Three-Dimensional Structure of Murine Anti-*p*-azophenylarsonate Fab 36-71. 2. Structural Basis of Hapten Binding and Idiotypy†

Roland K. Strong,*,‡ Gregory A. Petsko,§ Jacqueline Sharon,∥ and Michael N. Margolies⊥

Mail Stop 156-29, California Institute of Technology, Pasadena, California 91125, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254-9110, Departments of Pathology and Biochemistry, Boston University School of Medicine, 80 East Concord Street, Room K707, Boston, Massachusetts 02118, and Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114

Received August 29, 1990; Revised Manuscript Received December 26, 1990

ABSTRACT: Comparison between the structures and solvent-accessible surfaces of the antigen-binding fragments of two murine anti-*p*-azophenylarsonate monoclonal antibodies, one bearing a major cross-reactive idiotype of A/J strain mice (36-71) and one lacking the idiotype (R19.9; Lascombe et al., 1989), highlight the structural basis for the determination of hapten affinity and idiotypy. Since the sequence of R19.9 is identical with the germline-encoded sequence at 16 positions in both heavy-chain and light-chain variable regions where somatic mutations and junctional differences have occurred to produce the 36-71 sequence, the structure of R19.9 can be used to model the structure of the germline-encoded antibody (36-65) in the regions around these sites. These 16 sequence differences exclude the third heavy-chain complementarity-determining region because R19.9 utilizes a D gene segment not associated with the predominant idiotype, which is 4 residues longer than the canonical D gene segment utilized in the sequences of 36-71 and 36-65. This difference between the structures of R19.9 and 36-71 does not affect the validity of using the structure of R19.9 to model the structure of 36-65 since the third heavy-chain complementarity-determining region is highly solvent-exposed in both 36-71 and R19.9, and does not interact with any of these 16 sites. Comparing the structures of 36-71 and R19.9 suggests that only three of the differences in the heavy-chain sequences, and three of the differences in the light-chain sequences of 36-71 and 36-65, increase the affinity for hapten. The substitutions in the light chain appear to affect the binding constant for hapten by altering the overall conformation of the third heavy-chain complementarity-determining region. Regions where the solvent-accessible surfaces of 36-71 and R19.9 differ have been located. Presumably, these differences in accessible surface area account for the lack of reactivity with anti-idiotypic antibodies for R19.9 versus 36-71. Positions in the sequence where differences are known to affect the binding of anti-*p*-azophenylarsonate antibodies to monoclonal anti-idiotope antibodies generally map to the region around the hapten-binding site. Not unexpectedly, substitutions at these locations in the sequence do not always appear to affect anti-idiotope binding directly but may alter idiotype/anti-idiotype affinity indirectly by altering the conformation of contact residues.

The surfaces of the variable (V)[1] regions of antibodies bear collections of unique antigenic determinants (idiotopes) which taken together define the idiotype of a given antibody. Experimentally, the idiotype is defined through recognition by a polyclonal anti-idiotype antiserum; individual idiotopes may be defined on the basis of reactivity with specific monoclonal anti-idiotope antibodies [reviewed in Bona (1981) and Greene and Nisonoff (1984)]. These idiotopes can be recognized by anti-idiotypic antibodies, much as protein antigens can be recognized by anti-protein antibodies (Alzari et al., 1988). As "networks" of idiotypes/anti-idiotypes have been proposed to play an important role in the regulation of the immune response (Jerne, 1974), it is important to understand the structural basis of the determination of idiotype (Bentley et al., 1989).

When A/J strain mice are immunized with *p*-azophenylarsonate (Ars) conjugated to keyhole limpet hemocyanin, the majority of the antibodies produced share a cross-reactive idiotype, designated CRI_a (Kuettner et al., 1972) or Id^{CR} (Marshak-Rothstein et al., 1981). The evidence to data indicates that all Id^{CR+} anti-Ars antibodies are the result, in the secondary immune response, of the combination of a single set of gene segments encoding variable regions of immunoglobulins (Igs): V_H Id^{CR} (Siekevitz et al., 1982, 1983), $DFL_{16.1e}$ (Landolfi et al., 1986), J_H2 (Wysocki et al., 1986), and $V\kappa10$ and $J\kappa1$ (Sanz & Capra, 1987; Wysocki et al., 1987). However, monoclonal Id^{CR+} antibodies demonstrate heterogeneity in binding to rabbit anti-idiotypic antisera in binding assays (Lamoyi et al., 1980; Marshak-Rothstein et al., 1980a), and anti-idiotypic reagents detect distinct "private" idiotopes among Id^{CR+} monoclonal antibodies (Marshak-Rothstein et al., 1980b). This microheterogeneity arises largely

† This work was supported by the U.S. Public Health Service through NIH Grants CA 24432, HL 19259, and AI 23909. J.S. is a recipient of an ACS scholar award.

* Address correspondence to this author.
‡ California Institute of Technology.
§ Brandeis University.
∥ Boston University School of Medicine.
⊥ Massachusetts General Hospital and Harvard Mecical School.

[1] Abbreviations: CDR, complementarity-determining region; C_H1, first heavy-chain constant domain; C_L, light-chain constant domain; *F*, structure factor amplitude; Fab, antigen-binding fragment of an antibody consisting of V_H–C_H1 and V_L–C_L; H chain, heavy chain; I, intensity; Id^{CR+}, bearing the cross-reactive idiotype predominant among A/J anti-*p*-azophenylarsonate antibodies; Id^{CR-}, lacking the cross-reactive idiotype predominant among A/J anti-phenylarsonate antibodies; Ig, immunoglobulin; L chain, light chain; mAb, monoclonal antibody; Ars, *p*-azophenylarsonate; Ars-tyrosine, C^2-(*p*-azophenylarsono)-L-tyrosine; rms, root mean square; σ, standard deviation; SAS, solvent-accessible surface; V, immunoglobulin variable domain; V_H, heavy-chain variable domain; V_L, light-chain variable domain.

0006-2960/91/0430-3749$02.50/0 © 1991 American Chemical Society

```
          1                 10                  20                  30  CDR1            40
R19.9H    ( N O T   S E Q U E N C E D ) - - - - - - - - - - - - - - - - - - V - - - - - - - - - - - - - - -
36-65H    E V Q L Q Q S G A E L V R A G S S V K M S C K A S G Y T F T S Y G I N W V K Q R P G Q G L E W I G
36-71H    - - - - - - - - V - - - - - - - - - - - - - - - - - - - - - - N - - - - - - - - - - - - - - - - -

          50             CDR2 60                    70                  80                  90
R19.9H    - - - - - K - - L S - - - - - - - - - - - - - - R - - - - - - - - - - - - - - - - - - - - - - - -
36-65H    Y I N P G N G Y T K Y N E K F K G K T T L T V D K S S S T A Y M Q L R S L T S E D S A V Y F C A R
36-71H    - N - - - - - - I T - - - - - - - - - - - - - - - - - N - - - - - - - - - - - - - - - - - - - - -

           100    CDR3   a   b c d         110                 120
R19.9H    - F - · - - - H/D L A V - - - - S - - - - - - - - - - -
36-65H    S V Y Y G G S ·   · · · Y Y F D Y W G Q G T T L T V S S
36-71H    - E - - - - - ·   · · · - K - - - - - - - - - - - - - -


          1                 10                  20        CDR1 30                   40
R19.9L    - - - - - - - - - - - - - - - - - - - - - - - - ? - - - - - ? - - - - - - - - - - - - - - - - - -
36-65L    D I Q M T Q T T S S L S A S L G D R V T I S C R A S Q D I S N Y L N W Y Q Q K P D G T V K L L I Y
36-71L    - - - - - - I P - - - - - - - - - - - S - - - - - - - - - N - F - - - - - - - - - - S I - - - - -

          50 CDR2           60                  70                  80                  90 CDR3
R19.9L    - - - - - ? - - - - - - - - - - - - - - - - - - - - - - - - H - - - - - - - - - - - S - - - - - -
36-65L    Y T S R L H S G V P S R F S G S G S G T D Y S L T I S N L E Q E D I A T Y F C Q Q G N T L P R T F
36-71L    F - - - S Q - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - A - - - - -

           100
R19.9L    - - - - - - ? - ? -
36-65L    G G G T K L E I K R
36-71L    - - - - - - - - - -
```

FIGURE 1: Amino acid sequences of the variable domains of the H chain (top) and L chain (bottom) of 36-65 (Siekevitz et al., 1982; Wysocki et al., 1987; Parhami-Seren et al., 1989), 36-71 (Sharon et al., 1989), and R19.9 (Lascombe et al., 1989) are shown. The sequences are compared to that of the germline 36-65 antibody; *dashes* indicate sequence identity; *dots* denote gaps in the sequences introduced to maximize homology. The one-letter code for amino acids is employed (IUPAC–IUB); numbering is sequential relative to the 36-65 sequence; the insertion in R19.9 relative to 36-65 and 36-71 is numbered 105a–105d. Question marks in the R19.9 sequence indicate undetermined residues; a *slash* indicates an ambiguous residue assignment.

from the presence of somatic mutations within the V regions which accumulate during clonal expansion following hyperimmunization as confirmed by amino acid sequence comparisons (Estess et al., 1979; Margolies et al., 1981; Marshak-Rothstein et al., 1980b; Siegelman & Capra, 1981; Slaughter & Capra, 1983) and site-directed mutagenesis studies (Sharon, 1990a; Parhami-Seren et al., 1990a,b).

In this report, we compare the antigen-binding fragments (Fabs) of two anti-Ars antibodies: R19.9 (CRI$_d$ or Id^{CR-}; Lascombe et al., 1989) and 36-71 (CRI$_a$ or Id^{CR+}; Rose et al., 1990; Strong et al., 1991). In the 1.85-Å model of 36-71, residue Arg-96 in the light (L) chain and residues Asn-35, Trp-47, and Ser-99 in the heavy (H) chain contact the arsonate moiety of the hapten; an additional bond is found between the arsonate group and a tightly bound water molecule.[2] The phenyl moiety of the hapten packs against two tyrosine side chains at positions 50 and 106 in the H chain.[3]

Since the sequence of R19.9 (Lascombe et al., 1989) is identical with the germline-encoded sequence at 16 sites in both the H-chain and L-chain variable regions where somatic mutations and junctional differences have occurred to produce the 36-71 sequence (Sharon et al., 1989; see Figure 1), the structure of R19.9 can be used to model the structure of a germline-encoded antibody, 36-65 (Rothstein et al., 1980; Siekevitz et al., 1982; Wysocki et al., 1987; Parhami-Seren et al., 1989) in the regions around these sites. These 16 sequence differences exclude the third H-chain complementarity-determining region (HCDR3) because R19.9 utilizes a D gene segment not associated with the predominant idiotype, which is 4 residues longer than the canonical D gene segment utilized in the sequence of 36-71 and 36-65. Comparison of the structures of 36-71 and R19.9 thus allows the structural effects of these sequence changes to be determined, as well as their effect on hapten binding. By comparing the three-dimensional crystal structures of these two anti-Ars Fabs, it is also possible to elucidate some of the structural features responsible for the determination of idiotype by examining regions where the solvent-accessible surfaces of 36-71 and R19.9 differ either chemically or structurally. Presumably, these differences in the accessible surface are involved in the differences in reactivity of R19.9 versus 36-71 with polyclonal anti-idiotypic sera and with monoclonal anti-idiotype antibodies, since it is this surface which would bind to an anti-idiotype antibody.

[2] The numbering of amino acids in this report follows the sequential numbering of residues in the 36-65 molecule.

[3] See the preceding paper (Strong et al., 1991) for further discussion.

Materials and Methods

Accessible Surface Area Calculations. The solvent-accessible surface (SAS) was calculated by using the method of Connolly (1983) and the refined coordinates of Fabs 36-71[3] and R19.9 (Lascombe et al., 1989); the probe size is 2.8 Å in diameter, comparable in size to a water molecule. Only contact surfaces were calculated.

Superposition of 36-71 onto R19.9. The refined coordinates at 1.85-Å resolution, R factor = 24.8%, of Fab 36-71[3] were superimposed onto the 2.8-Å resolution, R factor = 29.9%, structure of Fab R19.9 currently available (Lascombe et al., 1989) in the Brookhaven Protein Data Bank using the least-squares minimization routine in the program HYDRA (Polygen Inc.). All residues in V_H and V_L were included except those residues within the CDRs. The rms deviation in atom position after docking is 2.2 Å (all variable domain atoms except those in CDR residues); the average distance between atoms is 1.6 Å, and the maximum between identical atoms is 10.5 Å. Every atom in the included residues was considered in the process, except for nonidentical atoms in residues at positions where the sequences differed.

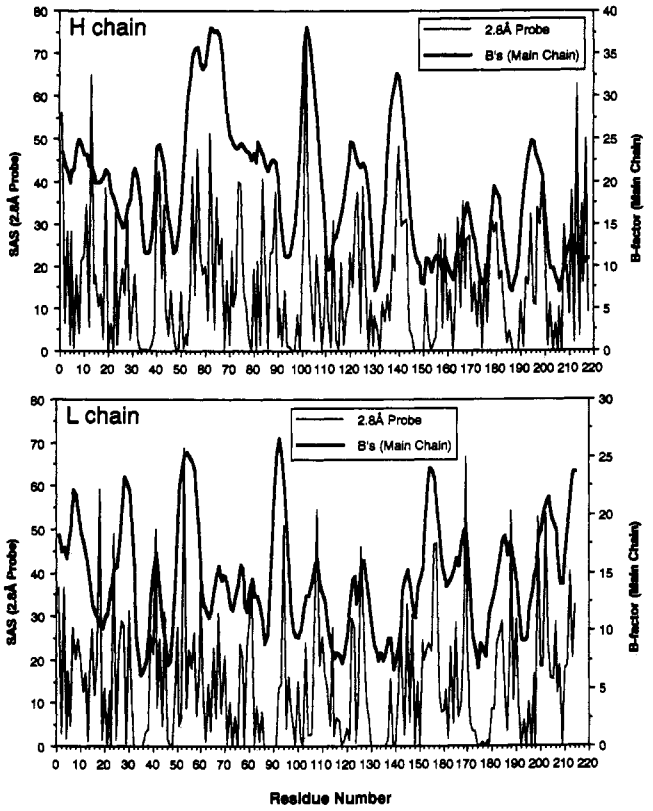

FIGURE 2: Solvent accessibility and *B* factor (the average over the main-chain atoms only) are displayed for each residue in the high-resolution refined structure of Fab 36-71. The units are Å^2 for both parameters; the numbering follows the sequential order of the residues in the sequence on the abscissa.

RESULTS

The Solvent-Accessible Surface of 36-71 Is Strongly Correlated with the Temperature Factors Calculated during Refinement. In Figure 2 is shown the SAS of 36-71 plotted with the refined individual *B* factors (Strong et al., 1991) averaged over the main-chain atoms. *B* factors are a measure of the positional uncertainty associated with a given atom in a crystal structure and are strongly correlated with thermal motion (Dodson, 1981). The correlation between the calculated *B* factors and the SAS is excellent and demonstrates that solvent accessibility is a major contribution to the observed thermal motion. However, the *B* factors associated with H-chain complementarity-determining region 2 (HCDR2; see Figure 1) are still higher than what can be explained by the SAS and may indicate that this region of the current model is not in perfect agreement with the observed X-ray data, or that there are several possible conformations for these regions. Table I shows that HCDR1 and LCDR3 display relatively low SASs and that HCDR2 and, to a lesser extent, LCDR2 dominate the total SAS of the CDRs. Almost a third of the total accessible surface of the CDRs is associated with HCDR2. The effect is not proportional to the length of these CDRs and is probably a result of the structure of these two CDRs; HCDR2 partially covers HCDR1,[3] and LCDR2 "loops" into the solvent (see Figure 4).

The Structures of Fabs 36-71 and R19.9 Are Very Similar Except for HCDR3. The superimposed coordinates of 36-71 and R19.9, shown in Figure 3, emphasize the small difference in "elbow" angle between the two structures (179° vs 178°). Note that the two structures are quite comparable throughout the framework regions in the variable domains; the only noticeable deviations occur in the CDRs. In Figure 4, these differences are emphasized. The chain trace is different in

Table I: Solvent Accessibility[a] of the Complementarity-Determining Regions

CDR[b]	solvent-accessible surface (Å^2)	SAS (% total CDR)	no. of residues in CDR	no. of residues (% total CDR)
HCDR1	21.8	1.9	5	8.2
HCDR2	354.0	31.4	17	27.9
HCDR3	227.6	20.2	12	19.7
LCDR1	172.2	15.3	11	18.0
LCDR2	240.6	21.4	7	11.5
LCDR3	109.2	9.7	9	14.8
total	1125.6	100	61	100

[a] The solvent-accessible surface is calculated by using the method of Connolly (1983); the probe size is 2.8 Å in diameter. [b] The amino acid residues included in each CDR are shown in Figure 1.

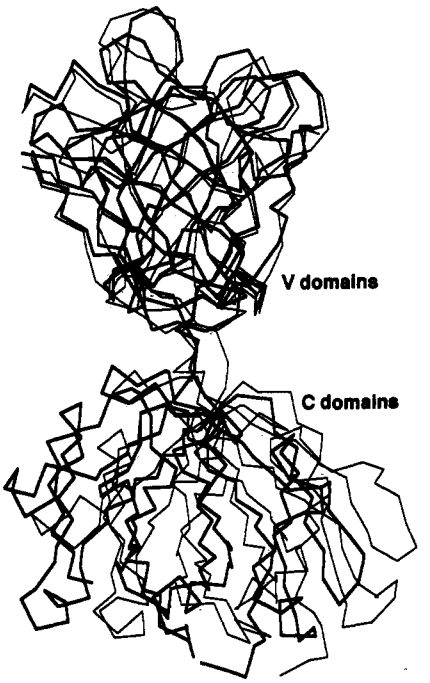

FIGURE 3: Cα backbones of R19.9 and 36-71 are superimposed to show the effect of the difference in "elbow" angle between the two molecules on the overall structure of the Fabs. The variable domains are at the top of the figure and were used in the least-squares alignment; the backbone of 36-71 is shown with *thick* lines. The "elbow" angle of 36-71 is 179° (Strong et al., 1991); the "elbow" angle of R19.9 is 178° (Lascombe et al., 1989).

the regions of LCDR1 and LCDR2; the largest deviation between the two structures occur in the middle of HCDR3. This is not unexpected as R19.9 utilizes a different (noncanonical) D gene than 36-71 which codes for a CDR four residues longer than that of 36-71 (see Figure 1), and the location of this insertion is in the middle of HCDR3. However, while the conformations of the residues at the base of HCDR3 are similar in the two structures, the chain trace in the region of the insertion in R19.9 appears to pass through the hapten as fitted in our model of the complex. Since Lascombe and co-workers (Lascombe et al., 1989) report that the electron density maps in this region were "less clear" than for the rest of the molecule, it is possible that with further refinement the position of this loop will no longer interfere directly with the hapten-binding site as determined for 36-71 from a 4.5-Å resolution difference electron density map between the complex with phenylarsonate and the native Fab.[3] Notwithstanding this discrepancy, the chain traces of the two molecules are very similar, even in HCDR2.

Figure 6 displays the residues in 36-71 identified as directly involved in binding the hapten[3,4] and the corresponding residues

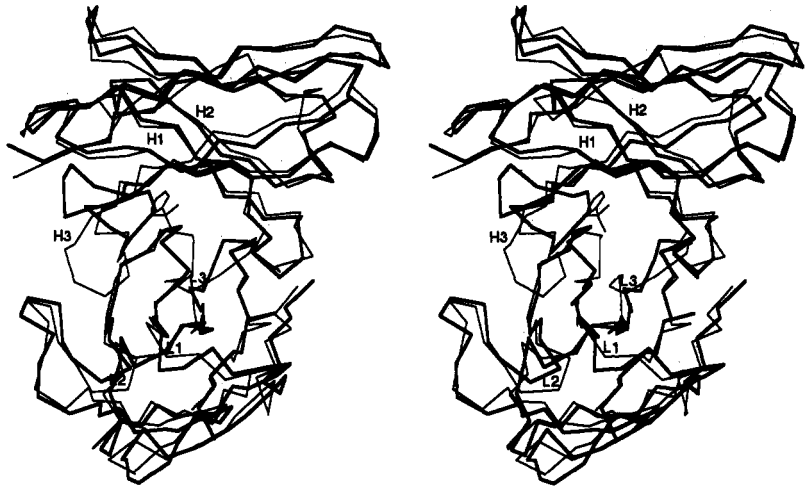

FIGURE 4: Superposition of the variable domains of Fabs 36-71 and R19.9 is displayed in this stereo diagram. V_H is at the top of the figure; the Cα backbone of R19.9 is shown with *thick* lines. The hapten (phenylarsonate) position has been included on the basis of the modeled complex with 36-71.[3]

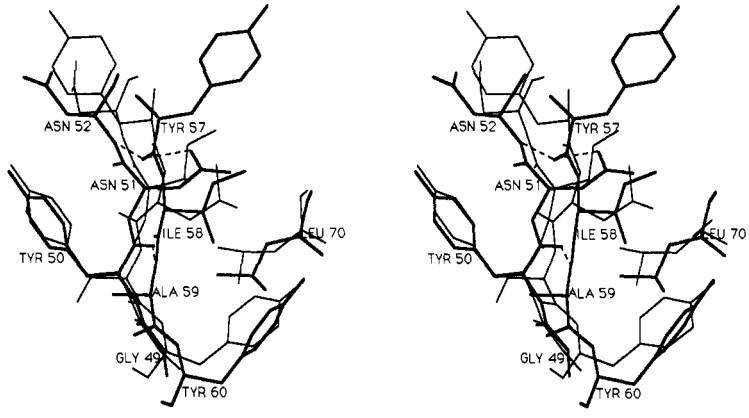

FIGURE 5: Residues in HCDR2 of 36-71 and R19.9 Fabs which are involved in the distorted β-sheet structure of this CDR are displayed in stereo; residues from 36-71 are shown with *thick* lines. Putative hydrogen bonds are shown with *dashed* lines. Residue 59H was built as an alanine instead of threonine because of the truncated side-chain density.

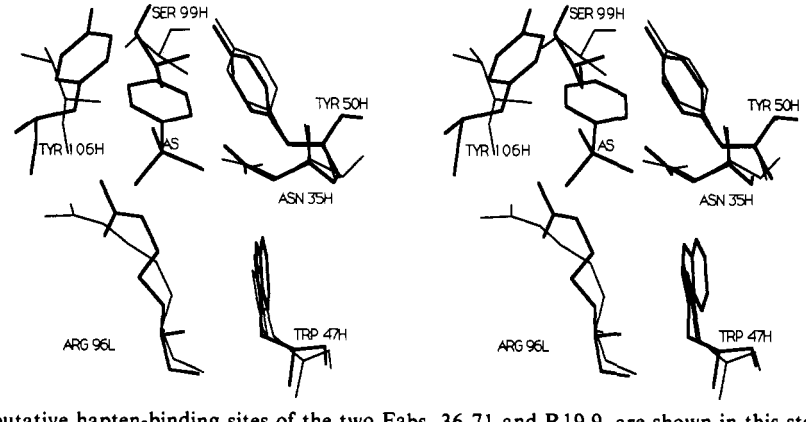

FIGURE 6: Superimposed putative hapten-binding sites of the two Fabs, 36-71 and R19.9, are shown in this stereodiagram, which includes the position of the hapten, phenylarsonate, based on the modeled complex with 36-71.[3] The arsenic atom of the hapten has also been labeled. Only residues contacting hapten from 36-71 (*thick* lines) and the corresponding residues from R19.9 (*thin* lines) based on the three-dimensional structure are displayed. The R19.9 residue superimposed on 36-71 Tyr-106H is Val-105dH in the R19.9 sequence (see Figure 1); this residue is the structural equivalent to Tyr-106H (36-71).

in the superimposed R19.9 structure. The R19.9 HCDR3 residues which would pass through our placement of the hapten are not shown in this figure. The agreement is again excellent; minor changes occur for Ser-99H and in the conformation of Tyr-50H, which is more "edge-on" in R19.9 to the Ars aryl moiety. In addition, the γ-O atom of Ser-99H potentially hydrogen-bonds to the hydroxyl of Tyr-50H in R19.9. The presence of this hydrogen bond between Ser-99H and Tyr-50H might affect affinity for hapten detrimentally by competing against the putative hydrogen bond between Ser-99H and the arsonate group of the hapten. The side chain of Arg-96L is in a different orientation in R19.9 as compared to 36-71. The ϵ-N atom of Arg-96L in R19.9 is closer to the position of the

[4] Tyr-107H in 36-71 is *structurally* equivalent to Val-105dH in R19.9 (see Figure 1) when the three-dimensional structures are superimposed.

terminal N of Arg-96L in 36-71. This is a less favorable position for Arg-96L to participate in a hydrogen bond with the arsonate group in R19.9. The equivalent residue in R19.9 at position 106H is a valine,[4] not a tyrosine; this change would also affect, probably lowering relative to 36-71, the affinity for ligand by altering the hydrophobic surface which packs against the phenyl moiety of the hapten.

Five of the Sequence Differences between Fab R19.9 and Fab 36-71 Alter the Structure of Fab R19.9 Relative to 36-71. The sequences of 36-71 and R19.9 (excluding the insertion) differ at 25 sites in both V_H and V_L;[5] of these, 16 of the residues in R19.9 at these positions are identical with the sequence of the unmutated antibody 36-65 (see Figure 1) and are considered under Discussion. Two of the sequence differences occur at the V_H–D and D–J_H gene junctions (positions 100H and 107H), and will also be discussed later. Two of the remaining seven changes either are conservative or occur at solvent-exposed sites, and therefore do not significantly alter the structure of the molecule: 92L and 74H. One change occurs at position 54H in a region where the main chain has a different conformation in the two Fabs. Residues 53–55H in HCDR2 follow slighly different paths in 36-71 and R19.9. Position 80L is solvent-exposed, but the glutamine at this position in 36-71 makes a hydrogen bond to the γ-O atom of residue 168L in C_L which the His in R19.9 does not. Since this bond bridges the gap between the variable and constant domains, the difference at this site may contribute to the slight observed difference in "elbow" angle (see Figure 3). The difference between Ile-34H (36-71) and Val-34H (R19.9) affects the conformation of the side chain of Tyr 50H. The additional methyl group pushes the main chain at 34H toward the ring of Tyr 50H, forcing the ring to pack "face-on" to phenylarsonate in the complex with 36-71, optimizing the binding surface of 36-71 (and 36-65) versus R19.9 for hapten.

Residues 58–59H are different in the three antibody sequences under consideration and lie in the distored β-sheet region of HCDR2 which packs against 49–52H. Examination of Figure 2 shows a "dip" in both the *B* factor and SAS plots at these sites (see the region of residues 58–59H). The side chain of the residue at 58H (either Ile or Leu) packs in a small hydrophobic "pocket" consisting of Tyr-57H, Tyr-60H, and Leu-70H (see Figure 5). Since this segment in HCDR2 is bordered by regions of greater disorder (as indicated by lower quality electron density and higher *B* factors), these latter residues may act as an "anchor" determining the conformation of this section of HCDR2. A Lys at 59H (as in 36-65) could interact with carbonyl oxygens at 93L or 94L, across the Ars binding site from 59H. The presence of Lys-59H could therefore sterically hinder Ars binding, or disrupt the structure of HCDR2, and might account for the nearly 4-fold increase in affinity observed when threonine (as in 36-71) replaces lysine at this position in site-directed mutagenesis experiments (Sharon et al., 1990; see Table II).

Table II: Binding Constants of Wild-Type and Mutant Antibodies for Ars-Tyrosine[a]

protein tested	$K_{a'} \times 10^5$ M^{-1} (relative affinity) with 36–65 L chain	with 36–71 L chain
wild-type H chain		
36–65	1.64 (1.0)	2.36 (1.0)
36–71	270 (165)	313 (133)
36–65 mutant H chain		
Ala-9 → Val	1.4 (0.9)	2.3 (1.0)
Tyr-32 → Asn	1.5 (0.94)	1.9 (0.8)
Ile-51 → Asn	2.2 (1.3)	4.5 (1.9)
Thr-58 → Ile	4.8 (2.9)	6.3 (2.7)
Lys-59 → Thr	6.5 (4.0)	9.1 (3.9)
Ser-77 → Asn	1.5 (0.9)	2.3 (1.0)
Val-100 → Glu	0.47 (0.29)	0.54 (0.23)
Tyr-107 → Lys	21 (13)	42 (18)
Thr-58/Lys-59/Tyr-107 → Ile/Thr/Lys	395 (241)	489 (207)
Thr-58/Lys-59/Val-100/Tyr-107 → Ile/Thr/Glu/Lys	40 (24)	244 (103)
Val-100/Tyr-107 → Glu/Lys	5.0 (3.0)	9.6 (4.1)
36–65 mutant L chain		
Arg-96 → Ala	none detected (0)	none detected (0)

protein tested	$K_{a'} \times 10^5$ M^{-1} (relative affinity)
hybridoma protein	
36–65	1.79 (1.0)
36–71	372 (227)

[a] Binding constants were determined by fluorescence quenching with Ars-tyrosine at 23–25 °C (Sharon 1990b; Strong et al., 1991).[3]

Discussion

The Structural Effects of Many of the Sequence Differences between 36-71 and 36-65 May Be Deduced by Using the Structure of R19.9 as a Model for the Structure of 36-65 around These Positions. The somatically mutated anti-Ars antibody 36-71 binds hapten with apparently 200-fold greater affinity (Rothstein & Gefter, 1983) as compared to antibody 36-65 which utilizes the same ("canonical")[6] gene segments but is unmutated. In addition, the two antibodies differ at the D gene junctions (see Figure 1). In Table II, the effects on hapten affinity of making mutations one at a time in the H chain of 36-65 that correspond to the sequence differences in the H chain in 36-71 as compared to 36-65 (Sharon, 1990b) are summarized. In these experiments, the mutated 36-65 H-chain genes were transfected into either 36-65 or 36-71 L-chain-producing cell lines. To elucidate the structural changes wrought by these mutations, we compared the structure of 36-71 to R19.9 in the regions about the positions where these changes occur and where 36-65 and R19.9 share a common primary structure and, by conjecture, a common tertiary structure.

At several sites, the mutations in 36-71 relative to 36-65 are conservative, or occur in solvent-exposed portions of the molecule: 20L, 30L, 43L, 93L, 51H, and 77H. Several of the other sequence differences entail only slight structural changes: the substitution of isoleucine in 36-71 for valine in 36-65 at position 44L affects the side chains of 95H and 111H as suggested by the Fab crystal structures of 36-71 and R19.9. Leu-54L (36-65) might pack internally in a hydrophobic pocket, rather than projecting into solvent as does Ser-45L in 36-71. Several sequence changes introduce "puckers" into the main chain. The Ile-Pro combination at positions 7L and 8L in 36-71 cooperates synergistically to pull residues 7L to 18L out of register between the two crystal structures by one amino acid in 36-71 as compared to R19.9 and, presumably, 36-65. The change at position 9H also introduces a "kink" in the main chain, but on a much smaller scale; this may explain the lack of an observed effect on the affinity of the Ala-9H → Val

[5] The sequence used in comparisons between R19.9 and 36-71 or 36-65 at indeterminate positions in R19.9 (Lascombe et al., 1990) is from the entry for the three-dimensional structure of Fab R19.9 in the Brookhaven Protein Data Bank.

[6] In this report, "canonical" refers to the specific set of gene segments found in murine anti-Ars antibodies expressing Id^{CR+}; "noncanonical" antibodies utilize different gene segments.

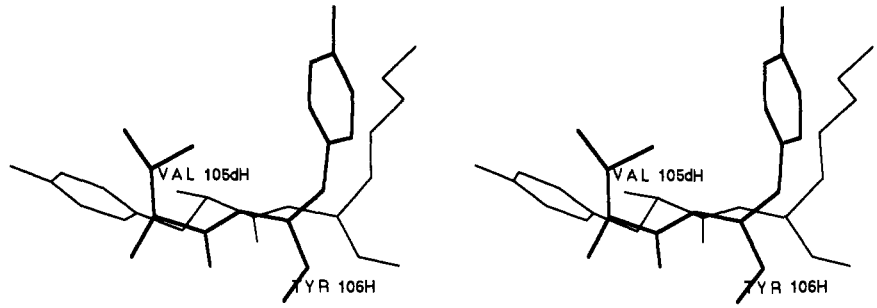

FIGURE 7: Superposition of residues 106H and 107H (36-71) onto 105dH and 106H (R19.9) is shown in this stereodiagram. Residues from R19.9 are shown with *thick* lines and are numbered. These residues illustrate the structural effect of the substitution of lysine (in 36-71) for tyrosine (in 36-65) at position 107H on the residue at position 106H, which affects hapten affinity. Residues 105dH and 106H in R19.9 are the structural equivalents of 36-71 residues 106H and 107H, respectively, based on the superposition of the variable domains.

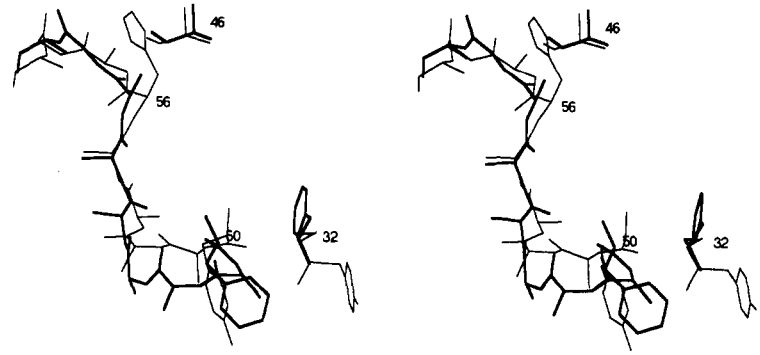

FIGURE 8: Residues 32L and LCDR2 in Fab 36-71 (*thick* lines) and Fab R19.9 are shown in this stereo figure. For clarity, all residues in both molecules except 32, 46, 50, and 56 have been displayed as alanines, including Gln-55L in 36-71.

mutation (see Table II). The substitution of asparagine in 36-71 at position 51H for isoleucine in 36-65 does not appear to have an effect on the structure when comparing the structure of 36-71 to R19.9 in the region surrounding this site. The observed 2-fold change in affinity corresponding to this substitution (see Table II) may be due to a change in the conformation of the adjacent Tyr-50H, a ligand to Ars, which might occur in the 36-65 background but is not observed in R19.9.

The residues at the V_H–D and D–J_H junctions (100H and 107H) affect hapten-binding indirectly. The ϵ-O atoms of Glu-100H in 36-71 hydrogen-bond to the carbonyl oxygens of 31H and 101H. A negative charge at this location (100H) probably acts to repel the phenolic hydroxyl group of Tyr-50H, forcing this side chain closer to the hapten, and reducing the observed affinity 3–4-fold through unfavorable steric interactions when the single mutation (Val-100H → Glu) is made in the 36-65 H chain (Table II). The hydrophobic residues found in 36-65 (Val) and R19.9 (Phe) at this position would not have this effect. The lysine at position 107H in 36-71 is involved in a salt bridge to Asp-109H on the "back" side of the V_H domain, on the other side of HCDR3 from the Ars-binding site. In R19.9, Tyr-106H, the structural equivalent to Lys-107H in 36-71, is a member of the "aromatic cluster" described by Lascombe and co-workers (Lascombe et al., 1989). The net result of the replacement of lysine (36-71) for tyrosine (36-65 and R19.9) is to move the Cα of the residue at position 106H in 36-71 (position 105H in R19.9; see Figure 1) which directly contacts the hapten (see Figure 7). In 36-71, residue 106H is pulled away fom the putative Ars site, allowing the aromatic side chain to pack optimally against the phenyl ring of the hapten. In R19.9, Val-105dH is pushed toward the hapten location, causing a steric clash in the model of the complex, explaining the 18-fold increase in affinity due to the single change (Tyr-107H → Lys) in 36-71 (see Figure 7 and Table II). The substitution of lysine for tyrosine at 107H seems to compensate for the valine to glutamate substitution at 100H (36-71 vs 36-65), yielding a net 4-fold increase in hapten affinity (see Table II). The influence of the D gene junctional residues on hapten affinity was noted previously in comparing sequences of primary response (unmutated) antibodies which differed from each other only at the V_H–D and D–J_H junctions (Parhami-Seren et al., 1989).

LCDR1, LCDR2, and HCDR3 Interact through a Cluster of Aromatic Residues; Changes in the Sequence of LCDR1 and LCDR2 Can Alter the Observed Affinity for Hapten by Altering the Conformation of HCDR3. The "aromatic cluster" in Fab R19.9 described by Lascombe and co-workers (Lascombe et al., 1989) consists of residues 32L, 49L, 50L, and 100–102H. In Fab 36-71, Tyr-102H is completely exposed to solvent and does not participate in such a structure. The substitution of phenylalanines for tyrosines at positions 32L and 50L (36-71 vs 36-65) has profound effects on the conformation of LCDR2 and this hydrophobic cluster, despite the fact that the substitution of a phenylalanine for a tyrosine would ordinarily be considered fairly conservative. The absence of a hydoxyl group on the side chain of residue 32L in 36-71 allows this phenylalanine to pack more closely into the "cluster", completely shielding it from solvent. The corresponding tyrosine in R19.9, and by analogy 36-65, is fairly solvent-exposed and not as intimately involved with the "aromatic cluster". The difference in ζ-C atom positions between the residues at position 32L in 36-71 and R19.9 is 5.7 Å (see Figure 8). A similar, though less dramatic, effect occurs at position 50L (see Figure 8). One aspect of these changes is to alter the relative position of HCDR3 in 36-71 as compared to R19.9 through the altered packing within the "aromatic cluster". This conclusion is based on the observed relative positions of the *base* of the hairpin loop of HCDR3, since the structure of this CDR varies significantly between

R19.9 and 36-71 at the *turn* in the hairpin loop (see Figure 4). In concert with the substitution at position 55L, in which the glutamine in 36-71 no longer hydrogen-bonds to the residue at position 46L as does histidine-55L in R19.9, and, analogously, 36-65, the substitution at position 32L alters the conformation of LCDR2, by moving it out of register one residue between the two structures (see Figure 8). It is unlikely that the substitution of glutamine in 36-71 for histidine as in 36-65 and R19.9 at position 55L in LCDR2 would have an effect on the observed affinity for hapten in and of itself.

By Use of the Structure of R19.9 To Model the Conformation of Residues in 36-65 at Positions Where the Sequences of 36-65 and 36-71 Diverge, Only 6 of the 19 Sequence Differences between 36-71 and 36-65 Have an Effect on the Structure Which Would Result in an Altered Affinity for Hapten. Of the 17 somatic mutations and 2 junctional differences between 36-71 and the germline sequence (36-65), only the 3 H-chain substitutions at positions 58, 59, and 107 appear to affect the three-dimensional structure of the molecule in such a way to increase affinity for hapten (see Figure 1). The only L-chain substitutions which might affect affinity are the three (32L, 50L, and 55L) which alter the structure of the "aromatic cluster", thereby altering the conformation of HCDR3. The structural change at the junctional position 100H has a negative effect on binding. The remaining mutations would not alter the conformation of the liganding residues, therefore having no direct or indirect effect on affinity. The structural observations are consistent with previous and concurrent binding studies of Fabs 36-71 and R19.9, as well as the experimental binding data for the H-chain substitutions produced by oligonucleotide-directed mutagenesis (see Table II). This agreement between the structural observations and the binding data supports the use of the structure of R19.9 as a "local" model for 36-65 in the regions where the sequences diverge.

The Utilization of either 36-65 or 36-71 L Chains Has a Pronounced Effect on the Affinity for Hapten Only When Certain Mutations Are Introduced into the 36-65 H Chain. The parent 36-71 and 36-65 L chains appear to have little effect on the observed affinity when associated with either the 36-71 or the 36-65 H chains (see Table II). The hypothesized effects from the structural comparison of 36-71 and R19.9 of the substitution of phenylalanines for tyrosines at positions 32L and 50L only become apparent when an intermediate set of mutations are made in the H chain of 36-65. A comparison of constructs containing multiple mutations at positions 58H, 59H, and 107H and those with mutations at positions 58H, 59H, 100H, and 107H shows the deleterious effect of a glutamic acid at position 100H. This residue is in a cleft behind the binding site between HCDR3 and HCDR1. The residue at this position is potentially capable of affecting the packing of the side chain of residue 50H, one of the residues contacting the hapten.[3] However, the magnitude of the effect of the substitution at position 100H is 2-fold when the two mutant H chains (Ile-58H/Thr-59H/Lys-107H or Ile-58H/Thr-59H/Glu-100H/Lys-107H) are associated with the 36-71 L chain, and 10-fold when associated with the 36-65 L chain. Clearly, the identity of the L chain has a major effect on affinity depending on the identity of the residue at position 100H. The best explanation is based on the structural observation that LCDR2 and LCDR3 affect the structure of HCDR3 through the residues at positions 32L and 50L (see above).[3]

The binding studies summarized in Table II were carried out with Ars-tyrosine (Sharon, 1990) and may include hapten/antibody interactions not identified in our model of the complex with the simpler molecule, phenylarsonate. However, the correlation between the hypothesized effects of the sequence substitutions based on the structural comparison and the experimentally observed binding constants is quite good.

The Interactions between Idiotype-Bearing Antibodies and Anti-idiotype Antibodies Are Determined by the Chemical and Physical Structure of the Solvent-Accessible Surface of the CDRs. There are two ways in which a mutation or gene junction difference can affect the antigenic surface of an Fab recognized by an anti-idiotype antibody. Changing the identity of a residue at a position in the sequence which is highly solvent-exposed in the three-dimensional structure can directly alter the chemical nature of the antigenic surface. Alternatively, substitutions of residues at positions which are "buried" in the structure can affect the SAS indirectly by altering the structure of other residues which are exposed to solvent. In considering the three-dimensional structural differences which are determined by differences in the sequence between Id^{CR+} and Id^{CR-} antibodies, locations with a high solvent accessibility do not play an major role since the side chains of these residues are interacting primarily with solvent molecules and not the rest of the protein. However, since it is the SAS which constitutes the antigenic surface, the nature of the residues at these positions becomes centrally important in discussions of idiotype. However, any given change in a highly solvent-exposed residue would not necessarily ablate reactivity with anti-idiotypic antibodies, nor is the converse necessarily true.

Clearly, the "ridge" surrounding the Ars-binding site depression is a major feature of the surface of the Fabs 36-71 and R19.9 and is composed of residues from HCDR2, HCDR3, LCDR2, and LCDR3, all regions of high solvent accessibility (see Figures 2 and 4). It was previously suggested based on correlation of sequence and serology that utilization of a canonical D gene segment (HCDR3) is tied to expression of Id^{CR} (Gridley et al., 1985; Meek et al., 1984) and the largest difference in the structures of 36-71 and R19.9 occurs in HDCR3. However, Parhami-Seren and co-workers (Parhami-Seren et al., 1990b) have shown, utilizing site-directed mutagenesis, that the loss of the cross-reactive idiotype (Id^{CR}) and several idiotopes defined by monoclonal anti-idiotypic antibodies is related to substitutions in HCDR2 (positions 52, 55, 58, and 59) independent of the D gene encoded segment. These results are consistent with the dominant role of HCDR2 in the SAS of the CDRs of 36-71 (see Table I). Furthermore, the single substitution of lysine for asparagine-55H (Parhami-Seren et al., 1990a) results in the loss of an idiotope defined by the monoclonal anti-idiotypic antibody 5Ci in the anti-Ars Id^{CR+} antibody 44-10 which differs in sequence from 36-65 at four sites. This residue is located on the first partially disordered turn in HCDR2 (see Figure 9). The three other sequence changes in 44-10 occur at residues 100H, 107H, and 110H, which are located on the "back" side of V_H in the structure of 36-71, and do not affect binding of the anti-idiotope antibody. An anti-idiotope antibody would have to "straddle" HCDR3 and HCDR2 to recognize both of these locations directly, covering a large portion of the V_H surface. In the structure of 36-71, this area is larger than the combining sites of most antibodies. In addition, of these four residues, 55H has the highest SAS in Fab 36-71 (see Figure 2; SAS of 55H = 41.1 Å^2, 100H = 19.7 Å^2, 107H = 13.5 Å^2, 110H = 23.0Å^2). Substitution of the "44-10"-like mutations at these latter three sites does not ablate 5Ci reactivity (Parhami-Seren et al., 1990a).

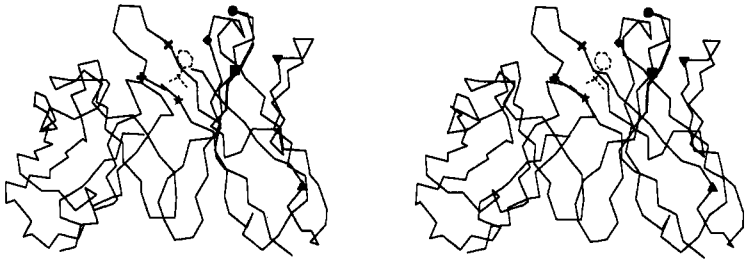

FIGURE 9: Variable domains of 36-71 are displayed in this stereodiagram; V_L on the left, V_H on the right. The hapten (phenylarsonate) is shown in *dashed* lines. The locations of residues which have been studied by site-directed mutagenesis to determine the effect of these sites on the binding of anti-idiotope antibodies as discussed in the text are identified as follows: 9H (▲); 32H (◆); 51H (■); 55H (●); 58 and 59H (*thick* lines); 77H (▼); 100H (×); 107H (+); 110H (★). See also Table III.

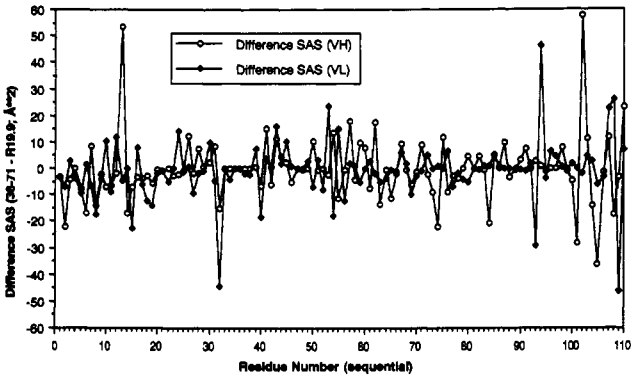

FIGURE 10: Difference in the solvent-accessible surfaces, calculated residue by residue, of part of the V domains of 36-71 and R19.9 is displayed (units of Å^2).

Smaller changes in the physical structure of the antigenic surface can be detected in plots of the difference in the solvent accessibility of each V domain residue between 36-71 and R19.9 (see Figure 10; we consider here only H-chain V residues 1–105 since the insertion in R19.9 occurs at this point). The average difference in solvent accessibility between residues in the H chains of 36-71 and R19.9 is 7.3 Å^2, in the L chains, the average difference is 6.5 Å^2, which indicates the overall similarity of the two structures. The curves also highlight regions where large variations occur in the detailed structures (see residues 15L, 32L, 52L, 93–94L, and 107–109L; 2H, 13H, 75H, 101–102H, and 105H in Figure 10 for the variations >20 Å^2). Residues at any of these positions of high differential accessibility, or positions indirectly affecting the accessibility of these residues, are potentially involved in defining the cross-reactive idiotype.

A study of the binding of four hapten-inhibitable anti-idiotope monoclonal antibodies (mAbs), which react with 36-65 but not with 36-71, to 36-65 engineered mutants (Sharon, 1990a) allows the mapping of idiotopes onto the surface of the 36-71 V_H (see Table III and Figure 9). In this study, single changes were made by oligonucleotide-directed mutagenesis in the 36-65 H chain which correspond to the mutations and junctional differences found in 36-71. The four anti-idiotope mAbs (107, MB, AI, and AD8) retain affinity for anti-Ars mutant antibodies despite changes in the framework residues 9H or 76H, which are distant from the hapten-binding site, and were therefore hypothesized as unlikely to be involved in defining these idiotopes. All anti-idiotope mAbs except MB react poorly or not at all with mutant anti-Ars antibodies which have Asn at position 32H, which is the only residue in HCDR1 that is appreciably accessible to solvent in R19.9 (and, by analogy, 36-65) and which occurs at a position in the sequence of the molecule where the SASs of 36-71 and R19.9 differ by 15.1 Å^2 (see Figure 10). All four anti-idiotypic mAbs also react poorly with a mutant

Table III: Percent Idiotype Expression of 36–71, 36–65, and Various 36–65 H-Chain Mutants Associated with 36–65 L Chains[a]

	anti-idiotype monoclonal antibodies			
	107	MB	AI	AD8
parental antibody				
36–65	100	100	100	100
36–71	<1	<1	<1	<1
hybrid antibody				
36–65H/36–65L	109	83	108	79
36–65 H-chain mutants				
Ala-9 → Val	75	73	73	110
Tyr-32 → Asn	7	41	<1	2
Ile-51 → Asn	120	81	76	88
Thr-58 → Ile	111	118	106	12
Lys-59 → Thr	109	142	109	135
Ser-76 → Asn	98	113	86	157
Val-100 → Glu	<1	12	<1	6
Tyr-107 → Lys	2	108	<1	139

[a] Percent idiotype expression is derived from inhibition radioimmunoassays and is expressed relative to idiotypes expressed by antibody 36–65 (Sharon, 1990a).

antibody which has a glutamic acid at position 100H which is located near the hapten in the modeled complex, like the residue at position 32H. The substitution of glutamic acid for valine at this position represents a drastic "chemical" change in the surface of 36-65. The residue at position 51H is buried, and a substitution here is not detected by any of these anti-idiotope mAbs. Residues 58H and 59H are moderately solvent-exposed, and are further from the hapten-binding site than residues 32H, 51H, 100H, and 107H. A substitution at 58H affects the binding of anti-idiotypic mAb AD8 only while substitution at 59H does not affect idiotypy. Thus, large changes in SAS for certain positions (58H and 59H) need not result in changes in idiotypy. In a previous study (Haba et al., 1989), it was proposed that residues 51H, 56H, and 58–59H were important for recognition by AD8.

The side chain of residue 107H, while solvent-exposed, projects out behind the V_H domain on the other side of HCDR3 from the hapten-binding site. Since it seems unlikely that an anti-idiotope antibody could directly interact with residues on both sides of HCDR3, the loss of binding observed with the anti-idiotypic mAbs 107 and AI when a lysine is substituted for a tyrosine at this position is probably due to the indirect effect this mutation has on the conformation of Tyr-106H, a ligand to the hapten. Since this mutation also has a profound effect on the affinity for hapten (see Table II), and MB and AD8 still bind to this mutant of 36-65 (Tyr-107H → Lys), the result of this sequence change demonstrates that anti-idiotypic antibodies are not necessarily "images" of the antigen. These results demonstrate that idiotype, as well as hapten binding, can be altered through indirect changes to the SAS, and not only through direct changes in surface residues.

ACKNOWLEDGMENTS

We thank Dr. Tom Smith (Purdue) for supplying his programs MacFrodo and MacPrep with which many of the figures were prepared and Dr. Robert Campbell (National Research Council Canada) for his advice.

REFERENCES

Alzari, P. M., Lascombe, M.-B., & Poljak, R. J. (1988) *Annu. Rev. Immunol. 6*, 555.

Bentley, G. A., Bhat, T. N., Boulot, G., Fischmann, T., Navaza, J., Poljak, R. J., Riottot, M.-M., & Tello, D. (1989) *Cold Spring Harbor Symp. Quant. Biol. 54*, 239.

Bona, C. (1981) *Idiotypes and Lymphocytes*, Academic Press, New York.

Connolly, M. (1983) *Science 221*, 709.

Dodson, G. G. (1981) in *Refinement of Protein Structure* (Machin, P. A., Campbell, J. W., & Elder, M., Eds.) p 95, Daresbury, London.

Estess, P., Nisonoff, A., & Capra, J. D. (1979) *Mol. Immunol. 16*, 1111.

Greene, M. I., & Nisonoff, A., Eds. (1984) *The Biology of Idiotypes*, Plenum Publishing Corp., New York.

Gridley, T., Margolies, M. N., & Gefter, M. L. (1985) *J. Immunol. 134*, 1236.

Haba, S., Lascombe, M.-B., Poljak, R. J., & Nisonoff, A. (1989) *J. Exp. Med. 170*, 1075.

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

Kuettner, M. G., Wang, A. L., & Nisonoff, A. (1972) *J. Exp. Med. 135*, 579.

Lamoyi, E., Estess, P., Capra, J. D., & Nisonoff, A. (1980) *J. Immunol. 124*, 2834.

Landolfi, N. F., Capra, J. D., & Tucker, P. W. (1986) *J. Immunol. 137*, 362.

Lascombe, M.-B., Alzari, P. M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A., & Poljak, R. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 607.

Margolies, M. N., Marshak-Rothstein, A., & Gefter, M. L. (1981) *Mol. Immunol. 18*, 1065.

Marshak-Rothstein, A., Benedetto, J. D., Kirsch, R. L., & Gefter, M. L. (1980a) *J. Immunol. 125*, 1987.

Marshak-Rothstein, A., Siekevitz, M., Margolies, M. N., Mudgett-Hunter, M., & Gefter, M. L. (1980b) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1120.

Marshak-Rothstein, A., Margolies, M. N., Benedetto, J. D., & Gefter, M. L. (1981) *Eur. J. Immunol. 11*, 565.

Meek, K. D., Jeske, M., Slaoui, O., Leo, T., & Capra, J. D. (1984) *J. Exp. Med. 160*, 1070.

Parhami-Seren, B., Wysocki, L. J., & Margolies, M. N. (1989) *J. Immunol. 143*, 409.

Parhami-Seren, B., Sharon, J., & Margolies, M. N. (1990a) *J. Immunol. 144*, 4426.

Parhami-Seren, B., Wysocki, L. J., Margolies, M. N., & Sharon, J. (1990b) *J. Immunol. 145*, 2340.

Rose, D. R., Strong, R. K., Margolies, M. N., Gefter, M. L., & Petsko, G. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 338.

Sanz, I., & Capra, J. D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1085.

Sharon, J. (1990a) *J. Immunol. 144*, 4863.

Sharon, J. (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4814.

Sharon, J., Gefter, M. L., Wysocki, L. J., & Margolies, M. N. (1989) *J. Immunol. 154*, 596.

Siegelman, M., & Capra, J. D. (1981) *Immunoglobulin Idiotypes* (Janeway, C., Sercarz, E., & Wigzell, H., Eds.) p 135, Academic Press, New York.

Siekevitz, M., Gefter, M. L., Brodeur, P., Riblet, R., & Marshak-Rothstein, A. (1982) *Eur. J. Immunol. 12*, 1023.

Siekevitz, M., Huang, S. Y., & Gefter, M. L. (1983) *Eur. J. Immunol. 13*, 123.

Slaughter, C. A., & Capra, J. D. (1983) *J. Exp. Med. 158*, 1615.

Strong, R. K., Campbell, R., Rose, D. R., Petsko, G. A., Sharon, J., & Margolies, M. N. (1991) *Biochemistry* (preceding paper in this issue).

Wysocki, L. J., Manser, T., & Gefter, M. L (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1847.

Wysocki, L. J., Gridley, T., Huang, S., Grandea, A. G., & Gefter, M. L. (1987) *J. Exp. Med. 166*, 1.