-D-J_H diversity, the total antibody diversity would be reduced by a factor of 4 (J_L) x 6 (V_L-J_L) x 5 (V_H-D) x 5 (D) x 5 (D-J_H) x 4 (J_H) = 1.2×10^4.

Combinatorial association.

Combinatorial association of heavy and light chains is a particularly powerful mechanism for diversity generation. 10^5 V_H regions and 4.8×10^3 V_L regions (see above) can generate at most 10^5 antibodies without combinatorial association, a reduction by a factor of 4.8×10^3 of the number of antibodies otherwise possible.

Somatic mutation.

The most difficult factor to assess is somatic mutation. In Chapter 3, it is suggested that certain patterns of variation arise repeatedly in the IgG and IgA pool (3). If these patterns of variation arise as a result of somatic mutation, then somatic mutation may well be quite reproducible. In addition, it has been suggested earlier in this summary that low level somatic variants may be effective at expanding the usable antibody diversity. Thus, organisms may be able to rely on generating a certain amount of functional antibody diversity by somatic mutation, eliminating the need to retain such diversity in the germline. Special mutational processes might well evolve to enhance the creation of such somatic diversity (23, 24).

In the above discussion I have emphasized the factors by which each mechanism

of diversity generation increases the antibody repertoire. In the cases of germline segment diversity, gene segment combinatorial joining and junctional diversity, and heavy-light chain combinatorial association, I feel that the factor by which each mechanism increases diversity is too large to be an evolutionary artifact. These mechanisms must make important contributions to immunity.

Somatic mutation, on the other hand, may increase immunity by a much smaller factor. Somatic mutations may only be present in clones which are repeatedly stimulated by antigen and hence undergo many rounds of division and clonal expansion and reduction. Thus somatic mutation is possibly the least important of all mechanisms of antibody diversity generation, even though it may be quite visible in the antibodies binding certain antigens.

In the next section I will briefly discuss experimental approaches which address some of the points raised in this summary.

Studies at the DNA level will eventually allow reasonable estimates of the number of germline V segments to be made. These estimates will permit a fairly accurate assessment of the germline V gene segment contribution to antibody diversity. Isolation and sequencing of germline D gene segments will probably also occur in the near future. It is important, however, to determine the extent to which D region combinatorial joining and D region junctional diversity actually occur during differentiation. This question could be investigated in a fairly straightforward way at the subgroup level using cDNA library of splenic H chain mRNA. This library could be screened with an appropriate probe and the resultant clones analyzed by restriction digests and DNA sequencing. By screening with a V_H segment probe only (no D segment) a broad collection of V_H-D-J_H cDNA combinations should be obtained allowing one to determine the extent of the D diversity associated with V gene segments from a particular group.

In addition, such a splenic heavy chain cDNA library could be used for an

extensive study of D segment variability. DNA complementary to C region sequences at the V_H-C_H junction could be used as a primer for DNA sequencing studies of D segments from cloned cDNAs.

Combinatorial association of light and heavy chains is an old idea which has never been satisfactorily investigated. λI light chains are encoded by only one or a few $V_{\lambda I}$ genes (25, 26) and presumably undergo combinatorial association to the same degree as κ light chains. Thus determination of the V_H sequences associating with λI chains would provide a measure of heavy-light chain combinatorial association. Partial V_H cDNA sequences from λI-producing hybridomas could be determined using the dideoxy nucleotide DNA sequencing technique and primers complementary to immunoglobulin C regions (7). This experiment would provide information on D segment variability as well as on combinatorial association.

A final question, and one which will be very difficult to answer, concerns the importance of each component of antibody diversity generation to immunity. The ideal experimental model would be a mouse lacking particular diversity mechanisms, e.g., mice with only a few V_L or V_H segments, one J_L or J_H segment, or one D segment. Unfortunately, no such mice have ever been found. The new awareness of the roles of D and J segments, however, may eventually lead to the discovery of their involvement in immunodeficiencies. Until then, the value of different mechanisms for generating antibody variability will continue to be judged on the basis of calculations such as those presented earlier. It is hard to imagine a way (short of stumbling across fortuitous mutants) of ever answering the fundamental question, "What are the contributions of the various mechanisms of antibody diversity generation to immunity?"

References

1. Gearhart, P., N. Johnson, R. Douglas, and L. Hood. (1980). IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. Manuscript in preparation.
2. Hubert, J., N. Johnson, J. Greengard, B. Eaton, and L. Hood. (1980). The complete amino acid sequences of six V_H regions from myeloma proteins binding phosphorylcholine. Manuscript in preparation.
3. Johnson, N., P. Gearhart, R. Douglas, and L. Hood. (1980). The amino acid sequence of nine V_H regions from hybridomas binding phosphorylcholine. Manuscript in preparation.
4. Cooper, M., J. Kearney, P. Lydyard, C. Grossi, and A. Lawton. (1976). Studies of generation of B-cell diversity in mouse, man, and chicken. Cold Spring Harbor Symp. Quant. Biol. **41**, 139.
5. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. (1978). Rearrangement of genetic information may produce immunoglobulin diversity. Nature (Lond.) **276**, 785.
6. 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 (Lond.) **283**, 35.
7. 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.
8. 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.
9. Seidman, J., A. Leder, M. Nau, B. Norman, and P. Leder. (1978). Antibody diversity. Science **202**, 11.

10. Max, E., J. Seidman, and P. Leder. (1979). Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. Proc. Nat. Acad. Sci., USA **76**, 3450.

11. Sakano, H., K. Hüppi, G. Heinrich, and S. Tonegawa. (1979). Sequences at the somatic recombination sites of immunoglobulin light chain genes. Nature **280**, 288.

12. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature **286**, 676.

13. Nossal, G., and B. Pike. (1975). Evidence for the clonal abortion theory of B-lymphocyte tolerance. J. Exp. Med. **141**, 904.

14. Parks, D., and W. Weigle. Regulation of B-cell unresponsiveness by suppressor cells. Immunol. Rev. **43**, 217.

15. Pernis, B., L. Forni, and A. Luzzati. (1976). Synthesis of multiple immunoglobulin classes by single lymphocytes. Cold Spring Harbor Symp. Quant. Biol. **41**, 175.

16. Jerne, N. (1973). The immune system. Scientific American **229**, 52.

17. Pierce, S., and N. Klinman. (1977). Antibody-specific immunoregulation. J. Exp. Med. **146**, 509.

18. Nisonoff, A., S.-T. Ju, and F. Owen. (1977). Studies of structure and immunosuppression of a cross-reactive idiotype in Strain A mice. Transplantation Reviews **34**, 89.

19. Klinman, N., and J. Press. (1975). The characterization of the B-cell repertoire specific for the 2,4-dinitrophenyl determinant and 2,4,6-trinitrophenyl determinants in neonatal Balb/c mice. J. Exp. Med. **141**, 1133.

20. Sigal, N., P. Gearhart, J. Press, and N. Klinman. (1976). Late acquisition of a germline antibody specificity. Nature **259**, 51.

21. Cosenza, H., and H. Köhler. (1972). Specific inhibition of plaque formation to phosphorylcholine by antibody against antibody. Science **176**, 1027.

22. Potter, M. (1977). Antigen-binding myeloma proteins of mice. Adv. Immunol. **25**, 141.

23. Baltimore, D. (1974). Is terminal deoxynucleotide transferase a somatic mutagen in lymphocytes? Nature **248**, 409.

24. Jerne, N. (1971). The somatic generation of immune recognition. Eur. J. Immunol. **1**, 1.

25. Tonegawa, S. (1976). Reiteration of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. Proc. Nat. Acad. Sci. USA **73**, 203.

26. Honjo, T., S. Packman, D. Swan, and P. Leder. (1976). Quantitation of constant and variable region genes for mouse immunoglobulin λ chains. Biochemistry **15**, 2780.

Appendix

Analysis of phenylthiohydantoin amino acids by high performance liquid chromatography on DuPont Zorbax cyanopropylsilane columns

This manuscript was published in Analytical Biochemistry

ANALYTICAL BIOCHEMISTRY 100, 335–338 (1979)

Analysis of Phenylthiohydantoin Amino Acids by High-Performance Liquid Chromatography on DuPont Zorbax Cyanopropylsilane Columns[1]

N. D. JOHNSON, M. W. HUNKAPILLER, AND L. E. HOOD

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

Received March 26, 1979

The phenylthiohydantoins (Pth) of the common amino acids can be resolved in a single analysis using a 25 × 0.46-cm DuPont Zorbax cyanopropylsilane (CN) column developed with a gradient of methanol/acetonitrile (17:3) in sodium acetate buffer, pH 5.4. The Zorbax CN columns exhibit greater durability, reproducibility, and sensitivity than do columns with an octadecylsilane (C_{18}) support when used for Pth amino acid analysis in automated polypeptide sequencing.

In the last few years, reverse-phase high-performance liquid chromatography (hplc)[2] has revolutionized protein sequencing analytical procedures (1–4). It provides rapid, sensitive, quantitative analysis of the Pth amino acids formed from the amino acids released during automated Edman degradation. Unlike the other techniques—thin-layer chromatography (5), gas chromatography (6), and back-hydrolysis to the parent amino acids (7,8)—commonly used for this analysis, hplc can resolve all the commonly encountered Pth derivatives in a single chromatographic run. In this report, we describe an hplc system for the Pth amino acid analysis using DuPont Zorbax CN columns rather than the more commonly used C_{18} columns. It provides greater column lifetime, higher sensitivity, increased column-to-column reproducibility, and more reliable resolution of Pth-Arg and Pth-His than previously reported hplc analyses of these derivatives.

[1] Supported by NIH Grant GM 06965.

[2] Abbreviations used: hplc, high-performance liquid chromatography; Pth, phenylthiohydantoin; CN, cyanopropylsilane; ODS and C_{18}, octadecylsilane.

MATERIALS AND METHODS

Reagents. Acetonitrile and methanol were Burdick and Jackson distilled-in-glass solvents. Sodium hydroxide and acetic acid were Baker analyzed reagents. Water was first deionized and then distilled. Pth amino acid standards were obtained from Pierce Chemical Company.

Chromatography equipment. The hplc unit consists of two Waters Associates Model 6000A solvent delivery systems, a Waters model 710 autosampler, a Waters Model 440 dual-channel (254 and 313 nm) absorbance detector, a Houston Instruments Omniscribe dual-channel chart recorder, a Caltech-constructed multilinear programmable gradient maker that controls the two Waters pumps, and a simple constant temperature oven. The oven houses a DuPont Permaphase ETH guard column (5 × 0.46 cm) and a DuPont Zorbax CN analytical column (25 × 0.46 cm) at 31 ± 1°C. The factory-installed sample loading loop of the Waters autosampler has been replaced with a 25-μl loop constructed of 0.0625-in.-o.d. × 0.009-in.-i.d. 316 stainless-steel tubing, and the line between the pump and the autosampler has been replaced

0003-2697/79/180335-04$02.00/0

with 0.0625-in.-o.d. × 0.020-in.-i.d. 316 stainless-steel tubing.

Pth amino acid analysis. The columns are equilibrated initially with pH 5.4 sodium acetate buffer (0.015–0.040 M) at a flow rate of 1.0 ml/min. The samples are dissolved in acetonitrile, loaded into the low-volume sample inserts for the autosampler, and injected into the flow to the column. The Pth amino acids are eluted with an increasing concentration of a methanol/acetonitrile (17:3) mixture with the total buffer plus solvent flow rate held constant at 1.0 ml/min. The acetate buffer is made up fresh daily by diluting a 0.10 M stock solution and degassing it by boiling for 5 min.

RESULTS AND DISCUSSION

Pth Amino Acid Fractionation

Figure 1 shows a typical elution profile of a mixture of 19 Pth amino acids from a Zorbax CN column. All 19 of the derivatives can be resolved (>90%) with a 35-min column recycle time. In addition, several other Pth amino acids can be resolved with the same gradient. These include the Pth derivatives of carboxymethylcysteine (elutes 20 s before Pth-Glu), succinyllysine (elutes 60 s before Pth-Asn), and carboxamidomethylcysteine (elutes 30 s after Pth-Glu).

Although Pth-Asn, -Ser, and -Thr elute within a 45-s period, they are >90% resolved because the peaks are very sharp. We have never experienced difficulty from their similar elution times in assigning a sequencer residue to these amino acids because (a) serine and threonine residues give multiple peaks in addition to Pth-Ser and -Thr and asparagine residues give some Pth-Asp in addition to Pth-Asn, and (b) there is always a background of most Pth amino acids in a sequencer sample that allows assignment of any large peak (such as Pth-Thr) relative to the background of smaller ones (the nearby Pth-Asn, -Ser, -Gln, and -Gly). Although the absolute retention times do vary by as much as 20 s from one run to another, the spacing between nearby peaks is nearly constant. Moreover, one can always cochromatograph a small amount of the authentic Pth derivative with the sample to confirm identification should there be no observable background.

This resolution is in several respects superior to that attainable with DuPont Zorbax ODS columns, the only other columns with which single-step resolution of all common Pth amino acids has been reported (1,4).

(i) The resolution of Pth-His and Pth-Arg from the other Pth amino acids is more easily maintained with the CN than with the ODS column. With the latter, Pth-His and Pth-Arg often elute as broad peaks that overlap with those of other Pth amino acids. The peak broadening increases rapidly with column use, and although the effects of column aging can be compensated for somewhat by increasing the starting buffer concentration, these derivatives frequently are resolved only by using a much steeper eluting gradient. With the CN columns, they elute as sharp peaks throughout the usable lifetime of the columns. Although they tend to elute later at a given buffer concentration as the columns age, this effect can be compensated for by increasing the buffer concentration from ~0.015 M for a new column to ~0.040 M for a column that has been used steadily for 2–3 months.

(ii) The usable lifetime (3 months with 2500–3000 injections) of the CN columns is considerably longer than that of the ODS columns (3–6 weeks with 750–1500 injections) if resolution similar to that shown in Fig. 1 is required. Although we have not examined the cause of this difference systematically, we feel it is most likely due to the high column temperature (55–60°C) required for comparable resolution of all the Pth amino acids on the ODS columns.

(iii) The sensitivity of the Zorbax CN columns as an analytical tool for quantitative analysis of Pth amino acid mixtures in automated protein sequencing experiments is superior to that of C_{18} columns that have

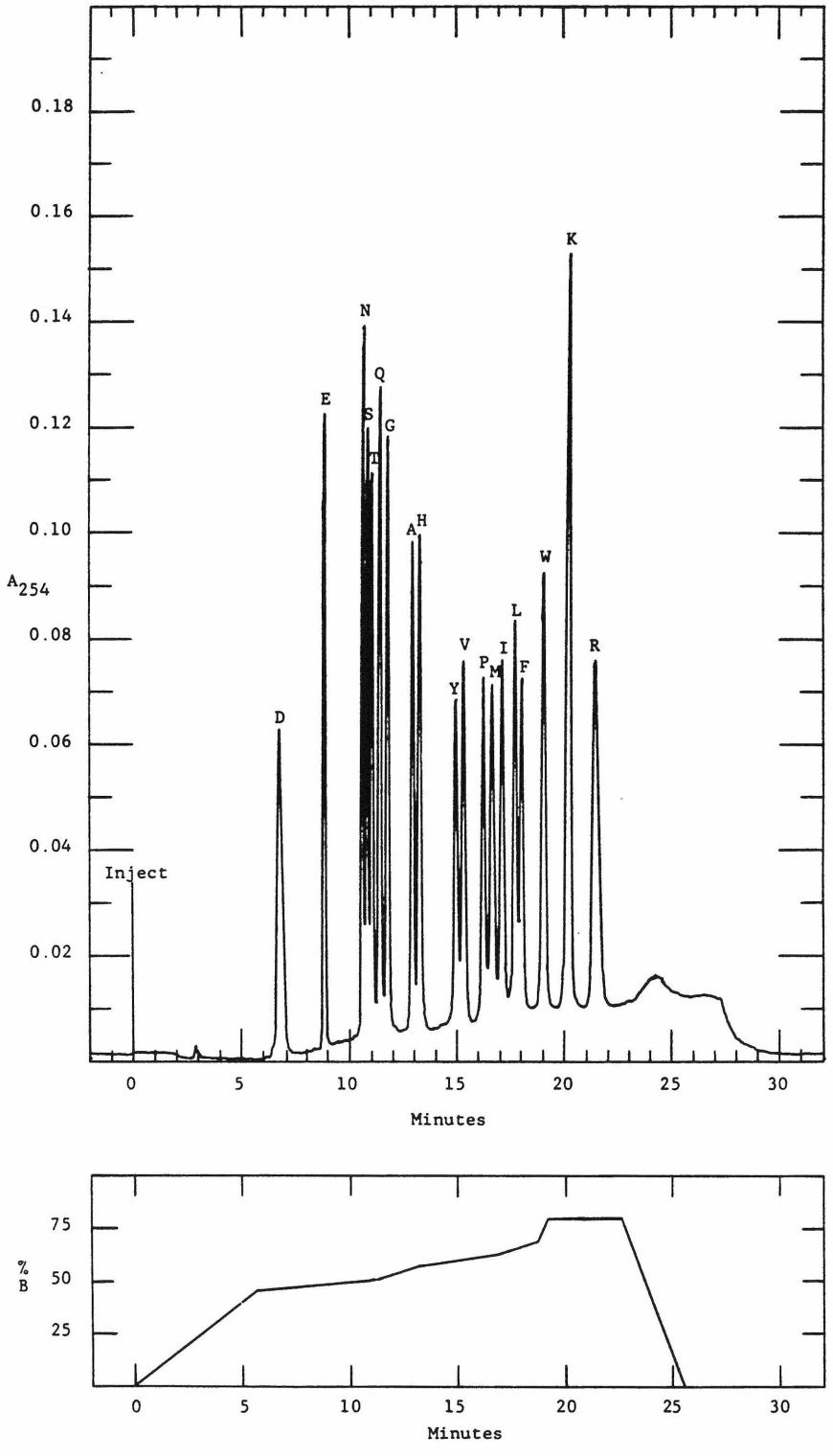


FIG. 1. High-performance liquid chromatographic separation of Pth amino acids on DuPont Zorbax CN. The initial solvent (A) was 0.024 M sodium acetate buffer, pH 5.4, and the secondary solvent (B) was methanol/acetonitrile (17:3). The sample consisted of 10 μl of acetonitrile containing 1 nmol of each of the Pth amino acids. Flow rate was 1.0 ml/min, and column temperature was 31°C.

been reported (1–4). The maximum peak absorbance per nanomole ranges from 0.06–0.15 A_{254}^{1cm} for the Zorbax CN columns compared to 0.03–0.09 A_{254}^{1cm} for C_{18} columns. This is due to the higher theoretical plate count of the CN columns and the gradient required for the Pth analysis.

(iv) The Zorbax CN columns give reproducible elution patterns for the Pth amino acids from one column to another. We have not had to change the gradient, solvent, buffer composition (other than concentration for Pth-His and Pth-Arg resolution), flow rate, or column temperature to achieve the resolution shown in Fig. 1 for the CN columns we have used over a period of 18 months. Nonreproducibility of the column-to-column performance of Zorbax ODS for Pth-His and Pth-Arg often forced us to devise separate gradients solely for their determination (4).

The chief disadvantage of using the Zorbax CN rather than the C_{18} columns is the higher baseline absorbance rise during the gradient development due to impurities in most batches of hplc-grade methanol. Investigators who need to use the higher sensitivity photometer scales should either ask their supplier for lots with an ultraviolet wavelength cutoff of 204 nm (or less) or should repurify the methanol before use. Treatment of hplc-grade methanol by boiling for 10 min with activated charcoal, filtration, and distillation through a 40-cm Widmer column under argon has proven effective in our laboratory. With the 254-nm photometer in our hplc system, the gradient baseline rise can be as low as 0.002 with the redistilled methanol compared to 0.005–0.020 for various batches of Burdick and Jackson methanol. The acetonitrile–buffer system used for Zorbax ODS columns (1,4) or the methanol–buffer systems used for Waters μBondapak C_{18} columns (2,3) typically show an absorbance rise of 0.005. Addition of acetone to the initial buffer to raise its 254-nm adsorbance to the level of the eluting solvent (and hence minimize baseline rise during the gradient), although successful with the μ-Bondapak C_{18} columns (2,3), is impractical with the Zorbax CN columns because of retention of acetone on the column during the initial equilibration with buffer and subsequent elution of the adsorbed acetone during the gradient.

REFERENCES

1. Zimmerman, C. L., Appella, C., and Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573.
2. Zeeuws, R., and Strosberg, A. D. (1975) *FEBS Lett.* **85**, 68–72.
3. Brown, A. S., Mole, J. E., Weissinger, A., and Bennett, J. C. (1978) *J. Chromatogr.* **148**, 532–535.
4. Hunkapiller, M. W., and Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133.
5. Edman, P., and Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91.
6. Pisano, J. J., Bronzert, T. J., and Brewer, H. B., Jr. (1972) *Anal. Biochem.* **45**, 43–49.
7. Smithies, O., Gibson, D. M., Fanning, E.M., Goodfleish, R. M., Gilman, J. G., and Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921.
8. Mendez, E., and Lai, C. Y. (1975) *Anal Biochem.* **68**, 47–53.