AMINO ACID SEQUENCE DIVERSITY WITHIN THE FAMILY OF ANTIBODIES BEARING THE MAJOR ANTIARSONATE CROSS-REACTIVE IDIOTYPE OF THE A STRAIN MOUSE*

BY CLIVE A. SLAUGHTER AND J. DONALD CAPRA

From the Department of Microbiology, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235

An uneasy consensus of opinion is being reached concerning the genetic basis of antibody diversity (1). This consensus, which includes elements of both the original "germline" and "somatic diversification" models, is being built in large part through the characterization of antibodies elicited by simple haptens in mice. The response to *p*-azophenylarsonate (Ars)[1] has been particularly well studied and illustrates the general features of the consensus model. It also poses, and helps to illuminate, some of the outstanding issues of the diversity problem.

When coupled to the protein keyhole limpet hemocyanin, Ars stimulates a vigorous antihapten response in all mouse strains. Monoclonal murine antibodies specific for Ars, isolated by the hybridoma technique (2, 3) may be classified into three families (4). These families, designated Ars-A, Ars-B, and Ars-C, are defined by the N-terminal amino acid sequences of their heavy chain variable regions (V_H regions),[2] which belong to different subgroups. The families delineated in this way are further distinguished both by serological characteristics and by the use of light chain variable regions (V_L regions) of different subgroups. The three families are believed to be encoded by different sets of germline genes, and the ability of a mouse strain to recruit a particular family into its response to Ars is an inherited characteristic. This situation exemplifies the frequently observed expression of multiple sets of germline genes in response to a single antigenic determinant.

Inbred strains may vary genetically in the spectrum of antibodies that are elicited by hapten. For example, A strain mice respond to Ars with antibodies of

* Supported by grant AI-12127 and fellowship I F32 GM07710 from the National Institutes of Health, and grant I-874 from the Robert A. Welch Foundation.

[1] *Abbreviations used in this paper:* Ars, *p*-azophenylarsonate; Ars-A, family of antibodies which bears the major antiarsonate cross-reactive idiotype of the A strain mouse; CDR, complementarity-determining region; CNBr, cyanogen bromide; CRI, cross-reactive idiotype; Dex, α-1,3-dextran; HPLC, high performance liquid chromatography; NP, (4-hydroxy-3-nitrophenyl) acetyl; PC, phosphoryl choline; V_H, heavy chain variable; V_κ, kappa chain variable; V_L, light chain variable.

[2] "V_H region" here refers to amino acid residues 1–121 of the heavy chain, which are encoded by three separate genes, V_H, D_H, and J_H. "V_H segment", in contrast, is used to denote amino acid residues 1–98 of the heavy chain, which are encoded by the V_H gene alone. Similarly, "V_L region" refers to amino acid residues 1–108 of the light chain, which are encoded by two separate genes, V_κ and J_κ (or V_λ and J_λ). "V_κ segment", in contrast, is used for amino acid residues 1–95, which are encoded by the V_κ gene alone.

which a major portion bears a cross-reactive idiotype (CRI) peculiar to that strain (5). The CRI is conventionally identified using a polyclonal rabbit antiserum that is raised against A/J mouse anti-Ars antibodies and rendered idiotypically specific by absorption on preimmune A/J immunoglobulin (5). The CRI is measured in a radioimmunoassay that determines the degree of inhibition between the rabbit anti-CRI and radiolabeled A/J anti-Ars antibodies (5). Breeding studies between the CRI-positive A/J mouse strain and CRI-negative strains such as BALB/c show that expression of the CRI in the anti-Ars response is inherited as if it were determined by a single autosomal dominant gene at a locus linked to the *Igh* gene complex (6). Linkage to *Igh* is a feature common to many such idiotypic systems (7). The expression of one of the three structurally defined anti-Ars families, Ars-A, shows the same strain distribution as the CRI, and so far all A/J antibodies bearing the CRI have been found to belong to this family (4). Allelic diversity of this general kind is commonly observed in the expression of antibody families and of shared idiotypic specificities.

The V_L regions of the Ars-A family have been characterized in detail. Eight complete amino acid sequences are available (8–10), seven from A/J and one a CRI-positive hybridoma product derived from the congenic strain C.AL 20, in which the heavy chain genes are contributed by an A strain genome and the light chain genes are from BALB/c. All the chains are closely similar in sequence and show <3% overall variation. They all appear to be the products of a single A/J germline V_κ gene or, in the case of C.AL 20, the BALB/c homolog of this gene. They probably also use a single germline J_κ gene, identified as the homolog of either $J_\kappa 1$ or $J_\kappa 2$ in BALB/c (11, 12). Thus, the BALB/c genome can contribute V_κ chains essentially identical to those of the A/J to make a CRI-positive antibody in response to Ars, provided that an appropriate V_H chain is available from the A strain genome. A light chain virtually identical to those of the Ars-A family has recently been reported (13) in an antibody elicited in BALB/c in response to the hapten 2-phenyl oxazolone. A similar chain has also been found in a BALB/c myeloma protein (MOPC-173) of unknown antigen specificity (14). Both proteins have heavy chains that differ from each other and both are unlike those of the Ars-A family. This is an example of the long-predicted production of novel antigen specificities by combinatorial association of V_L and V_H segment genes to form novel variable region pairs.

Although members of the Ars-A antibody family are similar in many respects, several lines of evidence suggest that structural differences exist between them. First, they differ in the degree to which they display the CRI in the conventional radioimmunoassay (15–17). Some are strongly CRI positive, some are only weakly positive, and some are CRI negative. Chain recombination experiments, using light and heavy chains from antibodies within the family that differ in their expression of the CRI, show that the heavy chain is largely responsible for the expression or nonexpression of the CRI and for the degree of its expression (18). Members of the family also differ in their expression of "private" or "individual" idiotypes (16), in their affinity for Ars (19, 20), their fine specificity (19), and in the N-terminal sequences of their heavy and light chains (15, 21, 22). Variation in such properties within antibody families is a general phenomenon and is sometimes very extensive. It has been demonstrated in structural

studies of antibodies against dextran (23), (4-hydroxy-3-nitrophenyl) acetyl (NP) (24), phosphoryl choline (PC) (25), oxazolone (13, 26), and the synthetic polymers GAT (copolymer of glutamic acid, alanine, and tyrosine) and GA (copolymer of glutamic acid and alanine) (27), and has also been characterized in families of closely related myeloma proteins sharing specificity for antigen, including galactan-binding myeloma proteins (28) and inulin-binding myeloma proteins (29, 30).

The genetic basis for such variation is the subject of intense interest. As part of the analysis of the variation in the arsonate system, Southern filter hybridization studies have been carried out using V_H segment DNA probes from Ars-A hybridomas to analyze the germline DNA of A/J mice. These studies have revealed several V_H genes closely related to members of the Ars-A family (31, 32). However, use of a J_H probe to detect the rearranged gene in hybridoma DNA has shown that productive rearrangements give rise to restriction fragments of the same size in most hybridomas making Ars-A antibodies (33, 34). This suggests that V_H DNA from only one of the restriction fragments observed in digests of genomic DNA is used to make antibodies of this family and that somatic mutation plays a determining role in the diversification of the V_H segment sequences.

However, the extent of sequence diversity in the complete V_H region, and the distribution of replacements between and within V_H, D_H, and J_H segments have not previously been determined for the Ars-A family. These issues are the subjects of the present report. The amino acid sequences of five V_H regions belonging to antibodies from the Ars-A family are described. The sequences of the corresponding V_κ regions have already been determined (8). The antibodies chosen for analysis were selected to represent the range of quantitative expression of the CRI seen among members of the family, extending from the most strongly CRI-positive HP 93G7 and HP R16.7, through the progressively more weakly CRI-positive HP 123E6 and HP 124E1, to the CRI-negative HP 91A3. These antibodies also differ in their expression of private idiotypes (16), in their reactivity with monoclonal antiidiotypic antibodies (35), and in their fine specificity for antigen (19), but show similar affinities for the immunizing hapten Ars (19). The present study was thus designed to develop structural correlates of the CRI and to illuminate the genetic mechanisms that give rise to sequence variability within the Ars-A family. The sequence of the prototype molecule for the series, HP 93G7, has already been published (31), and preliminary reports of the other sequences have also appeared (10, 36, 37). The data in this paper supercede all previous listings for these sequences.

The data show that the V_H regions are closely related, but the differences between them are sufficiently numerous to permit discrimination among all of them. The structural variation in the heavy chains together with the variation in the corresponding light chains (8) provide a basis for interpreting the differences in the serological and other properties that the five antibodies display.

Materials and Methods

Preparation and Characterization of Hybridomas. The five hybridomas secreting anti-Ars antibodies were prepared by fusion of spleen cells from A/J mice immunized with keyhole

limpet hemocyanin-Ars with the nonsecreting hybrid plasmacytoma cell line Sp2/0-Ag 14 (38). The generation of the hybridomas, which all produce antibodies that are IgG1, kappa, has been described in detail previously (15, 16). HP 93G7 and HP 91A3 were derived in one fusion, but since the spleens of two different mice were used, it is not known whether they arose from the same animal or different animals. HP 123E6 and HP 124E1 arose from a second fusion, and again, the spleens of two different mice were pooled. HP R16.7 was derived in a third fusion and was the gift of Dr. A. Nisonoff, Brandeis University, Waltham, MA. The hybridomas were injected into CAF_1 mice for the collection of ascites fluid containing large amounts of hybridoma antibody (2–5 mg/ml). In each case, the hybridoma product was isolated from the ascites fluid by affinity chromatography on a column of Sepharose 4B conjugated with human gamma globulin that had been derivatized with arsanilic acid (15). Pertinent data on the serological characteristics of the five antibodies are summarized in Table I.

Preparation and Sequencing of Cyanogen Bromide (CNBr) Peptides. Intact hybridoma immunoglobulins were cleaved with CNBr in 70% formic acid (39). The heavy chain fragment that consisted of the disulfide-bonded pair of peptides extending from position 21 to 81 and 82 to 142, respectively, was isolated by gel filtration on a Sephadex G100 column equilibrated with 5 M guanidine-HCl. Incomplete cleavage at Met-20, however, produced a copurifying fragment in which the 21–81 peptide was lengthened to comprise the complete N-terminal sequence 1–81. After desalting on a Sephadex G25M column equilibrated with 0.25 M ammonium hydroxide, and lyophilizing, the constituent peptides were dissociated by complete reduction and alkylation (40) and were desalted and lyophilized as before. For HP 93G7, HP R16.7, and HP 124E1, the peptide 82–142 was separated from 21–81 and 1–81 by cation-exchange chromatography on a sulphopropyl-Sephadex column equilibrated with 8 M urea and 0.001 M KCl at pH 3.4. Elution was performed with a gradient of 0.001–0.2 M KCl. All peptides were desalted after ion-exchange chromatography by dialysis against 0.02% trifluoroacetic acid in water using low molecular weight cut-off dialysis tubing, and were then lyophilized. The peptides 21–

TABLE I

Inhibition of Binding in Various Assay Systems by Purified Hybridoma Products

Unlabeled inhibitor	Nanograms required for 50% inhibition		
	A. Rabbit anti-CRI/CRI*	B. Rabbit anti-Id 123E6/93G7*	C. Rabbit anti-Id 91A3/93G7*
A/J anti-arsonate	15	200	25
HP 93G7	12	32	12
HP R16.7	9	Not done	10
HP 123E6	50	87	86
HP 124E1	2,900	48	16
HP 91A3	>20,000	>20,000	25
Arsonate	10 mM	2 mM	>100 mM

Assay A. Canonical assay for CRI, showing diversity of degree of CRI expression (15) and high sensitivity to hapten (17, 62). Antibodies capable of inhibiting by 50% (all except HP 91A3) are designated CRI-positive. Assay B. Heterologous assay for idiotypic specificities shared by HP 93G7 and HP 123E6, showing commonality among CRI-positive antibodies and high sensitivity to hapten inhibition (62). Assay C. Heterologous assay for idiotypic specificities shared by HP 93G7 and HP 91A3, showing commonality among all Ars-A antibodies (4) and relative insensitivity to hapten inhibition (E. Milner, personal communication).

* Radiolabeled. Assays used 10 ng ^{125}I-labeled tracer antibody and slightly less than an equivalent amount of antiidiotype.

81 and 1–81 were subsequently separated by gel filtration on a Sephadex G75 SF column equilibrated with 0.25 M ammonium hydroxide. In the case of HP 123E6, the peptide 82–142 could be only partially separated from 1–81 and 21–81 by cation-exchange chromatography, and required further fragmentation to produce homogeneous peptides for sequencing (see below). HP 91A3 had a Met at position 105 that was not present in the other antibodies. During CNBr cleavage, the 82–142 peptide was therefore split into two fragments. The 82–105 fragment remained disulphide bonded to the peptides 21–81 and 1–81 and after reduction and alkylation was successfully purified by the same procedure as was used for HP 93G7, HP R16.7, and HP 124E1. The 106–142 fragment, however, was recovered separately during the initial G100 gel filtration step and was subsequently isolated from co-chromatographing heavy chain constant region peptides by reverse-phase, high performance liquid chromatography (HPLC) (for conditions, see below).

Peptides extending from position 82 were sequenced to position 121, the end of the V_H region, except for the peptide 82–105 from HP 91A3, which was sequenced to its C-terminus. Peptides extending from position 21 could usually be sequenced only as far as position 70. Residues between positions 70 and 81 were identified after tryptic digestion (see below). Residues at positions 1–40 were identified by sequence analysis of intact heavy chains isolated by gel filtration on a G100 column equilibrated with 5 M guanidine-HCl after complete reduction and alkylation of the parent immunoglobulin.

Preparation and Sequencing of Tryptic Peptides. CNBr peptides 21–81 or 1–81 were digested with TPCK-Trypsin (Worthington Biochemical Corp., Freehold, NJ). Fragments were isolated by gel filtration on a Sephadex G50 SF column equilibrated with 0.25 M ammonium hydroxide, and by reverse-phase HPLC using an RP300 column (Brownlee Labs Inc., Santa Clara, CA). For the HPLC separation, initial conditions included an aqueous solution of 0.02% trifluoroacetic acid in water and elution was performed with a gradient of 0–40% of a mixture containing 70% acetonitrile and 30% isopropanol acidified with 0.02% trifluoroacetic acid.

The mixture of CNBr peptides 21–81 and 82–142 from HP 123E6, which was obtained during cation-exchange chromatography, was further processed to produce pure peptides for sequencing by citraconylating the mixture (40) and digesting with trypsin. This procedure fragmented the peptide 82–142 by cleavage at arginines in positions 84, 98, and 101, but left the peptide 21–81 intact. Separation of the various components of the resulting mixture was effected by gel filtration on a column of Sephadex G50 SF equilibrated with 0.25 M ammonium hydroxide. It was possible to obtain complete sequences of all tryptic peptides encompassing regions of the molecules not already sequenced using unfragmented CNBr peptides.

Sequencing Methodology. Automated amino acid sequence analysis was performed using Beckman 890C or 890D amino acid sequencers (Beckman Instruments, Inc., Palo Alto, CA). 0.25 M or 0.1 M Quadrol programs were used in conjunction with the nonprotein carrier Polybrene (41). The phenylthiohydantoin amino acid derivatives were identified by HPLC using both a Waters RCM C18 and RCM CN column (Waters Associates Inc., Milford, MA) for each determination, and by amino acid analysis after back-hydrolysis in selected cases. Quantities of peptides sequenced in individual runs varied from 200 pmol to 10 nmol and repetitive yields measured 92–95%.

Results and Discussion

The V_H region amino acid sequences of five A/J anti-Ars antibodies of the Ars-A family, which differ in the degree to which they express the CRI, are shown in Fig. 1. The sequences are complete except in two respects. First, uncertainties remain between positions 55 and 63 in HP 123E6. All analyses in this paper exclude from consideration this region of HP 123E6. Second, it has proved impossible to identify the residue at position 55 in HP 93G7 and HP 124E1 using conventional techniques (see Materials and Methods). In both cases,

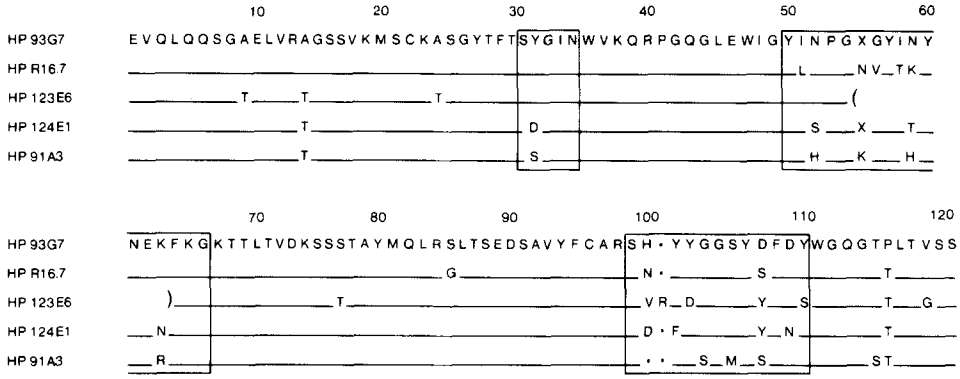

FIGURE 1. Antiarsonate hybridoma heavy chains differing with respect to their expression of the CRI. Comparison of amino acid sequences of heavy chain V regions with that of HP 93G7 (31). (——) identical residues; (■) gap introduced to maximize homology; parentheses indicate region in which no assignments are available. The heavy chains are given in order of their capacity to inhibit the reaction between A/J serum antiarsonate and its rabbit anti-idiotype (see Table I). Numbering is sequential. Complementarity-determining regions are outlined. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; X, modified residue (see text).

automated amino acid sequencing yielded no compound giving a consistent signal in the chromatographic systems used for identification, and there was a sharp fall in repetitive yield upon sequencing through this position. However, the amino acid sequence of the tryptic peptide of HP 124E1 extending from position 39 to 65 has also been analyzed in a procedure which uses a dipeptidyl peptidase for fragmentation of the peptide followed by gas chromatography-mass spectrometry for identification of the dipeptides (42, 43). This analysis yielded an identification of Asn at position 55 in this molecule (H. Krutzsch, personal communication). In addition, nucleotide sequence analysis of a V_H region cDNA clone made from the 93G7 cell line (31) showed that in HP 93G7, position 55 is encoded by an Asn codon. In both molecules, therefore, it appears that Asn is present at position 55 but is modified in a way that precludes identification by conventional procedures. In HP R16.7, however, Asn was identified at position 55 by normal means. In HP 91A3, Lys was found to be substituted for Asn at this position.

Ars-A V_H Regions are Closely Related Members of a Single Subgroup (V_HV) but Are Homologous to Chains of the V_HII Subgroup. Fig. 1 presents the five V_H region sequences in comparison with the sequence of the strongly CRI-positive molecule HP 93G7. All the chains fall within the V_HV subgroup (44) and constitute a closely related set in which none of the sequences differs from the prototype at more than 12 positions. The V_H region amino acid sequence of the CRI-positive A/J antiarsonate antibody HP 36–65, which has been deduced from a DNA sequence (32), also belongs to this family.

The V_H segments of these antibodies share residues that differentiate them from all other known antibody families. They, therefore, are believed to be encoded by a separate germline gene or set of germline genes. However, they are closely homologous to V_H segments of the V_HII subgroup (44), including the

antidextran (anti-Dex) antibody family in BALB/c bearing the Dex CRI (23) and the anti-NP antibody family in C57BL/6 that bears the NP^b idiotype (24). Representative comparisons are shown in Fig. 2. HP 93G7 differs from the prototype Dex-positive myeloma protein, MOPC-104E, by 21 residues of the 98 in the V_H segment, and differs from the prototype NP^b-positive antibody, B1-8, by 22 residues. The single most closely related chain that has been completely sequenced is that of the BALB/c myeloma protein MPC 11 (45). This also belongs to the V_HII subgroup and its V_H segment bears 20 differences from HP 93G7. The heavy chain N-terminal amino acid sequences of antibodies belonging to the Ars-B family also belong to the V_HII subgroup (4, 44) and they exhibit a level of homology to antibodies of the Ars-A family comparable to other V_HII sequences.

The 93G7 V_H segment is also related to, but distinctly different from, the sequence of a pool of antiarsonate antibodies derived from the serum of hyperimmune A/J mice after enrichment of the CRI-positive component by isoelectric focusing (46). The incidence of differences, which number 30 in the V_H segment (Fig. 2), is anomalously large when compared with the close relationship among monoclonal members of the Ars-A family. This parallels the situation in the light chains, where it was also found that the hybridoma products were more closely related to each other than they were to the serum sequence (8). This suggests that either a different set of germline genes encodes the CRI-positive antibodies expressed in A/J hyperimmune serum or that the serum antibody structure represents a consensus sequence for an inhomogeneous pool that contains CRI-negative antibodies of the Ars-B family as well as the CRI-positive antibodies of the Ars-A family.

The Different V_H Segments Probably Arise from a Single Germline V_H Gene by Somatic Mutation. In antibodies of the Ars-A family, the segment encoded by the V_H gene extends from position 1 to 98 of the mature heavy chain (34). Siekevitz et al. (34) have cloned *Eco R1* restriction fragments of embryonic A/J DNA and,

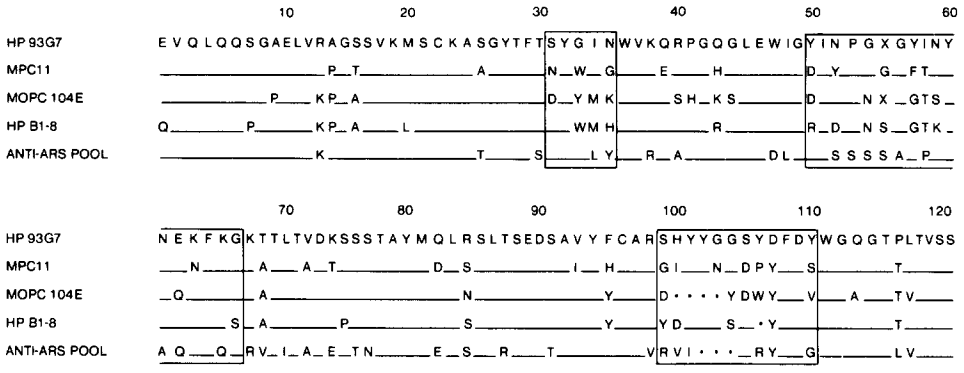

FIGURE 2. Relationship of heavy chains of V_HII subgroup and of A/J serum antiarsonate with an antiarsonate hybridoma heavy chain. Comparison of amino acid sequences of heavy chain V regions of MPC 11 (45), MOPC-104E (61), and HP B1-8 (24), and of A/J serum antiarsonate (46) with that of HP 93G7 (31). (——) identical residues; (■) gap introduced to maximize homology. Numbering is sequential. Complementarity-determining regions are outlined. For single letter code for amino acid residues, see legend to Fig. 1.

among inserts which hybridized strongly to a V_H segment DNA probe made from the CRI-positive cell line 36-65, they distinguished three different genes by restriction mapping and DNA sequencing. One had a sequence identical to the V_H segment expressed by the 36-65 cell line. The amino acid sequences described in the present paper are best interpreted as having been derived from this germline gene. The other two genes, although extremely similar to the first, encode at least eight amino acids that are different from it. Only one of these amino acids has been found in any sequences of members of the Ars-A family. This result represents the strongest evidence currently available that a single germline V_H gene is used by members of the Ars-A family, and that differences in their V_H segment sequences must be ascribed predominantly to somatic mutation. However, it does not exclude the possibility that sequence information from portions of more remotely related genes is occasionally used by the Ars-A family after unequal crossing over or gene conversion events. A thorough investigation of the possibility that multiple germline V_H segment genes may contribute to intrafamily variation has also been conducted for BALB/c anti-PC antibodies of the T15 family (47) and for C57BL/6 anti-NP antibodies of the NP^b family (24). In these cases, too, the amino acid or DNA sequences of the family of expressed V_H segments were compared with the nucleotide sequences of the set of germline V_H genes most closely related to them. Although several closely similar V_H genes were identified and sequenced in each family, it was apparent in both cases that the expressed proteins nearly always derived from only one germline gene of the set. These observations form the basis of an emerging consensus of opinion that somatic mutation plays a dominant role in diversification of the V segment sequences within all such antibody families.

Loss of CRI Is Due to Mutational Divergence from the Germline Sequence. Fig. 3

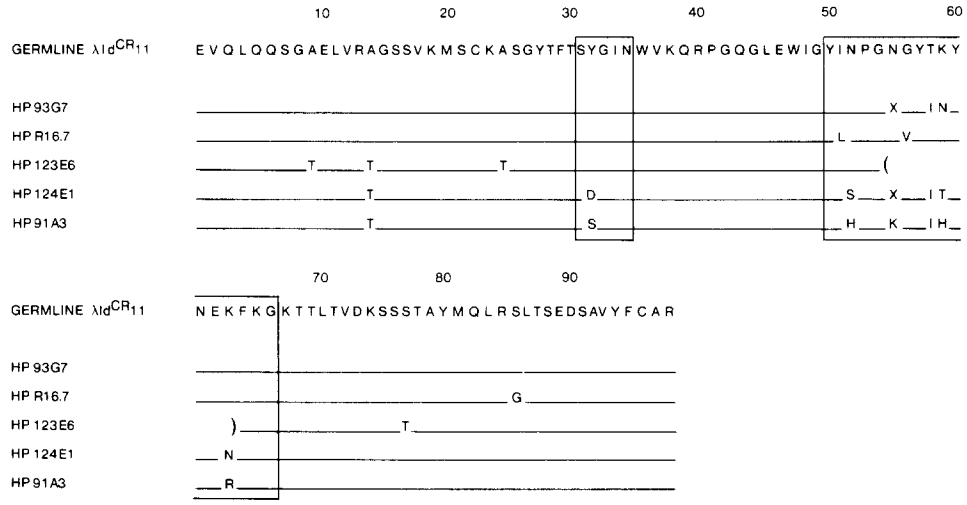

FIGURE 3. Relationship of antiarsonate hybridoma V_H segments to a germline V_H segment gene. Comparison of amino acid sequences of V_H segments with the sequence encoded by the genomic clone λId^{CR}11 (34). (——) identical residues. Numbering is sequential. Complementarity-determining regions are outlined. For single letter code for amino acid residues, see legend to Fig. 1.

shows the V_H segment amino acid sequences of the five antibodies compared with the sequence encoded by the genomic clone $\lambda Id^{CR}11$, a representative clone bearing the V_H gene used by members of the Ars-A family (34). The series of five expressed antibodies shows differences from the germline sequence at a total of 13 positions of the 98 in the V_H segment (13%). All of the replacements can be accounted for by single nucleotide substitutions. The number of differences shown by each antibody (excluding secondary modifications of Asn-55) ranges from two for HP R16.7 to seven for HP 91A3. It is noteworthy that the antibodies that are most strongly CRI-positive show fewer differences from the germline sequence than antibodies that are weakly CRI-positive or CRI-negative. This observation is consistent with the hypothesis that the germline encodes the amino acid sequence of a strongly CRI-positive antibody, and that reduction in the level of CRI expression in members of the Ars-A family is due to divergence from this sequence by somatic mutation.

This result has a parallel among the corresponding V_L segments (8). Direct information about DNA sequences of V_L genes used by members of the Ars-A family is not yet available. However, if it is assumed that a single germline gene encodes all known V_L segments in this family, and if it is further assumed that the most common amino acid at each position is encoded by this germline gene, then the most strongly CRI-positive antibodies are again found to have the fewest differences from the putative germline sequence (8, 9, 37).

There Is Clustering of Positions Bearing Substitutions in the V_H Segments. The linear distribution of V_H segment positions bearing substitutions from the germline sequence is clearly different from that expected on the basis of a random distribution of mutations. The incidence of amino acid replacements within the first and second complementarity-determining regions (CDR) (14.9%), calculated as the total number of residues different from the germline sequence in these regions divided by the total number of residues sampled, is 8.3 times higher than the corresponding incidence in the framework regions (1.8%), not counting the posttranslational modifications at position 55. CDR2 provides a special focus of variability, with an incidence of replacements (19.7%) that is 9.0 times higher than the incidence in the remainder of the V_H segment (2.2%). Historically, the hypervariable regions in the V_H and V_L chains were delineated by comparison of chains that belonged to the same variable region subgroup but which had been selected without regard to the antigen specificity of the parent myeloma proteins (48–50). The localization of differences could thus be interpreted as due, at least in part, to differences in antigen specificity, since most of the hypervariable regions were subsequently shown to contain complementarity-determining amino acids. More recent comparisons, such as the present study, have focused on antibodies that belong to the same family. These antibodies share not only V_H and V_L subgroups in common, but also specificity for the same antigen. In some of these studies, hypervariability in CDR3 of the heavy chain has been observed (23, 25), but variability in CDR1 and CDR2 has in no case been prominent. There is thus a general trend towards diminution of hypervariability in the CDRs and this has usually been attributed to selection against mutations in these regions based on the requirement for binding to the same antigen (51). The present set of antiarsonate V_H chains provides a significant departure from this trend, since

CDR2 (as well as CDR3; see below) provides a conspicuous focus for variability. It may be important that these chains, unlike the ones sequenced in other studies, have been drawn from antibodies showing a very broad range of idiotypic expression. This unusual result may thus be associated with variation in CRI expression. Such an interpretation receives support from the observation that the binding of CRI and anti-CRI antibodies is largely inhibitable by hapten, suggesting that residues in the CDR play a role in expression of the CRI (17).

Several V_H Positions that Bear Substitutions are "Hot Spots" of Variation. The 22 amino acid residues that differ from the germline V_H sequence are located at just 13 positions. Seven of these positions bear replacements in one V_H segment only, but three bear replacements in two V_H segments and three bear replacements in three V_H segments. There is thus a clustering of replacements in "hot spots" of variation. Examples of these hot spots are to be found in both framework and hypervariable regions. Once again, since this set of antibodies was selected to maximize variation in the degree of expression of the CRI, some positions may owe their high incidence of replacements directly or indirectly to selection for loss of the CRI.

At least two of the multiply substituted positions, 14 and 58, show identical repeated replacements. At position 14, three V_H segments (HP 123E6, HP 124E1, and HP 91A3) have a variant Thr in place of the germline-encoded Ala, and at position 58, three V_H segments (HP 93G7, HP 124E1, and HP 91A3) have a variant Ile in place of the germline-encoded Thr. In both cases, the substitutions are found in V_H segments drawn from at least two different animals in independent fusion experiments. Such repeated substitutions are likely to be a common feature of antibodies belonging to the Ars-A family since further substitutions identical to ones reported here have been discovered at five positions in chains sequenced independently by Margolies et al. (21, 34, and unpublished observations). They include the replacement of the germline Ala by Thr at position 9 in two further proteins, HP 45-223 and HP 3D10, and a further example of the replacement of the germline Ala by Thr at position 14 in the protein HP 45-223 (21, and unpublished observations summarized in 34). A repeat of the replacement of the germline Gly by Val at position 56, of germline Lys by Asn at position 63, and of germline Ser by Gly at position 85 have all been identified in another antibody, HP 1F6 (Margolies, personal communica-

$D_{FL16.1}$	Y Y Y G S S Y
HP 93G7	S H ____ G ___
HP R16.7	S N ____ G ___
HP123E6	S V R _ D _ G ___
HP124E1	S D F ___ G ___
HP 91A3	S ___ S G M _

FIGURE 4. Relationship of antiarsonate hybridoma D_H segments to a germline D_H segment gene. Comparison of amino acid sequences of D_H segments with the sequence encoded by the $D_{FL16.1}$ gene of BALB/c (52). (——) identical residues. For single letter code for amino acid residues, see legend to Fig. 1.

tion). Even in this wider series of repeated substitutions, however, no correlation between substitutions at different positions is yet discernible. Repeated substitutions of this kind have also been observed among the V_L segments of the Ars-A family (8, 37) and among chains from other systems such as the V_L segments from inulin-binding myeloma proteins (30). The V_H segments from antiarsonate antibodies of the Ars-A family, however, provide the best evidence so far that the phenomenon may arise from repeated yet independent somatic mutation events.

D_H Segment Diversity Adds Greatly to Structural Variability within the Ars-A Family and Is Probably Generated Somatically. The D_H segment, which encompasses a portion of CDR3, provides a second major focus for differences between the members of the Ars-A family. Its N-terminal boundary lies between positions 98 and 99, such that the codon for position 98 is supplied by the V_H gene and the codon for position 99 is supplied by the D_H gene (34). There is no junctional diversity at the V_H-D_H boundary since all five D_H segments have the same first amino acid, Ser. The C-terminus of the D_H segment is located at position 107, which marks the site of joining between the D_H and J_H genes. At this position, there are three different residues in the five molecules, an example of junctional diversity that will be discussed in the next section.

Between the boundaries of the D_H segment, the five chains differ in both sequence and length. After the introduction of gaps in the shorter sequences to maximize homology (Fig. 1), it is found that all of the variation in length is confined to the second and third D_H segment positions (positions 100 and 100a in the heavy chain). Also located at these two positions are five of the nine D_H segment residues that differ from the HP 93G7 prototype. The remaining four differences are scattered over four positions in a less variable region or "core" segment of D_H extending from position 101 to 106. Two of these differences occur in HP 123E6 and HP 124E1, respectively, and two occur in HP 91A3. One of the two differences in HP 91A3, the substitution of a Met for a Ser at position 105, would require a double-base change.

Although there exists no direct information on germline D_H genes in the A/J mouse, nucleotide sequences for 10 different D_H genes have so far been described in the BALB/c (52). These fall into three families, D_{FL16}, D_{SP2}, and D_{52}. The prototype gene for the D_{FL16} family, $D_{FL16.1}$, is the most closely homologous to the D_H segment core of the Ars-A antibodies. A comparison is shown in Fig. 4. All of the chains share a common difference from $D_{FL16.1}$, the substitution of a Gly for a Ser at position 104. Since this Gly is present in all the antibodies, it is most likely to be encoded in the germline gene or genes that give rise to this family of D_H segments. This suggests that a hitherto undescribed germline member of the D_{FL16} gene family is recruited in these antibodies. The number of germline D_{FL16} genes encoding the different D_H segments in the Ars-A family is unknown, but an attractive possibility is that a single gene, which encodes the prototype core sequence found in HP 93G7 and HP R16.7, is used by all the molecules. Under this hypothesis, the amino acid differences seen in HP 123E6, HP 124E1, and HP 91A3 result from somatic mutation, and the smallest numbers of mutations are once again found in the most strongly CRI-positive antibodies, as was the case in the V_H segments. The relationship of this hypothetical single

J_H2	Y F D Y W G Q G T T L T V S S
HP 93G7	D ——— P ———
HP R16.7	S ———
HP 123E6	—— S ——— G ——
HP 124E1	—— N ———
HP 91A3	S ——— S ———

FIGURE 5. Relationship of antiarsonate hybridoma J_H segments to a germline J_H segment gene. Comparison of amino acid sequences of J_H segment with the sequence encoded by the J_H2 gene of BALB/c (57, 58). (——) identical residues. For single letter code for amino acid residues, see legend to Fig. 1.

gene to $D_{FL16.1}$ is, of course, unknown. It may represent the A/J allelic counterpart to the BALB/c $D_{FL16.1}$ gene or it may define a new locus within the D_{FL16} complex.

In the comparison of the Ars-A D_H segments with the sequence encoded by $D_{FL16.1}$ (Fig. 4), amino acids at positions 99, 100, and 100a appear as hypervariable N-terminal extensions to the core sequence. Although $D_{FL16.1}$ encodes a Tyr immediately preceding the core sequence, this particular residue has not been identified in any of the antibodies at position 100a but is always replaced by a different amino acid. It is not known to what extent nucleotides from the germline codon at this position are used to create this diversity, but it is interesting to note that while the second and third nucleotides of a hypothetical Tyr codon are appropriate for creating a codon for the His in HP 93G7, the Asn in HP R16.7, and the Asp in HP 124E1, they are inappropriate for creating codons for the Arg in HP 123E6 and the Ser in HP 91A3. Extensions similar to the ones described here have been noted to occur in D_H segments in a variety of systems (52, 53). They are characterized by the presence of amino acids at either the N- or C-terminus of a D_H core sequence that are determined by codons unknown as parts of any germline D_H gene related to the one encoding the core. The genetic origin of amino acids in extensions of this kind is unknown, but some radical explanations are current. One possibility is the involvement of a second D_H gene that becomes attached to the core gene in a D_H-D_H joining process (52). However, this explanation is currently considered less attractive (54) than the alternative possibility that the amino acids result from the addition of nucleotides at the DNA level during V_H-D_H joining by the template-independent action of a terminal deoxynucleotidyl transferase, thus creating a so-called "N segment" (53). A possible objection to the latter mechanism in the present case is that it lacks the specificity required to place the same amino acid, Ser, at the N-terminus of the extension in every molecule. However, this objection may be circumvented by asserting that the Ser is essential in some way for the integrity of the antibody molecule or for its ability to combine with antigen. The most common length for the extension piece in the present set of antibodies is two residues, but HP 123E6 and HP 91A3 have extensions with the atypical lengths of three and one residue, respectively. It is not yet known whether any significance attaches to the fact that both these molecules also show reduced levels of

expression of the CRI. Variation in the number of residues present within CDR3 has been recognized for a long time and it is now known that variation in the lengths of D_H segment extension pieces contributes to this. However, there is a strong tendency toward conservation of the number of residues in CDR3 among antibodies of a family binding the same antigen (see, for example, 39, 56). The variation in length observed in the Ars-A family provides an interesting departure from this trend, the stereochemical implications of which remain to be determined.

Modest J_H Segment Diversity Is Somatic in Origin. The J_H segment extends from the site of D_H-J_H joining at position 107 to the C-terminus of the V_H region at position 121. All five J_H segments are different. However, they all have sequences that bear closest homology to the germline *J_H2* gene of BALB/c (57, 58). Although data about germline *J_H* genes in the A/J mouse are incomplete at the DNA level, this result suggests that all five J_H segments are the products of a single germline gene representing the A/J homolog of *J_H2*. A comparison of the five Ars-A J_H segments with the sequence encoded by *J_H2* is shown in Fig. 5. The chains differ from the germline sequence at one or two positions each. The total of eight amino acid differences are distributed over six positions and the only one of these positions that shows more than one replacement is the junctional position, 107. This distribution of substitutions is consistent with the hypothesis that the germline *J_H* gene used by these antibodies encodes the same amino acid sequence as BALB/c *J_H2* and that all differences from this sequence are generated somatically. In the J_H segment, there is no correlation between the numbers of differences from germline and the degree of expression of the CRI in the parent antibody.

The junctional diversity at position 107 cannot be accounted for solely by variation in the position of D_H-J_H joining (for a general discussion of gene segment joining, see 59). The *J_H2* gene has a Tyr codon, TAC, at the first position (57, 58) and Tyr is present in two of the five antibodies, HP 123E6 and HP 124E1. HP 93G7 has an Asp at this position, encoded by a GAC codon (31). This could have been generated by a joining event that brought a hypothetical G from the 3′ end of the *D_H* gene into apposition with AC from the Tyr codon at the 5′ end of the *J_H* gene, thus creating a hybrid codon for Asp. Alternatively, it could have been generated from the *J_H* Tyr codon by a mutational change of the first base from T to G. Both HP R16.7 and HP 91A3 have a Ser at position 107, the former encoded by a TCC codon (A. Maxam, personal communication). This codon could have been generated twice by two independent mutations of the second base in the Tyr codon from A to C. It could also have arisen by the formation of a hybrid codon between a *D_H* gene and a *J_H* gene. However, there is no way both the Asp codon in HP 93G7 and the Ser codons in HP R16.7 and HP 91A3 could arise by this mechanism if the same *D_H* and *J_H* genes were involved, whatever the position of joining might have been. The only valid alternatives to the mutational explanation would, therefore, be ones that relied on the existence of two *D_H* genes with different nucleotide sequences at their 3′ ends or two *J_H* genes with different nucleotide sequences at their 5′ ends.

Structural Basis of the CRI. The present study represents a sampling of V_H region substitutions that are associated with changes in expression of the CRI.

Some of the positions at which these substitutions occur, but almost certainly not all, form parts of the determinants comprising the CRI. Since the study excludes substitutions that compromise antigen binding, the observed variants may represent only a subset of residues that provide attachment sites for anti-CRI antibodies. However, it is this subset which is permitted to vary during an antigen-driven immune response and which therefore has potential physiological significance as the basis for selection between idiotypically distinct antibody forms.

In the three antibodies showing diminished expression of the CRI, there are, in total, 20 positions in the heavy chains and 7 positions in the light chains that bear residues not present in one or other of the strongly CRI-positive antibodies. These residues signal positions with potential involvement in the determinants comprising the CRI. Since the CRI itself is defined by a rabbit anti-CRI serum that is polyclonal, and the anti-CRI is conventionally prepared against A/J anti-Ars serum antibodies that are also polyclonal (5), it is impossible on the basis of serology alone to eliminate any of these positions as possible sites for attachment of anti-CRI antibodies. Some of the difficulties inherent in the conventional definition of the CRI can be avoided if idiotypic specificities defined only in terms of monoclonal antiidiotypic antibodies are considered. However, the problem is then merely transformed into one of identifying the antigenically dominant determinants comprising the CRI among the various determinants recognized by the monoclonal reagents, since any one of these determinants may contribute only a small part to the total antigenicity of the V regions.

Despite these difficulties there are indications of the way residues comprising the CRI are localized. For example, the ability of hapten largely to inhibit the interaction of CRI and anti-CRI suggests that residues in the CDR are important constituents of determinants comprising the CRI (17). Also, antibodies belonging to the Ars-A family but differing with respect to their expression of the CRI show, in chain recombination experiments, that substitutions in the heavy chains are predominantly responsible for variation in idiotypic expression (18). If attention is accordingly restricted to the CDR in the heavy chain, the number of positions which have potential roles as constituents of the CRI can be narrowed to 14. One of these positions is in CDR1, four are in CDR2, and nine are in CDR3. Some of these positions may eventually be eliminated from consideration if they are found to bear identical substitutions in further strongly CRI-positive antibodies. The role of others may be defined in the course of site-specific mutagenesis experiments. On the other hand, structural analysis of additional CRI-negative antibodies of the Ars-A family may result in the identification of further positions at which substitutions are able to modify CRI expression.

Factors Underlying the Pattern of Variation. Various aspects of the distribution of substitutions are associated with the loss of CRI in this set of molecules. These include the greater incidence of replacements in antibodies showing diminished expression of the CRI, the unusual degree of clustering of replacements in the CDR, and the localization of much of the variation in hot spots at which antibodies with weakened CRI expression can be distinguished from strongly CRI-positive antibodies. However, since the proportion of variants that have a direct effect on the CRI is unknown, and may be small, it is possible that these patterns may

be determined by factors only indirectly related to variation in the degree of expression of the CRI.

The first of these factors is mutation rate. In the absence of selection, mutations in the V regions accumulate through time to an extent that is believed to depend on the duration of developmental time in which the expressed variable region genes are subject to a high rate of somatic mutation (13, 25). The distribution of substitutions is controlled to an unknown extent by differences in the rates at which mutational changes and other somatic alterations arise at different sites. It is possible that mutations at positions affecting expression of the CRI occur only rarely relative to mutations at other positions and therefore arise predominantly in molecules sharing a high probability of bearing replacements at sites irrelevant to expression of the CRI. Selection for reduction in CRI expression will then give rise to molecules that bear replacements at several positions which share only a relatively high mutation rate in common. This may account for the otherwise anomalous observation of concordant variation in the numbers of substitutions in both the V_H and V_L chains, despite the lack of any detectable dependency of CRI expression on variation in light chain structure (18). It may also account for the existence of hypervariable positions at which the replacements do not appear precisely in accord with the level of CRI expression, such as positions 58 and 100 in the heavy chain. Hypervariability at positions that are believed to be unimportant for expression of the CRI may also be accounted for in this way. For example, position 14 in the heavy chain bears a substitution in all three molecules showing reduced or absent expression of the CRI, but is located in a framework region where it may be remote from the site of interaction with antiidiotypic antibodies.

The second factor that modulates the distribution of mutations is the stabilizing selection which suppresses amino acid substitutions at functionally crucial sites and permits only "conservative" changes (substitution of acidic residues by acidic, basic by basic, hydrophobic by hydrophobic, etc.) at other locations. The requirement for antigen binding is able to mediate such selection. The replacements that are found in the face of such stabilizing selection are ones which do not compromise the general integrity of the three-dimensional structure and cause no significant alterations in the specific sites that interact with antigen. An indication of the action of this and other kinds of selection may be obtained during sequencing studies at the nucleic acid level, where it will be possible to measure ratios of the incidence of silent and nonsilent base changes and to compare them to the ratios expected on the basis of unselected mutation. A significant elevation of base changes that give rise to amino acid substitutions can be taken as an indication of differential selection (60).

The third factor controlling the distribution of variation is selection for divergence of amino acid sequences. Although overt selection for loss of CRI expression has been imposed here during the choice of antibodies for sequence analysis, the patterns of variation that result from this selection may reflect not only the distribution of mutations affecting the CRI itself but also the incidence of mutations affecting other, as yet unidentified properties which are closely correlated with expression of the CRI in vivo. Possible examples of such properties are particular fine specificity patterns or the appearance or disappearance

of distinct idiotypic specificities.

These various factors are able to interact in a complex way to produce the patterns of variation seen in the Ars-A family. They are presumably no less actively at work in other systems of expressed antibodies, and it is perhaps the complexity of their interaction, more than any other component of the antibody diversity problem, which remains for the future to elucidate.

Summary

V_H region amino acid sequences are described for five A/J anti-*p*-azophenylarsonate (anti-Ars) hybridoma antibodies for which the V_L region sequences have previously been determined, thus completing the V domain sequences of these molecules. These antibodies all belong to the family designated Ars-A which bears the major anti-arsonate cross-reactive idiotype (CRI) of the A strain mouse. However, they differ in the degree to which they express the CRI in standard competition radioimmunoassays. Although the sequences are closely related, all are different from each other. Replacements are distributed throughout the V_H region and occur in positions of the chain encoded by all three gene segments, V_H, D_H, and J_H. It is likely that somatic diversification processes play a dominant role in producing the sequence variability in each of these segments. The number of differences from the sequence encoded by the germline is smallest for antibodies that express the CRI most strongly, suggesting that somatic diversification is responsible for loss of the CRI in members of the Ars-A antibody family. There is an unusual degree of clustering of differences in both CDR2 and CDR3 and many of the substitutions are located in "hot spots" of variation. The large number of differences between the chains prohibits the unambiguous identification of positions at which alterations play a major role in reducing the expression of the CRI. However, the data suggest that the loss of the CRI is associated with a definable repertoire of somatic changes at a restricted number of highly variable sites.

We are grateful to Dr. Pila Estess for providing the 93G7, 123E6, 124E1, and 91A3 hybridomas and Dr. Al Nisonoff and Dr. Edmundo Lamoyi for their gift of the R16.7 hybridoma. It is a pleasure to acknowledge the skilled technical assistance of Sandy Graham, Priscilla Presley, and Ana Garcia-Spain and the secretarial services of Kathy van Egmond.

Received for publication 5 July 1983.

References

1. Capra, J. D., and T. J. Kindt. 1981. Antibody diversity: something for everyone. Editorial. *Scand. J. Immunol.* 13:205.
2. Kohler, G., and C. Milstein. 1976. Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion. *Eur. J. Immunol.* 6:511.
3. Estess, P., A. Nisonoff, and J. D. Capra. 1979. Structural studies on induced antibodies with defined idiotypic specificities. VIII. NH_2-terminal amino acid sequence analysis of heavy and light chain variable regions of monoclonal anti-*p*-azophenylarsonate antibodies from A/J mice differing with respect to a cross-reactive idiotype. *Mol. Immunol.* 16:1111.

4. Milner, E. C. B., and J. D. Capra. 1982. V_H families in the antibody response to *p*-azophenylarsonate: correlation between serology and amino acid sequence. *J. Immunol.* 129:193.
5. Kuettner, M. G., A.-L. Wang, and A. Nisonoff. 1972. Quantitative investigations of idiotypic antibodies. VI. Idiotypic specificity as a potential genetic marker for the variable regions of mouse immunoglobulin polypeptide chains. *J. Exp. Med.* 135:579.
6. Pawlak, L. L., E. B. Mushinski, A. Nisonoff, and M. Potter. 1973. Evidence for the linkage of the IgC_H locus to a gene controlling the idiotypic specificity of anti-*p*-azophenylarsonate antibodies in strain A mice. *J. Exp. Med.* 137:22.
7. Eichmann, K. 1975. Genetic control of antibody specificity in the mouse. *Immunogenetics.* 2:491.
8. Siegelman, M., and J. D. Capra. 1981. Complete amino acid sequence of light chain variable regions derived from five monoclonal anti-*p*-azophenylarsonate antibodies differing with respect to a cross-reactive idiotype. *Proc. Natl. Acad. Sci. USA.* 78:7679.
9. Ball, R. K., J.-Y. Chang, S. S. Alkan, and D. G. Braun. 1983. The complete amino acid sequence of the light chain variable region of two monoclonal anti-*p*-azobenzenearsonate antibodies bearing the cross-reactive idiotype. *Mol. Immunol.* 20:197.
10. Slaughter, C. A., M. Siegelman, P. Estess, I. Barasoain, A. Nisonoff, and J. D. Capra. 1982. Antibody diversity and idiotypes: primary structural analysis of monoclonal A/J anti-arsonate antibodies. *In* Developmental Immunology: Clinical Problems and Aging. E. L. Cooper and M. A. B. Brazier, editors. Academic Press, Inc., New York. 45–67.
11. Sakano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination site of immunoglobulin light chain genes. *Nature (Lond.).* 280:288.
12. Max, E. E., J. G. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin K constant region gene. *Proc. Natl. Acad. Sci. USA.* 76:3450.
13. Kaartinen, M., G. M. Griffiths, A. F. Markham, and C. Milstein. 1983. mRNA sequences define an unusually restricted IgG response to 2-phenyl oxazolone and its early diversification. *Nature (Lond.).* 304:320.
14. Schiff, C., and M. Fougereau. 1975. Determination of the primary structure of a mouse IgG2a immunoglobulin. Amino acid sequence of the light chain. *Eur. J. Biochem.* 59:525.
15. Estess, P., E. Lamoyi, A. Nisonoff, and J. D. Capra. 1980. Structural studies on induced antibodies with defined idiotypic specificities. IX. Framework differences in heavy- and light-chain variable regions of monoclonal anti-*p*-azophenylarsonate antibodies from A/J mice differing with respect to a cross-reactive idiotype. *J. Exp. Med.* 151:863.
16. Lamoyi, E., P. Estess, J. D. Capra, and A. Nisonoff. 1980. Heterogeneity of an intrastrain cross-reactive idiotype associated with anti-*p*-azophenylarsonate antibodies of A/J mice. *J. Immunol.* 124:2834.
17. Lamoyi, E., P. Estess, J. D. Capra, and A. Nisonoff. 1980. Presence of highly conserved idiotypic determinants in a family of antibodies that constitute an instrastrain cross-reactive idiotype. *J. Exp. Med.* 152:703.
18. Milner, E. C. B., and J. D. Capra. 1983. Structural analysis of monoclonal anti-arsonate antibodies: idiotypic specificities are determined by the heavy chain. *Mol. Immunol.* 20:39.
19. Kresina, T. F., S. M. Rosen, and A. Nisonoff. 1982. Degree of heterogeneity of binding specificities of antibodies to the phenylarsonate group that share a common idiotype. *Mol. Immunol.* 19:1433.
20. Gefter, M. L., and T. L. Rothstein. 1983. Affinity analysis of idiotype-positive and

idiotype-negative Ars-binding hybridoma proteins and Ars-immune sera. *Mol. Immunol.* 20:161.

21. Margolies, M. N., A. Marshak-Rothstein, and M. L. Gefter. 1981. Structural diversity among anti-*p*-azophenylarsonate monoclonal antibodies from A/J mice: comparison of Id^- and Id^+ sequences. *Mol. Immunol.* 18:1065.
22. Alkan, S. S., R. Knecht, and D. G. Braun. 1980. The cross-reactive idiotype of anti-4-azobenzene-arsonate hybridoma-derived antibodies in A/J mice constitutes multiple heavy chains. *Hoppe-Seyler's Z. Physiol. Chem.* 361:191.
23. 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.
24. Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore. 1981. Heavy chain variable region contribution to the NP^b family of antibodies: somatic mutation evident in gamma-2a variable region. *Cell.* 24:625.
25. Gearhart, P. J., N. D. Johnson, R. Douglas, and L. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.).* 291:29.
26. Kaartinen, M., G. M. Griffiths, P. H. Hamlyn, A. F. Markham, K. Karjalainen, J. L. T. Pelkonen, O. Makela, and C. Milstein. 1983. Anti-oxazolone hybridomas and the structure of the oxazolone idiotype. *J. Immunol.* 130:937.
27. Ruf, J., C. Tonnelle, J. Rocca-Serra, D. Moinier, M. Pierres, S.-T. Ju, M. E. Dorf, J. Theze, and M. Fougereau. 1983. Structural bases for public idiotype specificities of monoclonal antibodies directed against poly ($Glu^{60}Ala^{30}Tyr^{70}$) and poly ($Glu^{80}Ala^{40}$) random copolymers. *Proc. Natl. Acad. Sci. USA.* 80:3040.
28. Pawlita, M., M. Potter, and S. Rudikoff. 1982. K-chain restriction in anti-galactan antibodies. *J. Immunol.* 129:615.
29. Vrana, M. S., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin-binding myeloma proteins. *Proc. Natl. Acad. Sci. USA.* 75:1957.
30. Johnson, N., J. Slankard, L. Paul, and L. Hood. 1982. The complete V domain amino acid sequences of two myeloma inulin-binding proteins. *J. Immunol.* 128:302.
31. Sims, J., T. H. Rabbitts, P. Estess, C. A. Slaughter, P. W., Tucker, and J. D. Capra. 1982. Somatic mutation in genes for the variable portion of the immunoglobulin heavy chain. *Science (Wash. D. C.).* 216:309.
32. Siekevitz, M., M. L. Gefter, P. Brodeur, R. Riblet, and A. Marshak-Rothstein. 1982. The genetic basis of antibody production: the dominant anti-arsonate idiotype response of the strain A mouse. *Eur. J. Immunol.* 12:1023.
33. Estess, P., F. Otani, E. C. B. Milner, J. D. Capra, and P. W. Tucker. 1982. Gene rearrangements in monoclonal A/J anti-arsonate antibodies. *J. Immunol.* 129:2319.
34. Siekevitz, M., S. Y. Huang, and M. L. Gefter. 1983. The genetic basis of antibody production: a single heavy chain variable region gene encodes all molecules bearing the dominant anti-arsonate idiotype in the strain A mouse. *Eur. J. Immunol.* 13:123.
35. Nelles, M. J., L. A. Gill-Pazaris, and A. Nisonoff. 1981. Monoclonal anti-idiotypic antibodies reactive with a highly conserved determinant on A/J serum anti-*p*-azophenylarsonate antibodies. *J. Exp. Med.* 154:1752.
36. Siegelman, M., C. A. Slaughter, L. J. McCumber, P. Estess, and J. D. Capra. 1981. Primary structural studies of monoclonal A/J anti-arsonate antibodies differing with respect to a cross-reactive idiotype. *In* Immunoglobulin Idiotypes. C. Janeway, E. E. Sercarz, and H. Wigzell, editors. Academic Press, Inc., New York. 135–158.
37. Capra, J. D., C. A. Slaughter, E. C. B. Milner, P. Estess, and P. W. Tucker. 1982. The cross-reactive idiotype of A-strain mice: serological and structural studies. *Immunol. Today (Amst.)* 3:332.

38. Schulman, M., C. D. Wilde, and G. Kohler. 1978. A better cell line for making hybridomas secreting specific antibodies. *Nature (Lond.).* 276:269.
39. Andrews, D. W., and J. D. Capra. 1981. Amino acid sequence of the variable regions of heavy chains from two idiotypically cross-reactive human IgM anti-gamma-globulins of the Wa group. *Biochemistry.* 20:5822.
40. Andrews, D. W., and J. D. Capra. 1981. Amino acid sequence of the variable regions of light chains from two idiotypically cross-reactive human IgM anti-gamma-globulins of the Wa group. *Biochemistry.* 20:5816.
41. Klapper, D. G., C. W. Wilde, III, and J. D. Capra. 1978. Automated amino acid sequence of small peptides utilizing polybrene. *Anal. Biochem.* 85:126.
42. Krutzsch, H. C., and J. J. Pisano. 1977. Analysis of dipeptides by gas chromatography-mass spectrometry and application to sequencing with dipeptidyl aminopeptidases. *Methods Enzymol.* 47:391.
43. Krutzsch, H. C. 1983. Polypeptide sequencing with dipeptidyl peptidases. *Methods Enzymol.* 91:511.
44. Kabat, E. A., T. T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. 1983. Sequences of proteins of immunological interest. U. S. Department of Health and Human Services.
45. Zakut, R., J. Cohen, and D. Givol. 1980. Cloning and sequence of the cDNA corresponding to the variable region of immunoglobulin heavy chain MPC 11. *Nucleic Acids Res.* 8:3591.
46. Capra, J. D., and A. Nisonoff. 1979. Structural studies on induced antibodies with defined idiotypic specificities. VII. The complete amino acid sequence of the heavy chain variable region of anti-*p*-azophenylarsonate antibodies from A/J mice bearing a cross-reactive idiotype. *J. Immunol.* 123:279.
47. Crews, S., J. Griffin, H. Huang, K. Calame, and L. Hood. 1981. A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell.* 25:59.
48. Milstein, C., and J. L. R. Pink. 1970. Structure and evolution of immunoglobulins. *Prog. Biophys. Mol. Biol.* 21:209.
49. Wu, T. T., and E. A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132:211.
50. Capra, J. D., and J. M. Kehoe. 1974. Variable region sequences of five immunoglobulin heavy chains of the V_HIII subgroup. Definitive identification of four heavy chain hypervariable regions. *Proc. Natl. Acad. Sci. USA.* 71:845.
51. Capra, J. D. 1981. Antibody diversity: a somatic model. *In* Immunoglobulin Idiotypes. C. Janeway, E. E. Sercarz, H. Wigzell, editors. Academic Press, Inc., New York. 825–830.
52. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* 155:201.
53. Alt, F. W., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-J_H fusions. *Proc. Natl. Acad. Sci. USA.* 79:4118.
54. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature (Lond.).* 302:575.
55. Kehoe, J. M., and J. D. Capra. 1971. Localization of two additional hypervariable regions in immunoglobulin heavy chains. *Proc. Natl. Acad. Sci. USA.* 68:2019.
56. Auffray, C., J.-L. Sikorav, R. Ollo, and F. Rougeon. 1981. Correlation between D region structure and antigen-binding specificity: evidences from the comparison of closely related immunoglobulin V_H sequences. *Ann. Immunol. (Paris).* 132D:77.

57. Bernard, O., and N. M. Gough. 1980. Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated V_H and constant region genes. *Proc. Natl. Acad. Sci. USA.* 77:3630.
58. 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.
59. Rudikoff, S., D. N. Rao, C. P. J. Glaudemans, and M. Potter. 1980. K chain joining segments and structural diversity of antibody combining sites. *Proc. Natl. Acad. Sci. USA.* 77:4270.
60. Loh, D. Y., A. L. M. Bothwell, M. E. White-Scharf, T. Imanishi-Kari, and D. Baltimore. 1983. Molecular basis of a mouse strain-specific anti-hapten response. *Cell.* 33:85.
61. Kehry, M. R., J. S. Fuhrman, J. W. Schilling, T. Rogers, C. H. Sibley, and L. E. Hood. 1982. Complete amino acid sequence of a mouse mu chain: homology among heavy chain constant region domains. *Biochemistry.* 21:5415.
62. Estess, P. 1980. Doctoral dissertation. University of Texas Health Science Center at Dallas.