MULTIPLE V_H GENE SEGMENTS ENCODE MURINE ANTISTREPTOCOCCAL ANTIBODIES

BY ROGER M. PERLMUTTER,[*] JOAN L. KLOTZ, MARTHA W. BOND,[‡] MOON NAHM,[§] JOSEPH M. DAVIE, AND LEROY HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125; the City of Hope Medical Center, Duarte, California 91010; and the Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110

Detailed analysis of genes encoding antibody molecules has provided considerable insight into the mechanisms responsible for the extraordinary structural diversity of antibodies (1). In particular, it is clear that combinatorial joining of V_H, D, and J_H gene segments (for heavy chains) and of V_H and J_H gene segments (for light chains), coupled with random combinatorial association of heavy and light chains, may permit the assembly of at least 10^7 different antibody combining sites (2–4).

Perhaps surprisingly, careful study of antibodies of restricted heterogeneity directed against specific antigens has revealed that much of the observed diversity in these antibody populations results from a process of somatic mutation superimposed on a small number of germline genetic elements. In the BALB/c antibody response to phosphorylcholine (PC),[1] for example, a single germline V_H gene segment (V1) and a single germline J_H gene segment (J_H1) direct the synthesis of virtually all observed heavy chains (5, 6). In this case, heavy chains containing as many as eight variant amino acids result from the superimposition of a localized hypermutational mechanism acting concordantly on V_H and V_κ genes (7, 10). Somatic mutation has also been observed in a family of genes encoding nitrophenacetyl (NP)-specific heavy chains in C57BL/10 mice (11) and in V_H genes encoding antibodies specific for azobenzenearsonate (ARS) (12), oxazolone (OX) (13), and the synthetic polypeptide, glutamic acid-alanine-tyrosine (GAT) (14). In each of these systems a family of closely related antibody molecules is apparently generated through the mutational alteration of a single germline gene segment that serves as a substrate for each of the variant sequences.

To further characterize the genetic basis of antibody diversity, we have conducted a detailed analysis of murine antibodies directed against group A streptococcal carbohydrate (GAC). This antibody family represents an ideal

This work was supported by grants AI18088, AI15926, and AI15353 from the National Institutes of Health.

[*] Recipient of New Investigator Research Award AI18088 from NIH.

[‡] Current address: DNAX Research Institute, Palo Alto, CA 94304.

[§] Current address: Division of Laboratory Medicine, Washington University School of Medicine, St. Louis, MO 63110.

[1] *Abbreviations used in this paper:* ARS, azobenzenearsonate; GAC, streptococcal group A carbohydrate; GAT, glutamic acid-alanine-tyrosine; IEF, isoelectric focusing; NP, nitrophenacetyl; OX, oxazolone; PC, phosphorylcholine; TAE, 0.04 M Tris-acetate, pH 7.5, 0.002 M EDTA.

model system for the study of antibody diversity, since each individual mouse produces one or at most a few predominant antibodies to GAC, while the strain repertoire of such antibodies is quite large, representing at least 200 different species by isoelectric-focusing (IEF) criteria (15–17). We have previously demonstrated that (*a*) murine anti-GAC antibodies are almost entirely restricted to IgM and IgG3 isotypes (18), (*b*) >50% of A/J mouse anti-GAC antibodies share a common light chain identified by spectrotypic (16) or idiotypic (19) criteria, and (*c*) since all A/J anti-GAC hybridomas are spectrotypically distinguishable although most contain an apparently identical light chain, much of the observed antibody diversity in this response may result from the combinatorial pairing of multiple heavy chains with a single light chain (16, 19). Here we report amino-terminal sequence analysis of GAC-specific antibodies as well as the cloning and sequencing of the V_H gene encoding one of the antibody heavy chains. Our analysis demonstrates the existence of multiple, highly homologous germline V_H gene segments encoding A/J anti-GAC antibodies, and underlines the importance of the germline antibody gene repertoire in the generation of antibody diversity.

Materials and Methods

Hybridoma Protein Purification. The production of GAC-specific hybridomas from hyperimmune mice has been described (19). Previously unreported hybridomas were HGAC63, HGAC72, and HGAC85, all $IgG3_\kappa$; and HGAC73, $IgG3_\lambda$. Anti-GAC hybridoma proteins, all $IgG3_\kappa$, were purified by adsorption to *N*-acetyl-glucosamine-conjugated Sepharose 4B and subsequent elution with 10% (wt/vol) *N*-acetyl glucosamine. *N*-acetyl glucosamine is the immunodominant determinant on GAC (20).

Separation of H and L chains. Purified proteins were reduced in 6.5 M guanidine, 0.5 M Tris (pH 8.2), and 0.2 M mercaptoethanol for 1 h and were subsequently alkylated with an equal volume of 0.36 M iodoacetamide, 1 M Tris (pH 8.2) for 15 min before dialysis against 3 M guanidine, 0.2 M ammonium bicarbonate solution as previously described (19). The fully reduced heavy and light chains were then separated by gel filtration using an AcA34 column (3 × 73 cm; LKB, Bromma, Sweden) equilibrated in the guanidine/ammonium bicarbonate solution.

Protein Sequencing. Automated Edman degradations were performed using the Caltech spinning cup sequenator (heavy chains) (21) or a modified Beckman 890B (Beckman Instruments, Inc., Palo Alto, CA) spinning cup sequenator (light chains) as previously described (22). Between 1 and 10 nmol were loaded per sequencing run and all residues were identified by high pressure liquid chromatography (23). Each chain was analyzed during multiple separate sequencing runs.

Library Construction. High molecular weight genomic DNA was isolated from GAC9 hybridoma cells according to the method of Blin and Stafford (24). This DNA was partially digested with *Mbo*I (New England Biolabs) (0.1 U/μg for 15 min) and fragments with an average size of 20 kb were purified by sucrose gradient centrifugation (25). These heterogeneous 20-kb inserts were then ligated into the *Bam*HI site of the L47.1 phage vector (26) and packaged in vitro before plating on KH802 host cells (27).

Library Screening. A single library of 10^6 phage was screened using a probe containing a 3.2-kb insert that included the entire J_H region from germline BALB/c DNA cloned into pBR322 (28), labeled to a specific activity of 10^8 cpm/μg by nick translation (29) using ^{32}P-nucleotide triphosphates (New England Nuclear, Boston, MA). Positive colonies were picked, rescreened, and grown in mass culture for DNA purification (30).

DNA Sequencing. Appropriate fragments were ligated into M13mp8 (Collaborative Research Inc., Lexington, MA) for sequencing using the dideoxynucleotide chain termination method of Sanger et al. (31) and Anderson et al. (32). Coding region sequences were determined on both DNA strands (see Fig. 2).

Southern Blotting. Genomic DNA was isolated from liver or sperm (24). 10 μg of DNA was completely digested with 10 U of restriction enzyme (Bethesda Research Laboratories, Gaithersburg, MD) before electrophoresis in 0.7% agarose in TAE buffer (0.04 M Tris-acetate, pH 7.5, 0.002 M EDTA) for 14 h at 40 V. The gel was then stained in 50 μg/ml ethidium bromide in water before denaturation in 1.5 M sodium chloride, 0.5 M sodium hydroxide for 1 h. Neutralization was accomplished in 1 M ammonium acetate, 0.02 M sodium hydroxide (33) for 1 h before blotting onto nitrocellulose (Schleicher and Schull, Keene, NH). After 3 h, the filter was baked at 80°C under vacuum for 1 h and then allowed to prehybridize at 37°C in 0.8 M sodium chloride, 0.1 M Tris, pH 7.5, 5× Denhardt's solution (34) containing 100 μg/ml boiled salmon sperm DNA as carrier (35), and 50% formamide. Hybridization was performed for 24 h at either 37 or 42°C in fresh prehybridization mix to which denatured ^{32}P-labeled probe (10^6 cpm/ml) had been added. Probes were generated by subcloning into M13mp8 (Collaborative Research Inc.) and were labeled using nick translation (29) or by primer extension from the M13 sequencing site (32). Blots were washed in 0.03 M sodium chloride, 0.003 M sodium citrate, 0.1% sodium dodecyl sulfate at 50°C before autoradiography on Kodak XAR-5 film. For high stringency washing, 0.015 M sodium chloride, 0.0015 M sodium citrate at 68°C was used for 3–5 h. In tests of V_H probes with known sequence, washing at this stringency eliminates hybrids that are <90% homologous (R. M. Perlmutter, unpublished observation).

Results

We have previously reported (16, 19) that the majority of anti-GAC antibodies raised in A/J mice use light chains that are spectrotypically and idiotypically indistinguishable. To further characterize the structural diversity of GAC-binding antibodies, we performed N-terminal sequence analyses of purified heavy and light chains from four GAC-specific hybridomas produced using GAC-primed A/J mouse spleen cells (19) (Fig. 1). The heavy chain sequences differ from one another by at most four residues and are identical to the sequence of A/J anti-GAC heavy chains purified from hyperimmune sera at all 21 positions that can be compared (36, 37). Two of the H chain sequences, HGAC39 and HGAC40, are identical throughout the first 60 residues despite the fact that these two hybridoma proteins are idiotypically distinct. This is consistent with previous localization of our GAC idiotype to V_κ (19). Interestingly, the HGAC39

Heavy Chains

```
                                          —HV1—              ——HV2——
        1        10        20        30        40        50        60     VK1 GAC
GAC9H   EVKLEESGGGLVQPGGSMKLSCVASGFTFSNYWMCWVRQSPEKGLEWVAEIRLKSDNYVT       +
GAC11H  ..................................H...................GS..A        +
GAC40H  ..................................D.......................A        +
GAC39H  ..................................D.......................A        -
ABPC4   ..................................N....................H..A.       -
```

Light Chains

```
                                ————HV1————                  —HV2—
        1        10        20        30        40        50        60        70
GAC9L   DIVMTQAAPSVPVTPGESVSISCRSSKSLLHSNGNTYLYWYLQRPGQSPQLLIYRMSQLASGVPKRFSNSGSG
GAC11L  .......................................F........................S...Q....TAFTLI
GAC40L  .......................................F.......................M....P....TA
GAC39L  ........F.N..L.T.A.........N......I.F....................N......H...?....
```

FIGURE 1. Amino-terminal sequences of heavy and light chains from four A/J anti-GAC hybridoma proteins. The residues are numbered sequentially from the amino terminus. Presence of the light chain idiotype $V\kappa_1^{GAC}$ is indicated with a "+". The heavy chain sequences are compared with the sequence of the ABPC4 heavy chain, derived from a BALB/c plasmacytoma that binds inulin (38). The positions of first (HV1) and second (HV2) hypervariable regions are marked with bars.

sequence differs at only two positions from the reported sequence of the heavy chain of ABPC4, a BALB/c plasmacytoma that binds inulin (38). Both of these substitutions could result from single nucleotide changes.

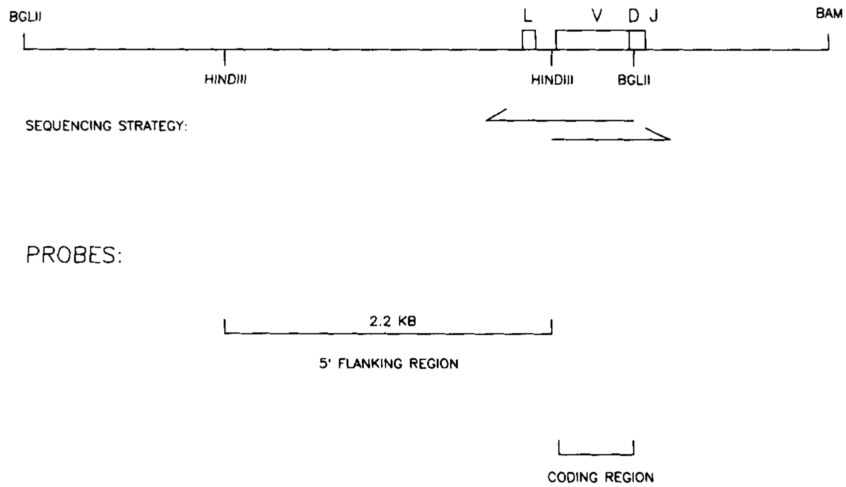
Three idiotype-positive light chains are also closely homologous, differing by only three substitutions out of the 70 positions compared (Fig. 1). One idiotype-negative light chain was examined and found to belong to a separate V_κ subgroup, differing from the idiotype-positive sequences at 10 positions in the first 70 residues, 5 of which represent alterations in the amino-terminal 23 residues.

Cloning of an Anti-GAC V_H Gene. Viewing the protein sequences shown in Fig. 1, it would appear that somatic mutation operating on a single V_H gene segment and two V_κ gene segments could produce much of the diversity in GAC-binding antibodies. In particular, the two to four substitutions observed in the heavy chain sequences are quite consistent with results obtained in BALB/c anti-PC heavy chains where a single V_H gene segment, altered by somatic mutation, encodes the entire repertoire of sequences (5, 6).

To characterize the genes responsible for A/J anti-GAC antibodies, we constructed a library of genomic DNA from the HGAC9 hybridoma in phage lambda and isolated the rearranged heavy chain gene using a BALB/c probe containing all J_H sequences (28, 30). Fig. 2 shows a partial restriction map of the HGAC9 V_H clone that we obtained and the nucleotide sequence of the leader, intron, and the V_H coding region from this clone. The translated amino acid sequence from this clone agrees entirely with the previously obtained protein sequence at all 61 positions available for comparison. The leader sequence, by analogy to other V_H sequences, is 15 residues in length and is followed by a 101-bp intron preceding the V_H coding region. The D segment nucleotide sequence is 15 bp in length and is not a member of any previously defined germline D family (39). This may reflect polymorphism between A/J and BALB/c D regions, high level somatic mutation within this particular rearranged D segment, or the existence of additional heretofore uncharacterized germline D segments in the mouse (13). The HGAC9 heavy chain uses the J_H2 segment that is sequence identical to the previously characterized BALB/c J_H2 segment (40).

Comparison of the translated sequence for the entire V_H region of HGAC9 with the sequence of the BALB/c inulin-binding plasmacytoma heavy chain, ABPC4, reveals only five amino acid substitutions, three of which are within hypervariable regions (Fig. 3). The two sequences differ dramatically in their D segment-encoded regions, and ABPC4 uses the J_H3 segment as opposed to J_H2 in GAC9 (38). Thus BALB/c inulin-binding antibodies and A/J GAC-binding antibodies use very similar V_H gene segments joined to different D and J_H segments and associated with different light chains.

V_HGAC Is a Member of a Small Multigene Family. Since BALB/c anti-inulin and A/J anti-GAC antibodies use such similar V_H sequences, it seemed possible that these different molecules are encoded by allelic V_H gene segments. In fact, comparison of a previously determined BALB/c germline V_H sequence that encodes a protein that differs by one amino acid residue from ABPC4 with the V_HGAC gene yielded 97% homology at the nucleotide level (S. T. Crews and R. M. Perlmutter, unpublished data). This level of homology is consistent with allelic differences in the T15 V_H family when different mouse strains are com-




```
  1    CCCGAATGCATCTTTCCTGATTCATACGACCTGGAACTCTAAGCTCTGACAGAGGACCTAACTCTGGACCTCACA    75
-19                                          M  D  L  R  L  S  C  A  F  I  I  V         -8
 76    GTCTTCCTCAGTGACCAACACGGACACAGACTATTCACCATGGACTTGAGACTGAGCTGTGCTTTTATTATTGTT   150
 -7    L  L  K                                                                          -5
151    CTTTTAAAAGGTAATTCATAGATAATAAGAGATATTGAGTGTGTGAGTGGACGGGAGTGAGAGAACACTGAATAT   225
 -4                                       G  V  Q  S  E  V  K  L  E  E  S  G  G          9
226    TTTGACGGCTTACTGAGAGGGTTCTCTGTGTTTGCAGGGGTCCAGAGTGAAGTGAAGCTTGAGGAGTCTGGAGGA   300
 10    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        34
301    GGCTTGGTGCAACCTGGAGGATCCATGAAACTCTCCTGTGTTGCCTCTGGATTCACTTTCAGTAACTACTGGATG   375
 35    C  W  V  R  Q  S  P  E  K  G  L  E  W  V  A  E  I  R  L  K  S  D  N  Y  V        59
376    TGCTGGGTCCGCCAGTCTCCAGAGAAGGGACTTGAGTGGGTTGCTGAAATTAGATTGAAATCTGATAATTATGTA   450
 60    T  H  Y  A  E  S  V  K  G  R  F  S  I  S  R  D  D  S  K  S  S  V  Y  L  Q        84
451    ACACATTATGCGGAGTCTGTGAAAGGGAGGTTCAGCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCAA   525
 85    M  N  S  L  R  A  E  D  T  G  I  Y  Y  C  T  D  L  G  T  S  Y  W  G  Q  G       109
526    ATGAACAGCTTAAGGGCTGAAGACACTGGAATTTATTACTGCACAGATCTCGGGACAAGCTACTGGGGCCAAGGC   600
110    T  T  L  T  V  S  S                                                             113
601    ACCACTCTCACAGTCTCCTCAGGTGAGTCCTTAAAACCACT
```

FIGURE 2. The structure of a V_H gene encoding an A/J anti-GAC heavy chain. Shown is a partial restriction map of a recombinant phage clone containing the V_H gene from the HGAC9 hybridoma cloned from a genomic library constructed in L47.1 (26, 30). The DNA sequencing strategy, using M13mp8 single-stranded bacteriophage and the dideoxynucleotide technique (31, 32) is also partially outlined, and the complete sequence of the leader, intron, and V_H coding regions is shown below. The protein sequence derived by translation is indicated immediately above the nucleotide sequence and is numbered sequentially with the first codon of the secreted heavy chain denoted number 1. The numbering of the nucleotide sequence is arbitrary. Also shown are the positions of coding region and flanking region probes derived by subcloning of the appropriate restriction fragments into M13mp8.

pared (R. M. Perlmutter, unpublished observation) or might reflect somatic mutation of the V_HGAC gene, and it further suggests that the V_H gene segments that encode inulin-binding heavy chains in BALB/c mice may direct the synthesis of GAC-binding heavy chains in A/J mice. To test this hypothesis, we analyzed genomic DNA by Southern blotting. Fig. 4 shows that a coding region probe derived from the V_HGAC gene (Fig. 2) identifies 10–12 bands estimated to be at least 90% homologous based on stringent washing in A/J, BALB/c, and C57BL/6 DNA cut with *Eco*RI or *Bgl*II, and a small number of bands in similarly digested Lewis rat DNA. The specific gene that is rearranged in the HGAC9

```
                                         VH Region
                              —HVI—               ————————HV2————————
        1        10        20        30        40        50        60        70        80        90
GAC9H   EVKLEESGGGLVQPGGSMKLSCVASGFTFSNYWMCWVRQSPEKGLEWVAEIRLKSDNYVTHYAESVKGRFSISRDDSKSSVYLQMNSLRAEDTGIYYCT
ABPC4   ..................................N....................H..A...........T...............N............

        D Region                                  JH Region
GAC9H   DLGTS                             GAC9H   YWGQGTTLTVSS
ABPC4   TGFA                              ABPC4   ......LV...A
```

FIGURE 3. A comparison of the heavy chains of HGAC9 and ABPC4. The deduced complete amino acid sequence of the HGAC9 heavy chain is aligned with the sequence of ABPC4 (38). The boundaries of the hypervarible regions (HV1 and HV2) are marked with bars. Sequences of V_H, D region, and J_H region encoded segments are shown separately.

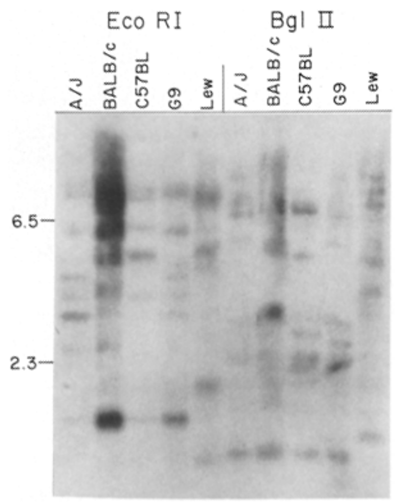


FIGURE 4. Multiple, closely homologous V_H gene segments comprise the V_HGAC gene family. Shown is a Southern blot analysis of 10 μg of sperm or liver DNA derived from A/J, BALB/c, or C57BL/6 mouse strains or from Lewis rats, compared with the pattern obtained using 10 μg of DNA derived from GAC9 hybridoma cells. The DNA samples were digested with either *Eco*RI or *Bgl*II to completion and fragments were separated by electrophoresis in 0.7% agarose. The probe used was the V_HGAC coding region segment shown in Fig. 2. Positions of standard DNA size markers are shown at the left of the figure (in kilobases).

hybridoma is not immediately identifiable. Using a 5′ flanking region probe from the HGAC9 V_H clone (which includes the leader and the intron sequences; see Fig. 2), however, a single germline band is observed in Southern blots of A/J DNA. As shown in Fig. 5, this probe permits the identification of the specific band in Fig. 4 that is rearranged in the HGAC9 hybridoma. The probe identifies two bands in *Eco*RI-digested BALB/c DNA, quite different in size from the band seen in *Eco*RI-digested A/J DNA, and hybridizes strongly with a 6.3-kb *Eco*RI fragment in the HGAC9 DNA that contains the rearranged V_H gene. The Lewis rat contains no sequences which hybridze with the 5′ probe at this stringency.

ABPC4 and V_H Genes Are Not Alleles. Examination of *Eco*RI (not shown) or *Bgl*II-digested DNA from the ABPC4 plasmacytoma demonstrates no rearrangement of the V_HGAC gene(s) defined by the 5′ flanking region probe (Fig. 6). Thus, despite the fact that the V_HGAC and ABPC4 protein sequences differ by only five residues, they are not encoded by allelic gene segments.

Multiple V_H Genes Are Used in the A/J Anti-GAC Response. Since there are at

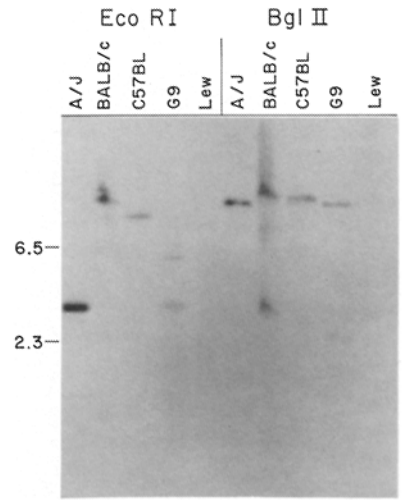


FIGURE 5. The V_HGAC9 gene segment can be defined by a flanking region probe. A Southern blot analysis using *Eco*RI- or *Bgl*II-digested DNA from A/J, BALB/c, or C57BL/6 mouse strains or from Lewis rats or the G9 hybridoma cells is shown with the V_HGAC flanking region probe (Fig. 2). The order of samples is identical to that shown in Fig. 4. Positions of standard size markers in kilobases are again indicated at the left of the figure.

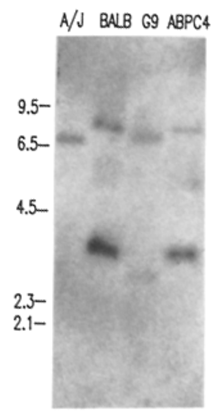


FIGURE 6. The V_HGAC gene and the ABPC4 V_H gene are not alleles. Shown is the pattern obtained when 10 μg of DNA from A/J or BALB/c liver or from GAC9 or ABPC4 cells is digested with *Bgl*II and analyzed by Southern blotting using the V_HGAC flanking region probe (Fig. 2). The positions of standard size markers are indicated in kilobases at the left of the figure.

least 10 germline bands with >90% homology to V_HGAC by Southern blotting in A/J mice (Fig. 4), it remains possible that more than one of these V_H gene segments might contribute to the diversity of the anti-GAC antibody response in these mice. Fig. 7 shows a Southern blot analysis of six A/J hybridomas, two BALB/c hybridomas, and one C57BL/6 hybridoma with specificity for GAC. Only three of the A/J anti-GAC hybridomas show a 6.3-kb rearranged band corresponding to the V_HGAC gene cloned from HGAC9 cells. The remainder of the hybridomas must use other V_H gene segments to generate GAC-binding

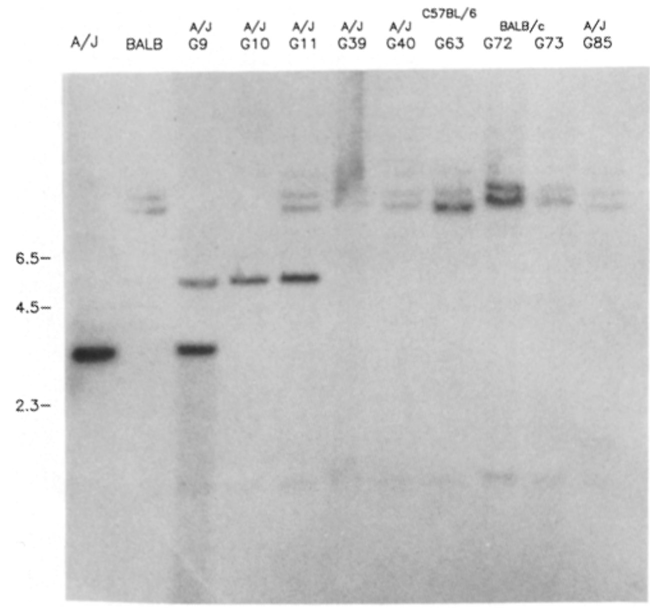


FIGURE 7. Multiple V_H gene segments contribute to the murine anti-GAC antibody repertoire. Shown is a Southern blot analysis of DNA derived from six A/J hybridomas, two BALB/c hybridomas, and one C57BL/6 hybridoma with specificity for GAC or from A/J or BALB/c liver. 10 μg of DNA was digested with *Eco*RI in each case and the V_HGAC flanking region subclone was used as a probe (Fig. 2). The positions of standard DNA size markers in kilobases are shown at the left of the figure.

activity. In particular, HGAC39 and HGAC40, which differ from HGAC9 at only two positions in the amino-terminal 60 residues, must use a distinct but very closely related V_H gene, while HGAC11, which differs at four positions, uses the same V_H gene segment as HGAC9 and probably the same D and J_H segments as well, as defined by additional Southern blots performed using other restriction enzymes (data not shown).

Discussion

Analysis of the molecular genetics of antibodies initially identified three fundamental mechanisms responsible for the generation of antibody diversity: germline repertoire, combinatorial joining of gene segments, and combinatorial association of heavy and light chains (1–4). More recently, it has become clear that a process of somatic hypermutation can operate on fully assembled V_H and V_L genes to amplify the already substantial germline diversity. In the BALB/c antibody response to PC, for example, fully half of the available heavy chain variable region sequences differ from one another and yet all are encoded by a single germline V_H gene segment (5, 6). Similarly, somatic mutation has been shown to operate on the V_κ167 gene segment that encodes several of the light chains which are used in the BALB/c anti-PC response (8–10). Heavy chains of A/J anti-ARS antibodies bearing a cross-reactive idiotype (CRI_A) also appear to be derived from a single germline V_H segment with superimposed somatic mutation (12), and similar results have been suggested for murine antibodies directed against GAT (14) and OX (13). In fact, in each of these systems the dominant mechanism responsible for antibody diversification appears to be somatic mutation. Thus, 70% of the substitutions in BALB/c PC-binding heavy chains reflect the action of the somatic mutation process (R. M. Perlmutter, S. T. Crews, and L. E. Hood, unpublished data).

To characterize further the relative importance of germline repertoire, combinatorial joining, combinatorial association, and somatic mutation mechanisms in the generation of antibody diversity, we have applied protein and DNA sequencing strategies to the analysis of a particularly heterogeneous population of antibodies, those directed against group A streptococcal carbohydrate. Although the individual anti-GAC repertoire of each mouse is limited to one or a few antibody species (15, 16), the typical anti-GAC repertoire of each mouse strain is quite large, comprising at least 200 antibodies distinguishable by IEF or other criteria (17). Since murine anti-GAC antibodies are for the most part restricted to IgM_κ and $IgG3_\kappa$ isotypes, the heterogeneity of these antibodies must reflect differences in variable region structure (18). Curiously, serologic studies have identified strain-specific idiotypes that are present on a large percentage of anti-GAC antibodies. The A5A determinant(s), for example, is associated with 20–60% of A/J anti-GAC antibodies and is inherited as a single Mendelian trait closely linked to immunoglobulin C_H allotype (41). Similar results have been reported with other antiidiotypic reagents in BALB/c (S117) (42) and SWR anti-GAC antibodies (43) and suggest that the diversity in this system is superimposed upon molecules with similar binding site structures. We have previously reported that >50% of A/J anti-GAC antibodies use a specific light chain defined by spectrotypic (16) and idiotypic ($V\kappa_1{}^{GAC}$) (19) criteria, suggesting that much of the binding site diversity of these anti-GAC antibodies results from the pairing of multiple heavy chains with a few light chains. In this report, we present a structural analysis of diversity in anti-GAC antibodies and the V_H genes that encode these antibodies.

Limited Heterogeneity of Anti-GAC Antibodies. The amino-terminal heavy and light chain sequences of HGAC9, HGAC11, and HGAC40 are quite similar, differing from one another by at most 4 of 60 residues for the heavy chain and 3 of 70 residues for the light (Fig. 1), although the IEF patterns of these antibodies are quite different (19). Most of the substitutions in these sequences are confined to hypervariable regions. Thus, in agreement with idiotypic data, the heterogeneity of A/J anti-GAC antibodies reflects substitutions superimposed on common framework heavy and light chain structures.

Structural Basis of $V\kappa_I{}^{GAC}$. Three light chains that bear the $V\kappa_1{}^{GAC}$ determinant(s) differ by only three residues, while a $V\kappa_1{}^{GAC}$-negative light chain differs by 10 residues from the most similar idiotype-positive sequence, including six framework substitutions (Fig. 1). It is likely that the $V\kappa_1{}^{GAC}$ idiotype depends at least in part on the presence of a characteristic framework structure, typified by HGAC9 and closely related to the BALB/c $V_\kappa 27$ subgroup (44, 45). The HGAC39 idiotype-negative light chain sequence resembles sequences of another BALB/c subgroup, $V_\kappa 25$. Herbst et al. (45) have recently reported the complete sequences of two anti-GAC light chains from CXBI mouse hybridomas, which we have aligned with the A/J sequences in Fig. 8. One sequence (7S34.1) differs

```
                                   ----------HV1----------            --HV2--
                1        10        20        30        40        50        60        70
         GAC9L  DIVMTQAAPSVPVTPGESVSISCRSSKSLLHSNGNTYLYWYLQRPGQSPQLLIYRMSQLASGVPKRFSNSGSG
         GAC11L .......................................F........................S...Q....TAFTLI
         GAC40L .......................................F.......................M....P....TA
V_κ27:   7S34.1 ......T...AL...........................F......C.........N.......D...G....TAFTLRI
```

```
                                   ----------HV1----------            --HV2--
                1        10        20        30        40        50        60        70
         GAC39L DIVMTQAAFSNPVTLGTSASISCRSSKNLLHSNGITFLYWYLQRPGQSPQLLIYRMSNLASGVPHRFS?SGSG
V_κ25:   2S1.3  ....................F......S.QQ.K...Y.....K...........Q.........D...G....T
```

FIGURE 8. Two distinct V_κ subgroups contribute to anti-GAC antibodies. The amino-terminal sequences of four A/J anti-GAC hybridoma light chains are compared with the sequences of two previously reported anti-GAC light chain sequences (7S34.1 and 2S1.3, 44) derived from hybridomas produced using GAC-immune spleen cells from CXBI recombinant mice. The sequences fall into two separate subgroups, V_κ25 and V_κ27 (43), as shown. Hypervariable regions (HV1 and HV2) are denoted with bars, The numbering is sequential from the amino terminus.

by seven substitutions from the consensus $V\kappa_1^{GAC}$-positive A/J light chain sequence, all but one of which are framework substitutions. The other sequence (2S1.3) differs by 10 substitutions from the HGAC39 light chain sequence, but 7 of these substitutions are in hypervariable regions. Clearly, similar light chains are used in anti-GAC antibodies from A/J mice and from the CXBI stains, in agreement with the observation that the $V\kappa_1^{GAC}$ idiotype/spectrotype can be found in BALB/c and C57BL/6 mice (16, 19). Attempts to define the V_κ gene segment(s) encoding $V\kappa_1^{GAC}$ are currently under way.

Shared V_H Segments in Anti-GAC and Anti-inulin Antibodies. The HGAC9 V_H segment and the ABPC4 (BALB/c) antiinulin V_H segment sequences differ at only five positions (Fig. 3) and are probably 97% homologous at the DNA level (R. M. Perlmutter and S. T. Crews, unpublished observation). This level of similarity strongly suggests that the same V_H segments that encode anti-GAC antibody heavy chains can also contribute to heavy chains in anti-inulin antibodies. Although V_HGAC and the ABPC4 V_H segments are not allelic (Fig. 6), it seems certain that other A/J GAC-binding antibodies could use a V_H segment that is an allele of the ABPC4 V_H gene segment (see below). This would be analogous to reports of the promiscuous use of a single V_κ gene segment in antibodies directed against ARS or OX (13).

Generation of Diversity in A/J Anti-GAC Antibodies. The heavy chain sequences of A/J anti-GAC antibodies in this report differ by as many as four substitutions in the first 60 residues (Fig. 1). Using a single-copy flanking region probe, we have shown that the two most dissimilar sequences, GAC9H and GAC11H, are encoded by the same germline V_H gene segment (Fig. 7); most likely, the same D and J_H segments are rearranged in these hybridomas as well (Fig. 7 and additional unpublished data). Thus, in anti-GAC antibodies, as previously described in the PC, NP, ARS, and GAT systems, a somatic mutation process, superimposed on a single germline V_H gene segment, acts to amplify the diversity of expressed heavy chains. Interestingly, three of these four substitutions require two nucleotide replacements ($C \rightarrow H$, $S \rightarrow G$, and $D \rightarrow S$) (Fig. 1).

In contrast to the results obtained in PC, ARS, and GAT systems, however, we have shown that some of the heavy chain diversity of anti-GAC antibodies results from the use of more than one V_H gene segment. Using a coding region probe, we estimate that the A/J and BALB/c germline repertoires include at least 10 V_H gene segments >90% homologous to V_HGAC (Fig. 4). The V_HGAC gene segment can be defined using a single-copy flanking region probe (Fig. 5) and is not rearranged in hybridomas HGAC39 and HGAC40 (Fig. 7), which share an identical heavy chain sequence that differs at only two positions from V_HGAC in the first 60 residues (Fig. 1). Only three of nine hybridomas that we examined use the V_HGAC gene segment defined by HGAC9 (Fig.7). Although it is remotely possible that several of the hybridomas that we examined had lost the chromosome encoding the GAC-binding heavy chain, in our experience these hybridomas are quite stable. Thus, the diversity of anti-GAC antibodies reflects a germline repertoire that includes at least two V_κ and at least two V_H gene segments and a superimposed somatic mutation process.

The HGAC9 D segment (GATCTCGGACAAGC), although unrelated to any previously described germline D segments (39), is similar to the D segment commonly observed in heavy chains from BALB/c and DBA/2 hybridomas binding OX (GATCGGGGG) (13). This observation supports the existence of another as yet uncharacterized D segment family in the murine germline.

Evolution of the V_HGAC Gene Family. It is interesting to speculate on the selective forces that maintain a germline family of perhaps 10 V_H segments that are >90% homologous. Unequal recombination events can be expected to result in expansion and contraction of this family over time (46), and indeed the BALB/c genome includes two copies of the V_HGAC gene that is apparently present in only a single copy in A/J and C57BL/6 mice (Fig. 5). This likely reflects gene duplication occurring since the time of strain divergence in mice or a gene deletion event in the ancestors to the A/J and C57BL6 mouse strains. Interestingly, multiple, closely homologous V_HGAC-coding region sequences are maintained in the context of quite divergent leader and intron sequences (R. M. Perlmutter and S. T. Crews, unpublished observation). Maintenance of closely homologous V_H segment sequences despite divergence of surrounding flanking regions may reflect gene conversion events (47). The V_HGAC gene family will provide an interesting proving ground for the study of short-term evolution in multigene families and, because of the large genetic distance between anti-GAC idiotypic markers and C_H allotype (41), should prove useful in the mapping of the murine V_H locus using easily detected restriction enzyme polymorphisms in the 5′ flanking region of V_HGAC (Fig. 5).

Summary

Most mouse strains are able to mount a diverse antibody response against group A streptococcal carbohydrate (GAC). We have previously reported that murine anti-GAC antibodies are for the most part restricted to IgM and IgG3 subclasses. In addition, despite extensive heterogeneity in their isoelectric focusing patterns, >50% of A/J anti-GAC antibodies share a common light chain defined by spectrotypic and idiotypic ($V\kappa_1^{GAC}$) criteria. We have used protein and DNA sequencing strategies to examine the genetic basis of diversity in murine anti-GAC antibodies. In particular, we report that, (*a*) multiple, closely homologous V_H gene segments contribute to the generation of anti-GAC anti-

bodies, (*b*) a common framework sequence, related to the $V_\kappa 27$ subgroup, probably defines $V_{\kappa 1}{}^{GAC}$, and (*c*) the A/J anti-GAC V_H regions and BALB/c anti-inulin V_H sequences are 95% homologous at the protein level and are likely encoded by overlapping V_H gene families. Lastly, we discuss the genetic mechanisms that might permit the evolution of multiple, closely homologous germline V_H gene segments in the context of highly divergent flanking region sequences.

We thank Alexander Shaffer, John Molkie, and Donna Venturini for technical assistance, Dr. Fred Mushinski for the gift of ABPC-4 DNA, Drs. Ellen Kraig, Sandra Chang, Mitchell Kronenberg, and Carol Readhead for critically reviewing the manuscript, and the secretarial staff of the Caltech Biology Division for word processing.

Received for publication 19 September 1983.

References

1. Honjo, T. 1983. Immunoglobulin genes. *Ann. Rev. Immunol.* 1:499.
2. Early, P., and L. Hood. 1981. Immunoglobulin genes. *Genet. Eng.* 3:157.
3. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature (Lond.).* 302:575.
4. Leder, P. 1983. Genetic control of immunoglobulin production. *Hosp. Pract.* 18:73.
5. 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 antibody. *Cell.* 25:59.
6. Gearhart, P., N. Johnson, R. Douglas, and L. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.)* 291:29.
7. Kim, S., M. M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. *Cell.* 27:573.
8. Gearhart, P., and D. Bogenhagen. 1983. Clusters of point mutations are found exclusively around rearranged antibody varible region genes. *Proc. Natl. Acad. Sci. USA.* 80:3439.
9. Gershenfeld, H., A. Tsukamoto, I. L. Weissman, and R. Joho. 1981. Somatic diversification is required to generate the V_κ genes of MOPC511 and MOPC167 myeloma proteins. *Proc. Natl. Acad. Sci. USA.* 78:7674.
10. Selsing, E., and U. Storb. 1981. Somatic mutation of immunoglobulin light chain varible region genes. *Cell.* 25:47.
11. 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 a $\gamma 2a$ variable region. *Cell.* 24:625.
12. 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.
13. Kaartinen, M., G. Griffiths, A. F. Markham, and C. Milstein. 1983. mRNA sequences defined an unusually restricted IgG response to 2-phenyloxazolone and its early diversification. *Nature (Lond.).* 304:320.
14. Rocca-Serra, J., H. W. Matthes, M. Kaartinen, C. Milstein, J. Theze, and M. Fougereau. 1983. Analysis of antibody diversity: V-D-J mRNA nucleotide sequence of four anti-GAT monoclonal antibodies. A paucigene system using alternate D-J recombinations to generate functionally similar hypervariable regions. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:867.

15. Perlmutter, R. M., D. E. Briles, and J. M. Davie. 1977. Complete sharing of light chain spectrotypes by murine IgM and IgG anti-streptococcal antibodies. *J. Immunol.* 118:2161.
16. Perlmutter, R. M., D. E. Briles, J. M. Greve, and J. M. Davie. 1978. Light chain diversity of murine anti-streptococcal antibodies: IgCH-linked effects on L chain expression. *J. Immunol.* 121:149.
17. Briles, D. E., and R. J. Carroll. 1981. A simple method for estimating the probable numbers of different antibodies by examining the repeat frequencies of sequences or isoelectric focusing patterns. *Mol. Immunol.* 18:29.
18. Perlmutter, R. M., D. Hansburg, D. E. Briles, R. A. Nicolotti, and J. M. Davie. 1978. Subclass restriction of murine anti-carbohydrate antibodies. *J. Immunol.* 121:566.
19. Nahm, M. H., B. L. Clevinger, and J. M. Davie. 1982. Monoclonal antibodies to streptococcal group A carbohydrate. I. A dominant idiotypic determinant is located on V_κ. *J. Immunol.* 129:1513.
20. Coligan, G. E., W. C. Schnute, Jr., and T. J. Kindt. 1975. Immunochemical and chemical studies on streptococcal group-specific carbohydrates. *J. Immunol.* 114:1654.
21. Hunkapiller, M. W., and L. E. Hood. 1980. New protein sequenator with increased sensitivity. *Science (Wash. DC).* 207:523.
22. 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.
23. Zimmerman, C. L., E. Appella, and J. J. Pisano. 1977. Rapid analysis of amino acid phenylthiohydantoins by high performance liquid chromatography. *Anal. Biochem.* 77:569.
24. Blin, N., and D. W. Stafford. 1976. A general method for the isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* 3:2303.
25. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.* 284–285.
26. Loenen, W. A. M., and W. J. Brammar. 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. *Gene (Amst.).* 10:249.
27. Hohn, B., and K. Murray. 1977. Packaging of recombinant DNA molecules in bacteriophage particles in vitro. *Proc. Natl. Acad. Sci. USA.* 74:3259.
28. Kraig, E., M. Kronenberg, J. A. Kapp, C. W. Pierce, A. F. Abruzzini, C. M. Sorenson, L. E. Samuelson, R. H. Schwartz, and L. E. Hood. 1983. GAT-specific T and B cells do not transcribe similar heavy chain variable region gene segments. *J. Exp. Med.* 158:192.
29. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. *J. Mol. Biol.* 113:237.
30. Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, and D. Quon. 1978. The isolation of structural genes from libraries of eucaryotic DNA. *Cell.* 15:687.
31. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
32. Anderson, S., M. J. Gart, L. Mayol., and I. G. Young. 1980. A short primer for sequencing DNA cloned into the single-strand phage vector M13mp2. *Nucleic Acids Res.* 8:1731.
33. Smith, G. E., and M. D. Summers. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. *Anal. Biochem.* 109:123.
34. Denhardt, D. T. 1966. A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* 23:641.

35. Mullins, J. I., W. J. Casey, M. O. Nicolson, and N. Davidson. 1980. Sequence organization of feline leukemia virus DNA in infected cells. *Nucleic Acids Res.* 8:3287.
36. Capra, J. D., C. Berek, and K. Eichmann. 1976. Structural studies on induced antibodies with defined idiotypic specificities. III. N-terminal amino acid sequence of the heavy and light chains of mouse anti-streptococcal antibodies, ASA, S8 and S117. *J. Immunol.* 117:7.
37. Perlmutter, R. M. 1978. Combinatorial diversity of anti-carbohydrate antibodies. Ph.D. thesis. Washington University, St. Louis, Missouri.
38. Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin-binding myeloma proteins. *Proc. Natl. Acad. Sci. USA.* 75:1957.
39. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* 155:201.
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. Eichmann, K. 1973. Idiotype expression and the inheritance of mouse antibody clones. *J. Exp. Med.* 137:603.
42. Berek, C., B. A. Taylor, and K. Eichmann. 1976. Genetics of the idiotype of BALB/c myeloma S117: multiple chromosomal loci for V_H genes encoding specificity for group A streptococcal carbohydrate. *J. Exp. Med.* 144:1164.
43. Briles, D. E., and R. M. Krause. 1974. Mouse strain-specific idiotypy and interstrain idiotypic cross reactions. *J. Immunol.* 113:522.
44. Potter, M., J. B. Newell, S. Rudikoff, and E. Haber. 1982. Classification of mouse V_κ groups based on the partial amino acid sequence to the first invariant tryptophan: impact of 14 new sequences from IgG myeloma proteins. *Mol. Immunol.* 19:1619.
45. Herbst, H., J. Y. Chang, R. Aebersold, and D. G. Braun. 1982. Murine V_κ25 isotype sequence, monoclonal antibody 2S1.3 specific for the group A streptococcal polysaccharide. *Hoppe-Seyler's Z. Physiol. Chem.* 363:1069.
46. Hood, L., J. H. Campbell, and S. C. R. Elgin. 1975. The organization, expression, and evolution of antibody genes and other multigene families. *Annu. Rev. Genet.* 9:305.
47. Baltimore, D. 1981. Gene conversion: some implications for immunoglobulin genes. *Cell.* 24:592.