INTER- AND INTRACLONAL DIVERSITY IN THE ANTIBODY RESPONSE TO INFLUENZA HEMAGGLUTININ

By STEPHEN H. CLARKE, KONRAD HUPPI, DIANE RUEZINSKY, LOUIS STAUDT,* WALTER GERHARD,* AND MARTIN WEIGERT

*From the Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111; and the *Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania 19104*

The initial antibody repertoire is determined by multiple heavy and light chain variable region (V_H and V_L)[1] gene segments and multiple joining gene segments (J_κ, J_λ, J_H, and D_H). These gene segments can join in various combinations (1–8) and complete V gene products can associate in different combinations. As a result of combinational joining and association, the size of the preimmune repertoire can in theory be as high as 10^7 different antibodies. This preimmune repertoire is further amplified by junctional diversity that arises by variation at the sites at which gene segments join, and by the addition of nucleotides (N regions) during gene segment joining (9). Somatic mutation may also contribute to the preimmune repertoire, since mutation of the rearranged V gene of a pre-B cell line has been observed (10). Thus the number of different antibodies in an individual may approach the number of preimmune B lymphocytes. As a result, preimmune B lymphocytes of independent origin can be uniquely defined by the nucleotide sequence coding for their antibody V regions, particularly that of their heavy chain diversity region (D_H) and N regions, since these are usually different even in antibodies of the same antigen specificity (11–15). On the other hand, sequence identity in these regions between lymphocytes of a single individual would indicate that they are derived from a single preimmune B lymphocyte.

The genotypes of individual lymphocytes are also unique at the unexpressed antibody loci. These silent alleles are often aberrantly rearranged gene segments (H^-, κ^-, λ^-), and hence can be as diverse as the productively rearranged genes (16). In addition, DNA between the V_κ and J_κ loci of rearranged L chain genes is often retained. The nature of this upstream (U) DNA depends on which V_κ and J_κ gene segments are joined, hence there is considerable variability in the nature of this upstream DNA (17). As these types of rearrangements occur independently of the productive rearrangements, the context of H^-, κ^-, and U DNA can also uniquely define a given lymphocyte.

In this report we describe the V region nucleotide sequence of a number of antihemagglutinin (anti-HA) antibodies generated from BALB/c mice after a

This work was supported by grants GM-20964-10, AI-13989, CA-06927, and RR-05539 from the National Institutes of Health, by the American Cancer Society Postdoctoral Fellowship PF-2132, and by NIH Postdoctoral Training Grant CA-09035.

[1] *Abbreviations used in this paper:* CDR, complementarity-determining region; D region, diversity region; DTT, dithiothreitol; H chain, heavy chain; HA, hemagglutinin; J region, joining region; L chain, light chain; N regions, nucleotide regions; SDS, sodium dodecyl sulfate; U, upstream.

secondary immunization with influenza virus A/PR/8/34 (PR8); we also identify the unexpressed V gene rearrangements in these hybridomas. We thereby can distinguish examples originating from different lymphocyte clones and identify examples originating from the same lymphocyte clone, generated in vivo. Interclonal comparisons measure the contribution of combinatorial joining, junctional diversity, and combinatorial association (V_H, V_κ) to diversity in this response. Intraclonal comparisons reveal the effects of a high rate of somatic mutation at both expressed V_H and V_L genes.

Materials and Methods

Antihemagglutinin Hybridomas. Hybridomas from mouse H36 and H37 were derived by fusion of spleen cells with Sp2/0-Ag14. Mice were immunized intraperitoneally 24 d before fusion and again, intravenously, 3 d before fusion, using 1,000 hemagglutinating units of influenza virus A/PR/8/34 (H1N1) each time (18).

Genomic Blot Hybridizations. Hybridoma and BALB/c liver DNA were prepared as described by Perry et al. (19). DNA were analyzed according to the method of Southern (20). 12 μg of DNA were digested with Eco RI (Bethesda Research Laboratories, Gaithersberg, MD), electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose (Schleicher and Schuell, Inc., Keene, NH), and hybridized with ^{32}P-labeled probes at 65°C. The probes were plasmid clones pJ11 (Fig. 5) (21), IVS (16), and PRI (17). pJ11 was used to identify H chain rearrangements, IVS to identify κ chain rearrangements, and PRI was used to identify U rearrangements. After a final wash in 30 mM NaCl/3.0 mM sodium citrate/0.1% sodium dodecyl sulfate (SDS) for 1 h at 65°C, filters were exposed to x-ray film at −70°C.

Isolation of Recombinant Clones. 300 μg of DNA from hybridoma H37–77 was digested with EcoRI. Digested DNA was loaded into trough of a 0.7% agarose gel and electrophoresed. Fractions of DNA were collected from a second trough located 10 cm from the origin. Fractions containing the 2.8 kb rearranged J_H fragment (unpublished observation) were identified by Southern blot analysis using pJ11 as a probe. DNA from the appropriate fraction was cloned into the EcoRI site of λgtWES (22, 23). Recombinant phage were identified by hybridization to pJ11 (24) and plaque purified. A restriction map was made by end labeling the EcoRI insert with α-[^{32}P]dATP with the Klenow fragment of PolI (Boehringer Mannheim Biochemicals, Indianapolis, IN), digested with various restriction endonucleases, and electrophoresed on 5% polyacrylamide gels. Appropriate fragments were cloned into M13mp8 or M13mp9 and sequenced by the dideoxy method (25).

RNA Preparation. RNA was prepared from cells grown in tissue culture or from subcutaneous tumors. ~3–5 × 10^8 cells were grown in Dulbecco's modified Eagle's medium (Gibco Laboratories, Grand Island, NY) with 10% fetal calf serum. Cells were washed twice in phosphate-buffered saline and resuspended in 10 ml of RSB solution (0.01 M NaCl, 3 mM $MgCl_2$, 0.01 M Tris, pH 7.4). 50 μl of Triton X-100 was added and the cells vortexed for 15 s. Cells were lysed by douncing and the nuclei and cell debris removed by centrifugation at 500 *g* for 10 min at 4°C. 2 ml of a solution containing 0.5 M NaCl, 50 μM EDTA, 10 m Tris, pH 7.4, and 2.5% SDS were added to the supernatant and the supernatant extracted exhaustively with chloropane (chloroform/phenol, 1:1 vol/vol, 0.05% 8-hydroxyquinoline, saturated with 0.01 M sodium acetate, 0.1 M NaCl and 1 mM EDTA). The RNA was then ethanol precipitated.

RNA was isolated from subcutaneous tumors grown in BALB/c × NZB F_1 mice as follows: Tumor tissue was rinsed in cold 0.9% NaCl, minced into small pieces, and transferred to 10 ml of cold 5% citric acid. The tissue was homogenized until an even suspension was obtained, and the suspension was passed through eight layers of cheesecloth. The cell debris was removed by centrifugation at 500 *g* for 10 min at 4°C. Supernatants were transferred to a polycarbonate tube (Beckman Instruments, Inc., Fullerton, CA) and centrifuged at 25,000 rpm for 2 h at 4°C. The pellet was washed with 0.01 M sodium acetate, pH 5, 0.05 M NaCl and dissolved in 25 ml of 0.01 M Tris, pH

7.4, 0.1 M NaCl, 1 mM EDTA, and 1.5% SDS. After a series of three chloropane extractions the RNA was ethanol precipitated.

Poly(A)$^+$ RNA was prepared as follows: RNA prepared from tissue culture cells or subcutaneous tumors was removed from ethanol by centrifugation and resuspended in 2 ml of 1 mM EDTA, 10 mM Tris, pH 7.6, 0.1% SDS. The RNA was heat-shocked at 68°C for 5 min, cooled to room temperature, and NaCl was added to a final concentration of 0.2 M. RNA solutions were passed through oligo(dT) Sepharose columns equilibrated in 0.2 M NaCl, 1 mM EDTA, 10 mM Tris, pH 7.6, 0.1% SDS. Poly(A)$^+$ RNA was eluted with 1 mM EDTA, 10 mM Tris, pH 7.6, 0.1% SDS, ethanol-precipitated overnight at −20°C, resuspended in sterile H_2O at 1 mg/ml, and stored at −70°C.

RNA Sequencing. RNA sequencing was carried out by extension of synthetic oligonucleotides homologous to κ or γ constant regions. The κ primer (5′-TGGATGGTGGGAAGATG-3′), the γ-crossreactive primer (5′-GGGGCCAGTGGATAGAC-3′), specific for γ1, γ2a, and γ2b H chain mRNA, and the γ3 primer (5′AGGGACCAAGGGATAGAC-3′) were made by the phosphotriester method (26).

Dideoxy sequencing was carried out using a modification of the procedure reported by Air (27). Each base reaction was done separately in a 5 μl volume containing100 mM Tris, pH 8.3, 140 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT). Cold dCTP was added to each reaction to a concentration of 3 μM and α-[^{32}P]dCTP was added to a concentration of 2.4 μM (670 Ci/mmol; New England Nuclear, Boston, MA). G-specific reactions contained 50 μM dATP, 50 μM dTTP, and 10 μM dGTP; A-specific reactions contained 50 μM dGTP, 50 μM dTTP, and 10 μM ATP; C-specific reactions contained 50 μM dATP, dGTP, and dTTP; T-specific reactions contained 50 μM dATP, 50 μM dGTP, and 10 μM dTTP (P-L Biochemicals Inc., Milwaukee, WI). In addition, each reaction contained 0.3 μM of the appropriate dideoxynucleotide (P-L Biochemicals Inc.). ~2.5 μg (2.5 μl) of poly(A)$^+$ mRNA and 10 ng of primer (0.5 μl) were mixed, boiled for 1 min, and quick-cooled on ice water. 2 μl of RNasin (20–30 U/ml; Promega Biotec, Madison, WI) and 1.5 μl of AMV reverse transcriptase (12–15 U/μl; Life Sciences, Inc., St. Petersburg, FL) were added and the mixture aliquoted equally into each of the four nucleotide mixes. After 30 min at 37°C, 1 μl of 1 mM dATP, dGTP, dCTP, and dTTP was added to each and the reaction mixtures returned to 37°C for 10 min. The reactions were stopped by the addition of an equal volume of 98% formamide, 0.3% xylene cyanol, 0.3% bromophenol blue. Samples were heat-shocked at 90°C and 3 μl were loaded on 80 cm, 6% polyacrylamide gradient gels (28). The first 50 bases were resolved by a second sequence reaction containing a fourfold increase in the dideoxynucleotide concentration. Any uncertainties in the sequence ladder were resolved by carrying out a third sequencing reaction using α-^{35}S-dATP. For this method cold dCTP and α-[^{32}P]dCTP were eliminated and replaced with 0.05 μm dATP and 4 μM α-^{35}S-dATP (500 Ci/mmol; New England Nuclear). Deoxynucleotide and dideoxynucleotide concentrations were the same except that dATP was replaced with dCTP.

A second method of obtaining H and L chain mRNA sequences, as described by Caton et al. (A. Caton, G. Brownlee, L. Staudt, and W. Gerhard, manuscript in preparation), was also used. H and L chain mRNA was completely extended using these oligonucleotides as primers. 40 ng of oligonucleotide were labeled with 10–20 U polynucleotide kinase (Boehringer Mannheim Biochemicals) and 13 μM γ-[^{32}P]dATP (5,000 Ci/mmol; Amersham Corp., Arlington Heights, IL) in 100 mM Tris, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, and 0.1 mM EDTA. After 1 h at 37°C the reaction was stopped by boiling for 5 min and quick-cooled on ice water. ~10–40 μg of poly(A)$^+$ mRNA was added to the labeled primer, bringing the volume to 25.5 μl. Solutions were boiled for 1 min and quick-cooled in ice water. Reverse transcriptase buffer (12.5 μl) containing 400 mM Tris, pH 8.3, 560 mM KCl, 40 mM $MgCl_2$ and 40 mM DTT, 5 μl of 5 mM dATP, dGTP, dCTP, and dTTP, 4 μl of RNasin (20 U/μl) and 3 μl of reverse transcriptase (12–15 U/μl) were added to the RNA solution. After 2 h at 37°C the mRNA was digested with 25 μl RNase (0.5 mg/ml in 50 mM EDTA) at 37°C for 30 min. The cDNA were extracted with phenol/chloroform (1:1 vol/vol, saturated with 10 mM Tris, pH 7.2, 10 mM EDTA, 10 mM NaCl), ethanol-precipitated, and electrophoresed on a 5% polyacrylamide, 7 M

urea gel for 3 h at 1,200 V. H and L chain cDNA transcripts were identified by exposure to film at room temperature. The cDNA were eluted from the appropriate gel slices in 2 M ammonium acetate by shaking overnight at 37°C. The cDNA were ethanol precipitated and washed extensively before we carried out Maxam and Gilbert (29) sequencing reactions.

Results

V_H and V_L Sequences. The complete V_H nucleic acid sequences of 10 antibodies specific for the Sb antigenic site of HA are shown in Figs. 1, 2, and 3. These V_H regions fall into four sets, each coded for by a unique combination of V_H, D_H, and J_H gene segments, as well as N region sequence. The V_κ regions of members of a set are also coded for by the same V_κ-J_κ gene segment combination (30 and Fig. 4). The set that includes hybridomas H36-1, 4, 15, and 18 expresses a V_H gene of the 7183 family (Fig. 1). The 7183 V_H family is one of seven families defined in the mouse and consists of 10–12 V_H genes (31). This V_H gene is assigned to the 7183 family based on having >80% sequence homology to the known members of this family. mRNA dot blot analysis supports this assignment, as mRNA isolated from these hybridomas hybridizes to a V_H gene from the 7183 family and not to V_H genes from other families (data not shown). The V_H gene from each member of this set has rearranged to the DFL16.2 (7) and J_H2 gene segments with identical V_H-D_H and D_H-J_H junctions. There are two differences at the D_H-J_H boundary (positions 311 and 313) compared with the germline DFL16.2 and J_H2 sequences. These substitutions may have been generated by the random insertion of bases during D_H-J_H rearrangement or they may have been generated by somatic mutation after rearrangement.

Hybridomas H36-5, 7, and 17 express a V_H gene of the S107 family (Fig. 2) (31, 32). This assignment is based on their having >80% sequence homology to the genes of this family. The expressed V_H gene of this family hybridizes in an mRNA dot blot analysis with a gene from this family but not to genes of other V_H families (data not shown). Of the four V_H genes in the S107 family (32), the expressed gene in these hybridomas is most homologous (94–97%) to V_H-V11. Thus, the hybridomas of this set express mutants of the V11 gene. This V_H gene has rearranged to a D gene closest in homology to the DQ52 gene (6) and to the J_H1 gene segment in each hybridoma. In addition, all three hybridomas have identical V_H-D_H and D_H-J_H junctions. The DQ52 germline gene segment consists of 10 nucleotides; however, in these examples only five nucleotides are spliced to the V and J genes, and the second codon of D is a hybrid codon formed by both the D and J genes.

The three hybridomas from mouse H37 can be assigned to two sets, based on both H and L chain gene usage. One set includes hybridomas H37–77 and 63. Both have rearranged the same V_H gene (Fig. 3). This gene is 90% homologous to the V_H gene rearranged in hybridomas H36-1, 4, 15, and 18, and the cloned H37-77 V_H gene, when used as a probe in a Southern blot analysis of BALB/c DNA, hybridizes to restriction fragments of sizes characteristic of the 7183 V_H family (data not shown). It is unlikely that the expressed V_H gene from hybridomas H36-1, 4, 15, and 18 and the expressed gene in H37-77 and 63 are the same, since the size of the productive J_H rearrangements in these hybridomas (as determined by Southern blot analyses) shows a difference in the location of the

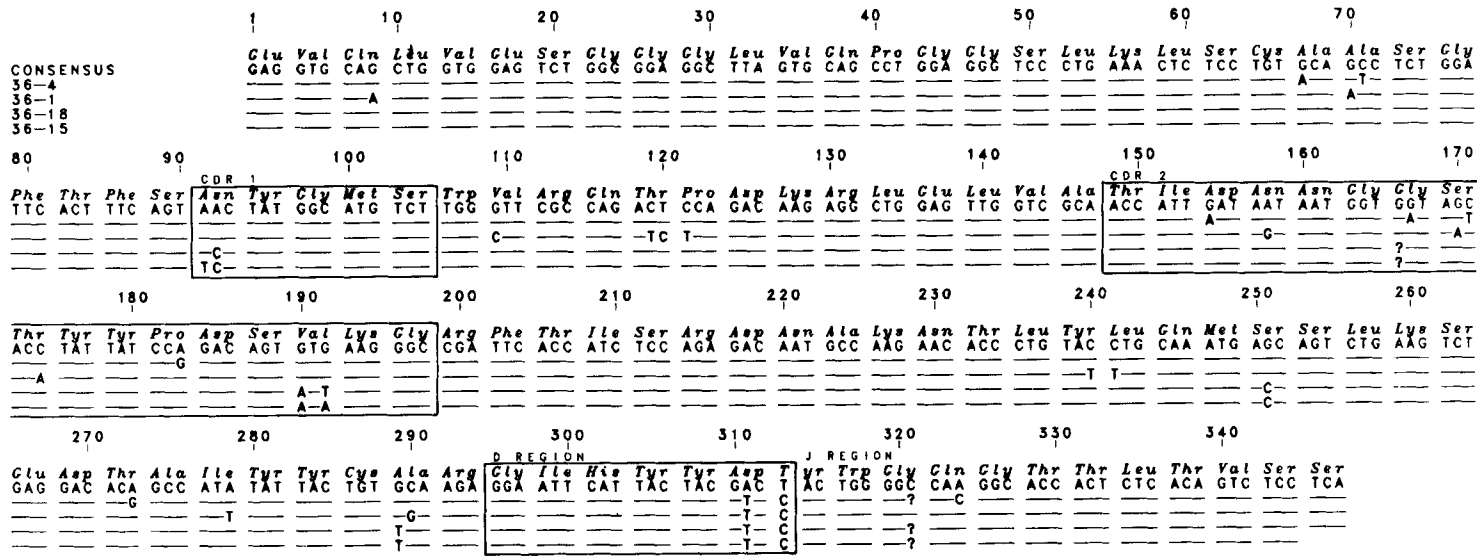


FIGURE 1. Comparison of V_H mRNA sequences from hybridomas H36-1, 4, 15, and 18. V_H regions are compared with a consensus V_H sequence, the germline DFL16.2 sequence, and the germline J_H2 sequence. The consensus V_H sequence is derived from the V_H sequences of these hybridomas in such a way as to require the fewest number of substitutions among these four V_H genes. At several positions (92, 190, 251, 289), these four hybridomas have divided equally between two bases. In these cases, the base shared by hybridomas H36-1 and H36-4 has been assigned to the consensus sequence. According to the genealogic tree derived from V_κ amino acid substitutions (30), H36-1 and H36-4 cannot share mutations that would not also be shared by H36-15 and H36-18 without requiring a second mutation at the same position. However, H36-15 and H36-18 can share mutations that are not also shared by H36-1 and H36-4 without requiring additional mutations at the same position. CDR1, CDR2, and the D_H segment are enclosed in boxes. Question marks indicate an unassigned base at that position.

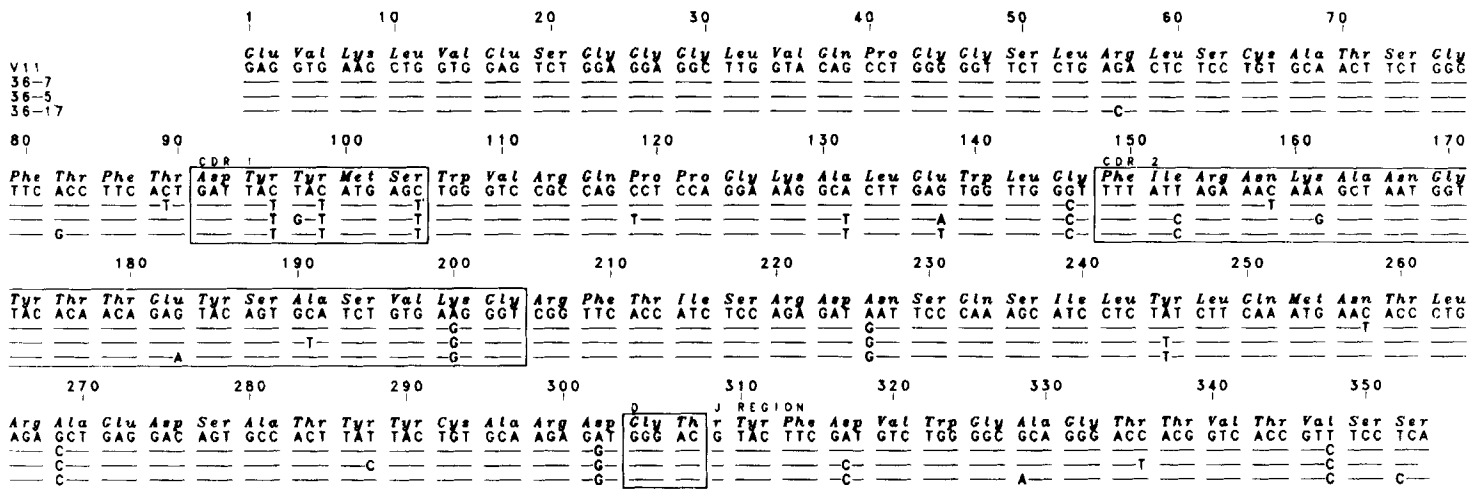


FIGURE 2. Comparison of V_H mRNA sequences from hybridomas H36-5, 7, and 17 to the germline V11, DQ52, and J_H1 sequences. The germline V11, DQ52 and J_H1 sequences have been spliced to create the same rearrangement found in these three hybridomas and serves as the prototype sequence. The J_H1 sequence taken from Sakano et al. (5) identifies a T at position 348, while Newell et al. (41) report a C at this position. If C is the correct nucleotide, then there is no substitution at position 348 in these hybridomas as indicated, thereby reducing by one the total number of V_H mutations in each hybridoma reported in Fig. 6. CDR1, CDR2, and the D_H segment are enclosed in boxes.

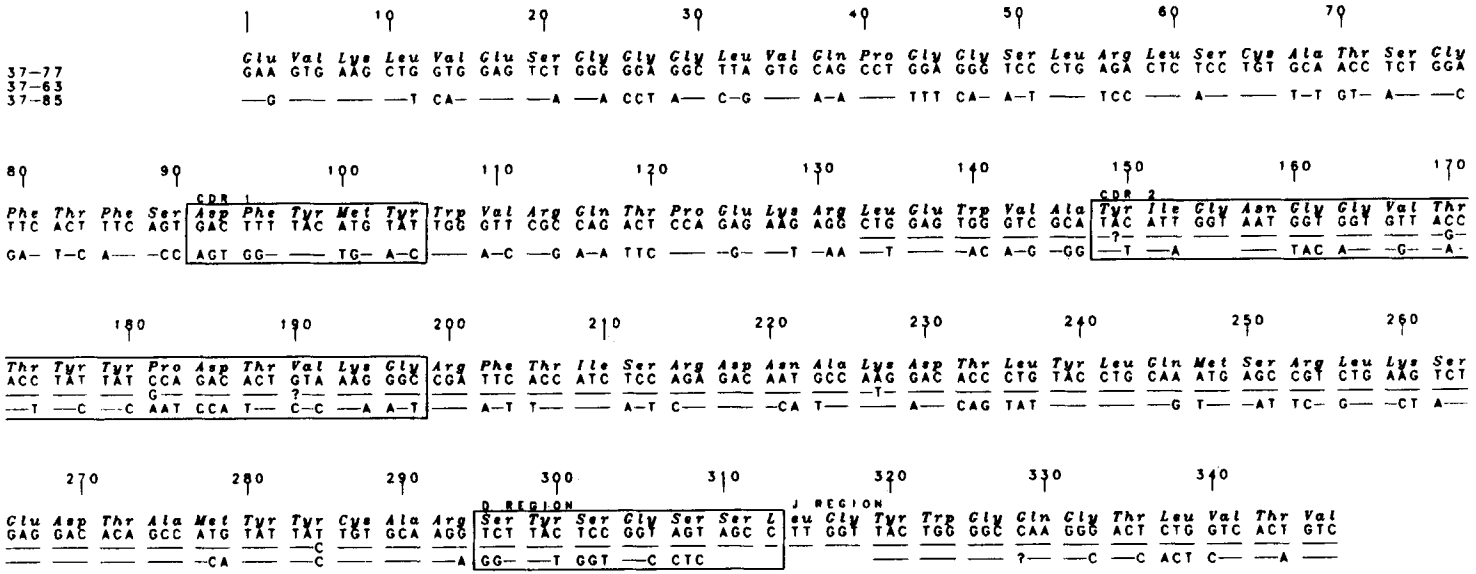


FIGURE 3. Comparison of V_H mRNA sequences from hybridomas H37-77, H37-63, and H37-85. H37-77 and H37-63 have rearranged a V_H from the 7183 V_H family to the DFL16.1 and J_H3 genes. H37-85 has rearranged a V_H from the 3660 V_H family to a D_H of unknown germline origin and to the J_H2 gene. CDR1, CDR2, and the D_H segment are enclosed in boxes.

TABLE I
H and L Chain Gene Rearrangements

	H^+	H^-	κ^+	κ^-	Upstream	
H36-5	2.9 kb		9.5 kb	3.8 kb		
H36-7	4.9 kb	2.4 kb	9.5 kb			
H36-17	4.9 kb		9.5 kb			
H36-15	2.3 kb	5.0 kb	9.5 kb	3.7 kb	12.0 kb	
H36-18	2.3 kb	5.0 kb	9.5 kb	3.65 kb	12.0 kb	
H36-1	2.3 kb	5.0 kb	9.5 kb	3.7 kb	12.0 kb	
H36-4	2.1 kb	5.0 kb	9.5 kb	3.7 kb	12.0 kb	
H37-63	2.8 kb		1.7 kb	5.6 kb	16.5 kb	7.8 kb
H37-77	2.8 kb		1.7 kb	5.6 kb	16.5 kb	7.8 kb
H37-85	2.4 kb				4.3 kb	

The J_H rearrangements (H^+, H^-), J_κ rearrangements (κ^+, κ^-), and upstream rearrangements were detected by Southern blot hybridization using the probes described (Materials and Methods). The H chain rearrangement for all hybridomas were detected using EcoRI-digested hybridoma DNA. L chain rearrangements for the H36 hybridomas were detected using HindIII-digested hybridoma DNA (Huppi, manuscript in preparation), whereas L chain rearrangements for the H37 hybridomas were detected using HindIII/BamHI-digested hybridoma DNA. All upstream rearrangements were detected after BamHI digestion of hybridoma DNA. Brackets indicate clonally related hybridomas.

EcoRI site 5′ of the rearranged V_H genes (Table I). Therefore, at least two members of the 7183 family are used in the response to the Sb site of HA. Both H37-77 and 63 have rearranged the expressed V_H gene to the DFL16.2 and J_H3 gene segments.

H37-85 expresses a V_H gene with only 62–65% homology to the V_H genes of the S107 and 7183 V_H families (Fig. 3). It is closest in homology (96.5%) to the V_H gene used in the BALB/c antiarsonate response (1210.7 V_H) (33). This BALB/c antiArsonate V_H gene hybridizes in a Southern blot analysis to a set of restriction fragments whose sizes are characteristic of the 3660 V_H family (31, 33). Thus, a third V_H family can contribute to the anti-HA (Sb) response. This V_H gene has rearranged to a D_H gene with ≤40% homology to any known germline D_H gene and to the J_H2 gene segment. This D_H segment may be coded for by an as yet unidentified germline D_H gene or it may have been derived from a known germline gene, such as DSP2 (7), that has been altered by the addition of nucleotides and somatic mutation.

All seven H36 hybridoma antibodies express V_κ21C-J_κ2 L chains. H37-77, 63, and 85 express V_κ genes that are 97–99% homologous to the V_κ21C germline gene sequence (34) (Fig. 4) and encode amino acids that distinguish V_κ21C from all other V_κ21 subgroups (35, 36). This gene has rearranged to J_κ4 in H37-85 and to J_κ5 in H37-77 and 63. The V_κ21C-J_κ5 junction in H37-77 and 63 is unusual in that at least two and possibly as many as five extra nucleotides are contributed by the V_κ21C germline gene to create a proline codon not seen in any other V_κ21C sequence.

Genetic Rearrangements at H and L Chain Loci. The nature of rearranged DNA at both expressed and unexpressed loci is for the most part consistent with the idea that these 10 hybridomas originated from four different precursors (see Fig. 6). A previous study (30) suggested that four hybridomas from mouse H36

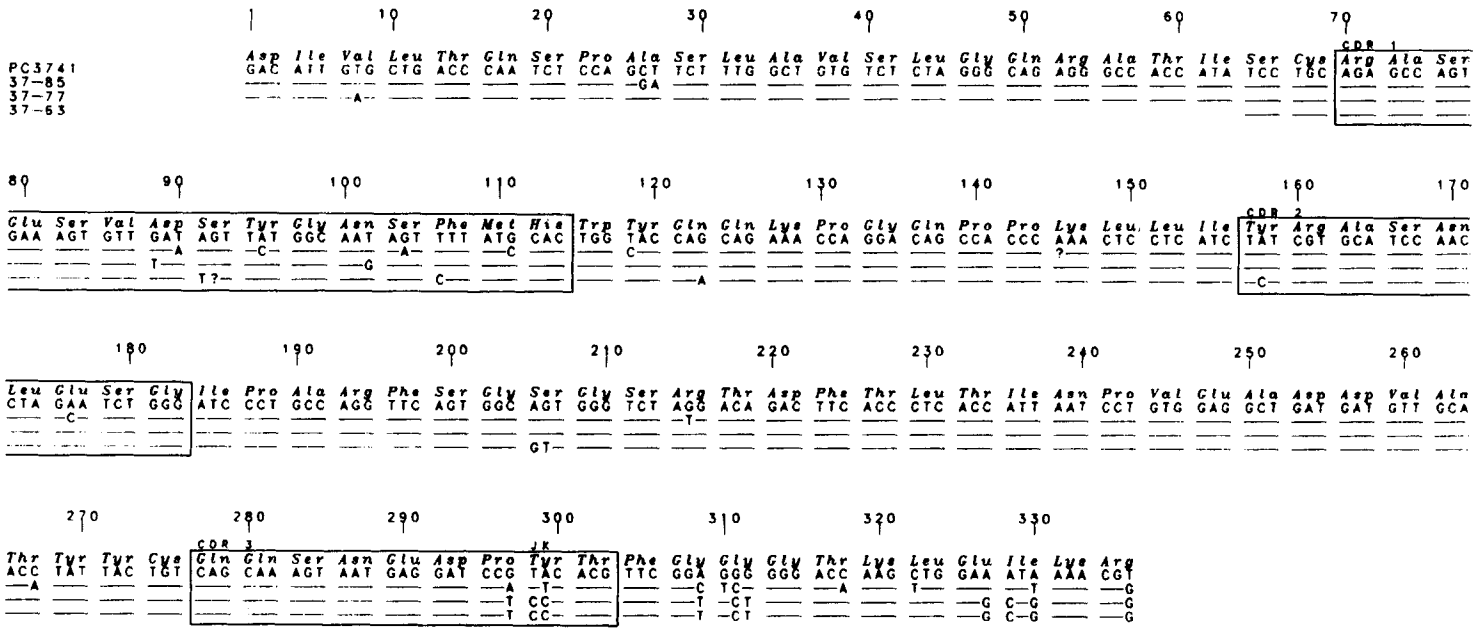


FIGURE 4. Comparison of $V_\kappa 21C$ mRNA sequences from hybridomas H37-85, H37-77, and H37-63 to a germline $V_\kappa 21C$-$J_\kappa 2$ DNA sequence from plasmacytoma PC3741 (Huppi, manuscript in preparation). H37-85 has rearranged to the $J_\kappa 4$ gene, and H37-77 and 63 have rearranged to the $J_\kappa 5$ gene. CDRs are enclosed in boxes.

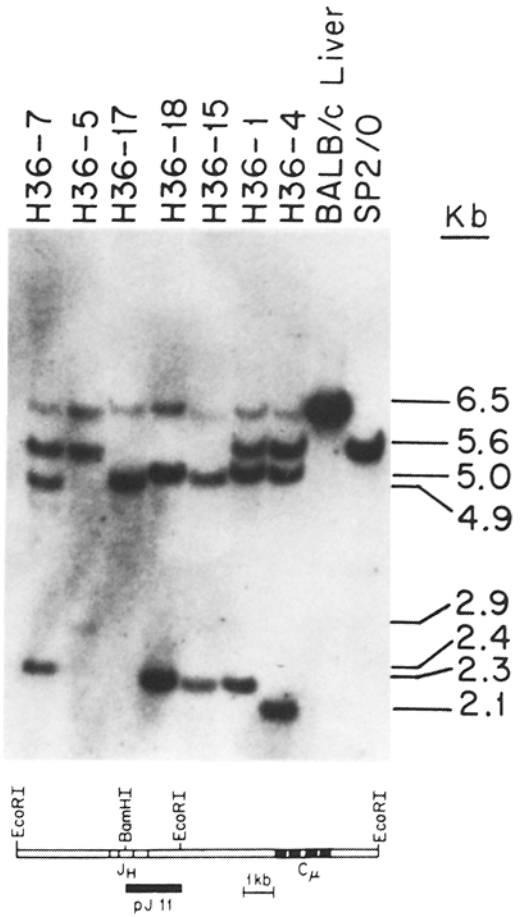


FIGURE 5. Southern blot hybridization of a J_H probe, pJ11, to DNA from the seven V_κ21C-associated hybridomas of mouse H36, BALB/c liver DNA, and Sp2/0-Ag14 DNA after digestion with EcoR1. All hybridomas share a 6.5 kb band that corresponds to an unrearranged J_H locus, which may be derived from either the hybridoma or from contaminating host tissue. In addition, some hybridomas have a 5.6 kb rearrangement derived from the fusion partner Sp2/0-Ag14. H36-5 DNA shows a faint band at 2.9 kb.

(H36-1, 4, 15, and 18) are derived from a common lymphocyte precursor. This conclusion was based on shared κ^+, κ^-, and U rearrangements (Table I). The remaining three hybridomas from this mouse (H36-5, 7, and 17) share certain V_κ amino acid substitutions (30) but, as they lack κ^-, and U rearrangements (Table I), their relationship to each other and to the other four H36 hybridomas was uncertain. Of the three H37 hybridomas, H37-63 and H37-77 appear to be clonally related, since they share a κ^+, κ^-, and two U rearrangements (Table I). The 1.7 kb J_κ rearrangement is the expected size for the V_κ21C-J_κ5 joining and likely represents the κ^+ allele. The 5.6 kb band must therefore represent a shared κ^- allele.

J_H gene rearrangements were examined for all 10 hybridomas (Fig. 5 and Table I). Hybridomas H36-4, 1, 15, and 18 each have two rearrangements. All four share a 5.0 kb rearrangement, three share a 2.3 kb rearrangement, and

H36-4 has a unique rearrangement of 2.1 kb. The 5.0 kb rearrangement was cloned into the EcoRI site of λgtWES and used as a probe in a Southern blot of EcoRI-digested BALB/c liver DNA. This DNA clone hybridizes to a set of restriction fragments characteristic of the DFL16/SP2 families of D genes (data not shown), indicating that the 5.0 kb J_H rearrangement shared by H36-1, 4, 15, and 18 is a nonproductive D-J rearrangement, thus supporting the clonal relatedness of these four hybridomas. The 2.3 kb rearrangement of H36-1, 15, and 18 and the 2.1 kb rearrangement of H36-4 must therefore be the productive V_H-D_H-J_H rearrangements.

H36-17 and 5 each have just one hybridoma-specific J_H rearrangement (4.9 kb and 2.9 kb, respectively; Fig. 5 and Table I). Therefore, these must be productive rearrangements. H36-7 also has a 4.9 kb J_H rearrangement. As this rearrangement is shared by H36-17, it is probably the productively rearranged V_H in this hybridoma. In addition, H36-7 has a second rearrangement of 2.4 kb, which must be nonproductive. The hybridomas of mouse H37 also share rearrangements at the J_H locus (Table I). H37-77 and H37-63 each have a single rearrangement of 2.8 kb, while hybridoma H37-85 has a unique rearrangement of 2.4 kb. Since these are the only rearrangements in these hybridomas, each must represent a productively rearranged V_H gene.

Several unique rearrangements among members of a set have been found. For instance, hybridomas H36-1, 15, and 18 share a 2.3 kb productive H chain gene rearrangement while H36-4 has a 2.1 kb productive rearrangement (Table I). The reason for the differences in size of these restriction fragments is not certain. Southern blot analysis of HindIII-digested DNA hybridized with pJ11 showed the same difference in size between the productive rearrangements, suggesting that it is not due to point mutation but more likely to either a deletion of DNA from H36-4 or an insertion of DNA into H36-1, 15, and 18, before their divergence. A similar pattern exists among the second set of hybridomas from mouse H36. H36-7 and 17 share a 4.9 kb productive rearrangement, while H36-5 has a 2.9 kb productive rearrangement (Table I). This also may be explained by insertion, deletion, or point mutation. These three hybridomas also differ at their nonproductive loci. The lack of H^- and κ^- rearrangements in two of these three hybridomas (Table I) can be explained by chromosome loss either before or after cell fusion. Alternatively, lymphocytes may continue to rearrange their H and L chain gene loci after productive rearrangements have occurred, thereby accounting for unshared rearrangements.

Discussion

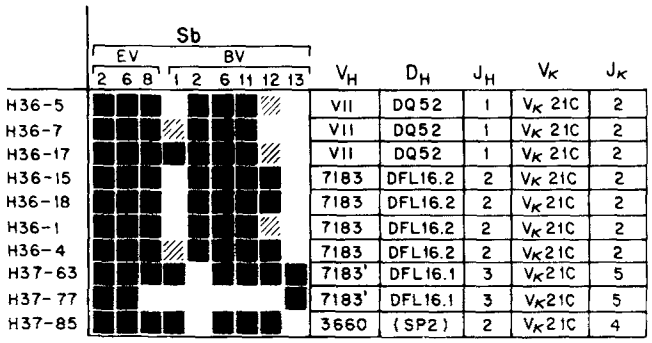
The secondary antibody response to the infleunza HA molecule is highly diverse. It is estimated (18) that the BALB/c mouse has the potential of producing at least 500 different antibodies to any of the four major antigenic sites on the HA molecule. To understand the genetic basis of this diversity, we have analyzed the V region sequences of hybridomas secreting HA-specific antibodies, focusing on examples directed to the antigenic site Sb [HA (Sb)]. This survey of diversity includes analysis of the contribution of germline genes, combinatorial joining, and somatic mutation. Germline diversity is assessed by the V region differences found between members of different clones of lymphocytes; somatic mutation is

assessed by differences between members of the same clone of lymphocytes. Hybridomas from different individuals, i.e., mouse H36 or H37, are one source of examples of independent origin, while hybridomas from a single mouse may include examples of the same or different clonal origin. We are able to determine whether hybridomas from a single individual have originated from the same or different precursors on the basis of V_H region sequence. Because of the considerable flexibility in forming a complete V_H region gene, lymphocytes that have joined the same V_H, D_H, and J_H gene segments and form identical junctions must be clonally related. Since these rearrangements are fixed, lymphocytes that differ in these respects must be of separate clonal origin.

Interclonal Diversity Among HA (Sb) Antibodies. The V_H and V_κ loci of the mouse are highly complex. Both loci are estimated to consist of a few hundred genes. These genes fall into families in which the average nucleic acid sequence homology between families is ~60% (31). The homology of genes within a family is much higher, averaging >85%. The size of families varies from ~40–50 genes (the V_H-J558 family) to just one (such as V_κ167 [37] or $V_{\lambda 1}$ and $V_{\lambda 2}$ [38]). These germline complements of V genes provide an important source of diversity, one that is further amplified by the combinatorial joining of a given V gene with different D_H, J_H, or J_κ gene segments. The interclonal comparison of HA (Sb) antibodies reveals the extent of germline diversification. These antibodies have V_H regions drawn from three different families (designated S107, 7183, and 3660 [31]). Different D_H genes and J_H genes are also represented within these V_H regions, and the junctions of these gene segments show both flexibility of joining and apparent nucleotide additions (N sequences; Fig. 1, 2, and 3). Different V_κ gene products are also found among HA (Sb) antibodies. The V_κ21C subgroup, other V_κ21 subgroups, and at least one other V_κ group (V_κ8) are represented (18 and S. Clarke, W. Gerhard, and M. Weigert, manuscript in preparation).

To assess the contribution of combinatorial joining, one must compare examples that share a particular V gene segment. In this survey all examples express the product of a single V_κ gene, V_κ21C. Combinatorial joining does add to the diversity of these antibodies; in this sample V_κ21C has joined to either J_κ-2, J_κ-4, or J_κ-5 (Fig. 6). Combinatorial joining also diversifies V_H regions. Primary and secondary HA (Sb) antibodies are frequently encoded by the V_H-V11 gene, which has been found to be joined to a variety of D_H gene segments and two different J_H gene segments (Fig. 2 and Clarke, manuscript in preparation). A further amplification of germline diversity arises from the combinatorial association of different V_H and V_L regions. This is shown in the comparison of the V_κ21C-associated HA (Sb) antibodies (Fig. 6). These V_κ21C L chains are paired with H chains from the three V_H families described above.

The effect of germline diversity on the specificity of these antibodies can be evaluated. These monoclonal antibodies have been tested for binding to mutants of the PR8 virus (18), each mutant differing from PR8 by a single amino acid substitution (39). This analysis provides a reactivity pattern for each antibody from which one can infer the site of binding. Further, as these antibodies usually show reduced binding to different mutant viruses, these reactivity patterns also define the fine specificity of a given antibody. The extent to which sequence differences affect specificity and fine specificity is obviously limited by the assay.



	V_H	D_H	J_H	V_K	J_K
H36-5	VII	DQ52	1	V_K 21C	2
H36-7	VII	DQ52	1	V_K 21C	2
H36-17	VII	DQ52	1	V_K 21C	2
H36-15	7183	DFL16.2	2	V_K 21C	2
H36-18	7183	DFL16.2	2	V_K 21C	2
H36-1	7183	DFL16.2	2	V_K 21C	2
H36-4	7183	DFL16.2	2	V_K 21C	2
H37-63	7183'	DFL16.1	3	V_K 21C	5
H37-77	7183'	DFL16.1	3	V_K 21C	5
H37-85	3660	(SP2)	2	V_K 21C	4

FIGURE 6. Reactivity patterns of the 10 V_κ21C-associated antibodies against a panel of Sb mutant PR8 viruses as reported by Staudt and Gerhard (18). Binding that is equal to that observed against parental virus PR8 is indicated by a black box; partially reduced binding (50–75%), by a cross-hatched box; and strongly reduced binding (>75%, usually not detectable) by an open box. In addition, the gene segments used by each hybridoma are summarized. V11 denotes the use of a specific V_H gene, whereas 7183 and 3660 denote the use of a member of the 7183 and 3660 V_H families. H37-77 and 63 use a different member of the 7183 family than do H36-1, 4, 15 and 18, as indicated by 7183'.

For example, antibodies with the same reactivity patterns might have distinguishable specificities if tested against additional PR8 mutants. However, within the context of this panel of mutant PR8 viruses, it is apparent that the V_κ21C region plays a dominant role in antigen binding (Fig. 6). All of these antibodies bind to the Sb site in spite of major sequence differences in V_H regions. In particular, H36-5 and H36-1 have equivalent reactivity patterns even though they differ vastly with regard to their V_H regions (Fig. 6). It seems likely that these antibodies were primarily selected for by immunization with PR8, because of the contribution of V_κ21C to their binding sites. This interpretation is supported by the pattern of somatic mutation in the V_κ21C and V_H genes coding for these V regions.

Somatic Mutation and Intraclonal Diversity. The V_H region sequences have identified sets of clonally related hybridomas from mice H36 and H37. These clonal relationships are substantiated by shared nonproductive κ and H chain gene rearrangements and upstream κ rearrangements (Table I). Nevertheless, clonally related examples differ significantly in their V_H and V_κ sequences (Figs. 1–4). By definition these differences are the result of somatic mutation. By comparing the sequences and reactivity patterns of hybridomas of a single clone, one can reach several conclusions concerning the nature of somatic mutation and can assess the effect of somatic mutation on fine specificity.

The H36 hybridomas are members of two clones. Members of each clone have both shared mutations and unique mutations, because of sequential accumulation of mutations during clonal expansion. Two comprehensive genealogic trees that include all mutations in V_H and V_L coding regions are shown in Fig. 7. The clone that includes hybridomas H36-5, 7, and 17 stems from a precursor that rearranged the V_H-V11 and V_κ21C genes. These three hybridomas have accumulated 20, 12, and 19 V_H-coding region mutations and 9, 8, and 15 V_κ-coding region mutations (Fig. 7). The clone that includes hybridomas H36-1, 4, 15, and 18 is derived from a precursor that has rearranged a gene of the 7183 V_H family and

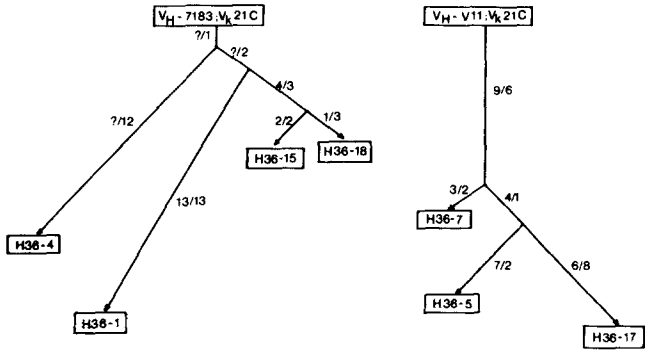


FIGURE 7. Genealogic trees of the two lymphocyte lineages generating the seven V_κ21C hybridomas from mouse H36. Each lineage is constructed to require the fewest number of parallel mutations. Shared mutations occurred early in clonal expansion, while the unique mutations occurred later in clonal expansion. One lineage (H36-1, 4, 15, and 18) derives from a lymphocyte that has rearranged a V_H gene of the 7183 family (V_H-7183) and the V_κ21C gene. The other lineage (H36-5, 7, 17) derives from a lymphocyte that has rearranged the V_H-V11 and V_κ21C genes. Numbers alongside branches are number of V_H nucleic acid substitutions from Figs. 1 and 2 over the number of V_κ nucleic acid coding region substitutions taken from Huppi et al. (manuscript in preparation). Question marks indicate where the number of V_H substitutions is not known, since the V_H-7183 germline sequence has not been determined.

the V_κ21C gene. The total number of V_H mutations cannot be determined for this clone in the absence of a germline sequence. However, we can compare the number of V_H and V_κ mutations by comparing the points at which different progeny in this clone have diverged. The number of mutations that have accumulated in V_H and V_κ after divergence are nearly the same (Fig. 7).

These observations suggest that the rate of somatic mutation must be similar at both V_H and V_L loci. To account for the extent of mutation at the V_κ locus of these cells, the mutation rate must be at least 10^{-3} mutations per base pair per division (30). The rate of mutation can be estimated at the V_H locus of hybridomas H36-5, 7, and 17, since all V_H mutations for these are known. These three hybridomas have accumulated an average of 17 coding region mutations. If these lymphocytes had divided every 18 h from the time of primary immunization to the time of fusion (24 d), then the rate of mutation at this V_H gene would be 1.5×10^{-3} per base pair per division. A major uncertainty in this estimate is the number of divisions during which mutation takes place. For example, mutations may accumulate before antigen selection, as suggested by studies on a pre-B cell line (10). Furthermore, the precursors of these clones may have been stimulated by unknown antigens, thereby accumulating mutations before exposure to influenza virus. However, even if the number of divisions were doubled, the calculated rate of mutation would remain high (7.5×10^{-4}), and would still be five to six orders of magnitude above the spontaneous mutation rate in eukaryotic DNA (40).

The nature and distribution of mutations in these V regions suggests that many of them are selected for by antigen. Analysis of the noncoding mutations, i.e., silent and flanking sequence substitutions, has shown these to be randomly distributed in and around these V_κ21C genes (K. Huppi, S. Clarke, S. Litwin, L.

Staudt, W. Gerhard, and M. Weigert, manuscript in preparation). If replacement mutations also occur at random in the V_κ21C gene, then selection of replacement mutations should influence both the ratio of replacement to silent mutations (R/S) and the distribution of replacement mutations, since antigen-selected mutations must be replacement types and should be located predominantly in complementary-determining regions (CDR). In addition, these features of antigen-selected mutations should be found in the V_κ genes and not in the V_H genes, if the V_κ21C plays the dominant role in determination of the specificity of these antibodies, as argued above. Indeed, the V_κ21C mutations strongly indicate antigen selection: first, the R/S ratio is 10, and, second, 80% of the replacement mutations are in CDRs (Huppi, manuscript in preparation). In contrast, the V_H regions of H36-5, 7, 17 (for which a germline sequence is known) have an R/S ratio of 1, and only 33% of the replacement mutations are in CDRs. Neither value is significantly different from that expected by a random distribution of mutations.

Surveys of V_κ21C-associated anti–HA (Sb) antibodies (Huppi, manuscript in preparation) show that all cases have accumulated at least one CDR mutation, including H2-6C4, a rare example of an IgM antibody. Thus, it appears that mutation of V_κ21C is required to create this specificity, an idea consistent with the recent observation (Staudt and Gerhard, manuscript in preparation) that V_κ21C antibodies are rare in the primary response to HA, but highly represented in the secondary response. That somatic mutation can influence specificity is seen by a comparison of the reactivity patterns between members of all three clones. The most striking differences are between hybridomas H37-77 and 63. H37-77 is unable to bind six of the nine Sb mutant virus, whereas H37-63 can bind all but one (Fig. 6). If somatic mutants of V_κ21C were indeed exclusively selected for by antigen, then at the initial stages of diversification, mutations should be predominantly replacement types in CDR. At later stages of diversification, additional CDR replacements may confer a selective advantage because of further modification of a combining site. However, it seems likely that most mutants with additional CDR amino acid substitutions would be at a selective disadvantage once a well-fitting combining site is formed. Hence, at later stages of diversification one would expect a preponderance of flanking and silent mutations. The order in which mutations have accumulated in these clones fits this model. The shared V_κ21C mutations presumed to have occurred early during clonal expansion, have a high R/S ratio and are predominantly replacement mutations in CDRs (Huppi, manuscript in preparation). In contrast, the unique mutations in each hybridoma have a low R/S ratio and are predominantly flanking region or silent coding region mutations.

Summary

This study focuses on 10 BALB/c anti–influenza virus (A/PR/8/34) hemagglutinin antibodies that have light chains encoded by the same variable region κ chain (V_κ) gene, V_κ21C. A comparison of antibodies from lymphocytes of independent origin reveals the contribution of germline diversity (combinatorial joining and association) to this response. Although combinatorial joining and

association contribute to sequence diversity, they appear to have little effect on the fine specificity of these antibodies.

Somatic mutation, in addition to contributing to the sequence diversity of these antibodies, creates differences in their fine specificity. The extent of mutation and its effect on fine specificity can be seen by comparing antibodies of lymphocytes from the same clone. These intraclonal comparisons also indicate that somatic mutation is an ongoing process occurring at a high rate (estimated to be at least 10^{-3} mutations per base pair per division) in the expressed V region heavy chain (V_H) and V_κ genes. Furthermore, both the nature and distribution of these mutations suggest that amino acid replacement mutations in the light but not the heavy chain are selected for by antigen.

Received for publication 12 November 1984 and in revised form 28 December 1984.

References

1. Sakano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature (Lond.).* 280:288.
2. Max, E., J. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. *Proc. Natl. Acad. Sci. USA.* 76:3450.
3. Bernard, O., N. Hozumi, and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell.* 15:1133.
4. Early, P. W., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D_H and J_H. *Cell.* 19:1981.
5. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Toonegawa. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature (Lond.).* 286:676.
6. Sakano, H., Y. Kurosawa, M. Weigert, and S. Tonegawa. 1981. Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. *Nature (Lond.).* 290:562.
7. Kurosawa, Y., H. von Boehmer, W. Haas, H. Sakano, A. Traunecker, and S. Tonegawa. 1981. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature (Lond.).* 290:565.
8. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* 155:201.
9. 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.
10. Wabl, M., P. D. Burrows, A. von Gabain, and C. Steinberg. Hypermutation at the immunoglobulin heavy chain locus in a pre-B cell line. *Proc. Natl. Acad. Sci. USA.* In press.
11. Clarke, S. H., J. L. Claflin, M. Potter, and S. Rudikoff. 1983. Polymorphism in antiphosphocholine antibodies reflecting evolution of immunoglobulin families. *J. Exp. Med.* 157:98.
12. Gearhart, P., N. D. Johnson, R. Douglas, and L. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.).* 291–29.
13. Rao, D. N., S. Rudikoff, H. Krutzsch, and M. Potter. 1979. Structural evidence for independent joining region gene in immunoglobulin heavy chains from anti-galactan

myeloma proteins and its potential role in generating diversity in complementarity-determining regions. *Proc. Natl. Acad. Sci. USA.* 76:2890.

14. 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.
15. Kaartinen, M., G. M. Griffiths, A. F. Markham, and C. Milstein. 1983. mRNA sequences define an unusually restricted IgG response to 2-phenyloxazalone and its early diversification. *Nature (Lond.).* 304:320.
16. Coleclough, C., R. P. Perry, K. Karjalainen, and M. Weigert. 1981. Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature (Lond.).* 290:372.
17. Van Ness, B. G., L. Coleclough, R. P. Perry, and M. Weigert. 1982. DNA variable and joining gene segments of immunoglobulin κ light chain is frequently retained in cells that rearrange the κ locus. *Proc. Natl. Acad. Sci. USA.* 79:262.
18. Staudt, L. M., and W. Gerhard. 1983. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. I. Significant variation in repertoire expression between individual mice. *J. Exp. Med.* 157:687.
19. Perry, R. P., D. E. Kelley, U. Schibler, K. Heubner, and C. M. Croce. 1979. Selective suppression of the transcription of ribosomal genes in mouse-human hybrid cells. *J. Cell Physiol.* 98:553.
20. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.
21. Marcu, K. B., J. Banerji, N. A. Penncavage, R. Lang, and N. Arnheim. 1980. 5′ Flanking region of immunoglobulin heavy chain constant region genes displays length heterogeneity in germlines of inbred mouse strains. *Cell.* 22:187.
22. Leder, P., D. Tiemeier, and L. Enquist. 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. *Science (Wash. DC).* 196:175.
23. Hohn, B., and K. Murray. 1977. Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. *Proc. Natl. Acad. Sci. USA.* 74:3259.
24. Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science (Wash. DC).* 196:180.
25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
26. Gough, G. R., M. J. Brunden, J. G. Nadeau, and P. T. Gilham. 1982. Rapid preparation of hexanucleotide triester blocks for use in polydeoxyribonucleotide synthesis. *Tetrahedron Lett.* 23:3439.
27. Air, G. M. 1979. Nucleotide sequence coding for the "signal peptide" and N terminus of the hemagglutinin from Asian (H2N2) strain of influenza virus. *Virology.* 97:468.
28. Biggin, M. D., J. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA.* 80:3963.
29. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 74:560.
30. McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA.* 81:3180.
31. Brodeur, P. H., and R. Riblet. 1984. The immunoglobulin heavy chain region (*Igh-V*) locus in the mouse. I. 100 *IgH-V* genes comprise seven families of homologous genes. *Eur. J. Immunol.* 14:922.
32. Crews, S., J. Griffin, H. Huang, K. Calame, and L. Hood. 1981. A single V_H gene

segment encodes the immune responses to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell.* 25:59.

33. Near, R. I., E. C. Juszczak, S. Y. Huang, S. A. Sicari, M. N. Margolies, and M. L. Gefter. 1984. Expression and rearrangement of homologous immunoglobulin V_H genes in two mouse strains. *Proc. Natl. Acad. Sci. USA.* 81:2167.
34. Heinrich, G., A. Traunecker, and S. Tonegawa. 1984. Somatic mutation creates diversity in the major group of mouse immunoglobulin κ light chains. *J. Exp. Med.* 159:417.
35. McKean, D. J., M. Bell, and M. Potter. 1978. Mechanisms of antibody diversity: multiple genes encode structurally related mouse κ variable regions. *Proc. Natl. Acad. Sci. USA.* 75:3913.
36. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature (Lond.).* 276:785.
37. Selsing, E., and U. Storb. 1981. Somatic mutation of immunoglobulin light-chain variable-region genes. *Cell.* 25:47.
38. Brack, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell.* 15:1.
39. Caton, A. J., G. G. Brownlee, J. W. Yewdell, and W. Gerhard. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell.* 31:417.
40. Vogel, F. 1970. Spontaneous mutations in man. *In* Chemical Mutagenesis in Mammals and Man. F. Vogel and G. Röhrborn, editors. Springer-Verlag, Heidelberg, 16–68.
41. Newell, N., J. R. Richards, P. W. Tucker, and F. R. Blattner. J genes for heavy chain immunoglobulins of mouse. *Science (Wash. DC).* 209:1128.