Modeling the antigen combining site of an anti-dinitrophenyl antibody, ANO2

DONNA BASSOLINO-KLIMAS, ROBERT E. BRUCCOLERI,
AND SHANKAR SUBRAMANIAM¹

Department of Macromolecular Modeling, Bristol-Myers Squibb Pharmaceutical Research Institute,
Princeton, New Jersey 08543

(RECEIVED December 17, 1991; REVISED MANUSCRIPT RECEIVED July 1, 1992)

Abstract

A model structure has been constructed for a monoclonal anti-dinitrophenyl antibody. The antibody, ANO2, has been sequenced and cloned (Anglister, J., Frey, T., & McConnell, H.M., 1984, *Biochemistry* 23, 1138–1142). Its amino acid sequence shows striking homology with the anti-lysozyme Fab fragments HyHel5 (83%) and HyHel10 (73%). Based on this homology, a model for the ANO2 variable heavy and variable light chain framework was constructed using a hybrid of the HyHel5 light chain and the HyHel10 heavy chain backbone, omitting the hypervariable loops. These coordinates were used as scaffolds for the model building of ANO2.

The CONGEN conformational sampling algorithm (Brucoleri, R.E. & Karplus, M., 1987, *Biopolymers* 26, 127–196) was used to model the six hypervariable loops that contain the antigen-combining site. All the possible conformations of the loop backbones were constructed and the best loop structures were selected using a combination of the CHARMM potential energy function and evaluation of the solvent-accessible surface area of the conformers. The order in which the loops were searched was carried out based on the relative locations of the loops with reference to the framework of the β -barrel, namely, L2-H1-L3-H2-H3-L1. The model structures thus obtained were compared to the high resolution X-ray structure (Brünger, A.T., Leahy, D.J., Hynes, T.R., & Fox, R.O., 1991, *J. Mol. Biol.* 221, 239–256).

Keywords: antibody modeling; canonical structures; conformational search; homology modeling; hypervariable loops

Antibodies of the IgG class consist of heavy and light polypeptide chains linked by disulfide bonds and folded into independent β -sheet domains. The light chain is approximately 220 residues and contains a variable N-terminal domain (VL) and a constant C-terminal region. The heavy chain is made up of an N-terminal variable domain (VH) and three or four constant domains. The heavy chain can contain from 450 to 575 residues. The fragment containing the noncovalent dimer of VL and the VH domains is called the Fv. The VL and VH are each antiparallel β -barrels, and the contact between them forms an eight-stranded β -barrel with four strands from each domain (Richardson, 1981; Novotny et al., 1983). In each variable region, three hypervariable loops are attached to a constant β -sheet framework region (FR). The specificity

of the antibody is determined by the hypervariable loops that form the surface of the antigen-binding sites. These six hypervariable loops are also referred to as complementarity determining regions (CDRs). Although the sequences of several thousand antibodies have been determined (Kabat et al., 1987), the structures of only a few dozen antibodies have been determined by X-ray crystallography. However, there is a high degree of homology in the tertiary structure of immunoglobins as well as in their primary sequence. The structural and functional diversity of antibodies arises from sequence diversity on only about 10% of the molecule (Wu & Kabat, 1970). For this reason homology modeling is a logical choice for antibodies and several model building studies have been attempted (Kabat & Wu, 1972; Chothia et al., 1986; de la Paz et al., 1986; Fine et al., 1986; Snow & Amzel, 1986; Shenkin et al., 1987; Bruccoleri et al., 1988; Kussie et al., 1991; Anchinn et al., 1992; Davies et al., 1990, and references therein; Nell et al., 1992). It has become apparent that the tertiary structures of the nonhypervariable or framework

Reprint requests to: Donna Bassolino-Klimas, Department of Macromolecular Modeling H3812, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543.

¹ Present address: Department of Physiology and Biophysics, Beckman Institute, University of Illinois at Urbana, Urbana, Illinois 61801.

regions are very similar. Thus, most modeling attempts assume the same framework for the VL and VH domains as that found in the crystal structures (Davies et al., 1990). Several computational methods have been used to model the hypervariable loops. One approach uses known CDR structures as templates. The other tries to predict the conformation based on energetics.

Recently, analysis of known structures by Chothia et al. (1989) revealed that there are a small number of main chain conformations found for five of the six hypervariable loops. Evidence indicates that a few conserved residues determine the canonical conformation. Kabat et al. (1977) noted that there are conserved residues even within the hypervariable loops. Data-base search and template prediction methods were devised based on the canonical structure concept of Chothia (Martin et al., 1989). Chothia et al. (1986) modeled the CDRs of the antibody D1.3 based on canonical structure motifs before the experimental structure determination was completed. Comparison with the crystal structure showed that four of the six loops were correctly predicted.

The alternative method of reconstructing the hypervariable loops based on energetics was first attempted by Stanford and Wu (1981), who constructed a model of the backbone structure of the combining site of MOPC315, a phosphorylcholine binding antibody (Dower et al., 1977). Their predictions cannot be assessed because the structure of MOPC315 Fv is not yet available. Fine et al. (1986) randomly generated a large number of conformations for the backbone of the loop followed by either molecular mechanics or molecular dynamics. The lowest energy conformation was then selected. They applied this to four of the CDRs in McPC603, another phosphorylcholine antibody, and obtained root-mean square deviations (RMSDs) of about 1.0 Å with the crystal structure. More recently, Brucoleri et al. (1988) reconstructed the

CDRs in HyHel5 and McPC603 by conformational search. Loop conformations are generated by sampling the energetically favorable regions of torsion angle space and imposing the conditions that the loop ends fit onto the framework and bad steric contacts are avoided. The backbone RMSDs of the loops were 1.4 Å for HyHel5 and 1.7 Å for McPC603. The all atoms RMSDs were 2.6 Å and 2.4 Å for HyHel5 and McPC603, respectively.

In this study the dinitrophenyl (DNP) antibody, ANO2, will be modeled using a combination of homology modeling and conformational sampling. The model will then be examined using canonical structure analysis.

Results and discussion

To determine the best Fab structure to use in the homology modeling, the sequences of ANO2 heavy and light chains were compared to those of 10 Fab structures whose high resolution X-ray coordinates were available. The sequence alignment is shown in Figures 1 and 2. These results show that the VH chain of ANO2 is 73% homologous to the VH chain of HyHel10, and the VL chain is 83% homologous to the VL chain of HyHel5. These homologies are higher than those between most previously modeled antibodies. Thus a hybrid of HyHEL5 and HyHel10 should be used to model ANO2. Within the loops, from 38 to 83% of the residues are homologous.

The 2.54-Å resolution X-ray coordinates for HyHel5 (Sheriff et al., 1987) and 3.0-Å resolution coordinates for HyHel10 (Padlan et al., 1989) were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977) (2HFL and 3HFM, respectively). Coordinates of the ANO2 Fv framework were derived in two alternative ways. One hybrid structure was created by superimposing the light chain of HyHel10 (H10L) onto the light chain

Fig. 1. Sequence alignment of the light chain with other known Fabs. The loop regions are labeled. Uppercase, aligned non-identical residues; lowercase, unaligned residues; — — —, aligned identical residues; . . . ., gap in alignments.

	1	H1	H2	58
ANO2	DVQLQESGPGLVKPSQSLSLTCTVTGYSITSDYAWNWIQFPNGKLEWMGYMSYSGS.T.			
HY10	-----S-----TL-----S-----D-----S-----K-----R-----Y-----V-----Y			
D1.3	Q---K-----A-----L-I-----S-F-L-GYG.V--V--P--KG---L-MIWGD-Nt..			
HY5	-----Q--AE-M--GA-VKIS-KAS--TFSDYW.IE-VK-R--HG---I-EILPGSGst.			
J539	E-K-L---G---Q-GG-LK-S-AAS-FDFSKYW.MS-V--A--KG---I-EIHPDSGti.			
KOL	E---VQ---G-V-Q-GR-LR-S-SSS-FIFS-YA.MY-V--A--KG---VAIIWDD--dq.			
MCPC	E-K-V---G---Q-GG-LR-S-ATS-FTFSDF-.ME-V--P--KR---IAASRNK-N.Ky			
NC41	QI--VQ---K-K--GETVKFS-KAS--TF-NYG.M--VK-A--KG-----WINTNTGep.			
NEWM	Q---EQ-----R---TL-----S-STFSN--YT-V--P--RG---I--VF-H-T.Sd			
	59	H3	105	
ANO2	..RYNPSLRSRISITRDTSKNQFFLQLKSVTTEDTATYFCARG.....WPLAYW			
HY10	...-----K-----YY-D-N-----Y--N-----DGD--			
D13	..D--SA-K--L--SK-N--S-V--KMN-LH-D--R-Y--Erdyrl.....D--			
HY5	..N-HERFKGKATF-A---SSTAYM--N-L-S--SGV-Y-LH-nydfd.....G....			
J539	..N-T---KDKFI-D--NA--SLY--MSK-RS---L-Y---Lhyy.....GYN--			
KOL	..H-AD-VKG-FT-S-ND---TL---MD-LRP---GV-----DgghgfcssascFGPD--			
MCPC	ttE-SA-VKG-FIVS---QSILY--MNALRA---I-Y---Nyygst.....YFDV-			
NC41	..T-GEEFKG-FAFSLE--ASTAN--INNLRN--K--F-----ednfg.....SLSD--			
NEWM	dtT...P---VTMLV-----S-R-S---AA---V-Y---Nliag.....CIDV-			
	106	115		
ANO2	GQGTQVSVSE			
HY10	----L-T--A			
D13	----TLT--S			
HY5	----TLT--S			
J539	----L-T--A			
KOL	----P-T--S			
MCPC	-A--T-T--S			
NC41	----TLT--S			
NEWM	----SL-T--S			

Fig. 2. Sequence alignment of the heavy chain with other known Fabs. The loop regions are labeled. Uppercase, aligned non-identical residues; lowercase, unaligned residues; ---, aligned identical residues; . . . ., gap in alignments.

of HyHel5 (H5L) using a least-squares fit, then splicing the heavy chain of HyHel10 (H10H) and H5L into one coordinate set. This model will have the characteristics (i.e., elbow angle and dyad axis) of HyHel10 and will be called model structure A. Conversely, a second hybrid was created by superimposing the heavy chain of HyHel5 (H5H) onto H10H and then splicing the H5L chains and H10H chains into another coordinate set. This model will have the characteristics (i.e., elbow angle and dyad axis) of HyHel5 and will be called model structure B. In both cases backbone coordinates of the parent molecule were retained. Conserved side chain coordinates were also retained in the ANO2 framework. Thirty-nine amino acid replacements and six amino acid insertions were required. The amino acid replacements are made using the SPLICE feature of CONGEN. The gaps created by residue insertions and deletions can then be closed by sampling the conformational space of the surrounding residues keeping the rest of the molecule fixed. The new side chains were then sampled with CONGEN. This is done by iteratively varying each of the χ angles 60° for Lys and Arg and 30° for all other amino acids. The conformation of lowest energy was saved to be used in later calculations. Visual inspection of the two initial complexes showed that all the amino acid substitutions are on the exterior of the protein and are highly exposed.

These two coordinate sets, omitting the coordinates for atoms in the hypervariable loop regions, were then used as the framework for loop structure generation. Because

loops L2 and H2 are nearly identical in sequence to H5L loop L2 and H10H loop H2, respectively, only the substituted side chains were sampled for these loops. The other loops were sampled in the order H1, L3, H3, L1. Because L3 and H1 are closer to the midpoint of the β -barrel interface and do not interact with each other, these loops are built first. Loops H3 and L1 interact closely and are built last. This is the same protocol followed by Bruccoleri et al. (1988) in the modeling of McPC603.

In order to model the hypervariable loops, a search is performed over the entire conformational space of the residues involved in the loop. The backbone and side chain torsions are sampled together. The program employs a modified Gō and Scheraga (1970) algorithm to assure chain closure (Bruccoleri & Karplus, 1985). Energy considerations also provide many constraints that eliminate conformations. Because the atoms in the remainder of the protein contribute to the energy of each conformation, any partial conformation that results in a strong repulsive contact will be discarded, thereby eliminating many possible conformations.

The best conformation of each loop based on a combination of low energies (as evaluated by the CHARMM potential function) and low solvent-accessible surface areas was selected to be incorporated into the final model in the following way. First, from the set of conformations generated for each loop, the conformation with the lowest energy was found. Then, the solvent-accessible surface area was calculated for all conformations within 3.0 kcal/mol

of the lowest energy conformation. The best conformation (i.e., lowest energy with lowest solvent-accessible surface area) was selected to be incorporated into the model structure. This structure was then used as the starting structure for the modeling of the next loop.

Modeling the loops

Like all other antibodies, ANO2 contains six hypervariable loops. These loops include three regions on the light chain, residues 26–31 (L1), residues 50–56 (L2), and residues 91–97 (L3), and three regions on the heavy chain, residues 26–34 (H1), residues 50–56 (H2), and residues 98–104 (H3). These loop regions are indicated on Figures 1 and 2. In this paper, all residues are in terms of the sequential numbering of the ANO2 sequence. The conversion table to the Kabat numbering is given in Brünger et al. (1991).

Loops L2 and H2

Loop L2 incorporates residues 50–56 on the light chain of ANO2. This loop is identical to that in H5L except for the replacement of Lys by Asn-52 in the ANO2 sequence. Loop H2 incorporates residues 50–56 on the heavy chain of ANO2 and is identical to loop H2 in H10H except for the replacement of Met for Val-52. In these cases only the side chain conformation of the replaced residue was sampled using a 30° grid. All other backbone and side chain atoms were fixed and identical to the parent coordinates in H5L and H10H. The best conformation in each case was chosen to be incorporated into the model.

Loop H1

Loop H1 is a long loop of nine residues from Gly-26 to Ala-34. This loop includes two residues not found in H10H. The Asp in H10H has been replaced by Tyr-27 in ANO2 and an Ala-34 has been inserted between the Tyr and the Trp. This loop was found to be too long for a single conformational search. Instead the loop was divided into two overlapping segments (residues 26–30 and 29–34). Each segment was sampled using a 30° grid and the best backbone conformation was selected and built into the model. Then all the side chains in this loop were sampled on that backbone.

Loop L3

Loop L3 is a short loop from residue Ser-91 to Thr-97. In ANO2, this loop differs from H5L by the insertion of two residues Pro-95 and Ile-96 into the sequence. This insertion results in two consecutive prolines, which make this loop highly strained. In this case cis/trans isomerization of the prolines was permitted. This loop was initially sampled with a 30° grid for all residues, but no low energy structures were obtained. The sampling was repeated with a 15° grid for proline, and acceptable conformations were found.

Loop H3

Loop H3 is a short loop from Arg-98 to Tyr-104 on the heavy chain. This loop is very different from that in the parent H10H chain. Only two residues, Trp-100 and Tyr-104, are the same. It also contains a proline not found in other H3 loops. Thus, the whole loop was sampled using a 30° grid as in loop H1. However, no low energy conformations were found so the search was repeated with a 15° grid. The best conformation was selected to be built into the final model.

Loop L1

Loop L1 is a short loop incorporating residues Ser-26 to Tyr-31. This loop is the same as H5L except for the replacement of Asn by Tyr-30 in the ANO2 sequence. However, because this loop interacts closely with loop H3 all six residues were sampled using a 30° grid. The best conformation was selected to be built into the final model.

Analysis of the loops

After generating a final structure containing all the new loop conformations, each loop was minimized for 10 steps of steepest descents minimization to reduce any residual steric strain with nearby atoms. The RMSD before and after minimization was 0.02 Å. The resulting structures were then used for the following analysis. Figure 3a,b shows space filling views of model structures A and B, respectively. Superimposition of the two modeled structures indicates they are very similar. The RMSD of the backbone atoms is 0.26 Å for the light chain and 0.46 Å for the heavy chain between the two model structures. The RMSD of the backbone of the whole complex taken at once is 0.71 Å. This larger RMSD is attributable to the difference in VL/VH orientations of the starting scaffold. The RMSD between the loops in the two model structures are given in Table 1. Table 2 shows the energies and solvent-accessible surface areas for the selected conformations in each model structure as well as for the parent HyHel5 and HyHel10 structures.

Loops L2, L3, H2, and H3 connect adjacent strands of antiparallel β -sheet. L1 and H1 connect strands of differ-

Table 1. Root mean square deviations (Å) between model structure A and model structure B

Loop	Backbone	All atoms
L1	0.05	0.06
L2	0.04	0.11
L3	0.77	1.28
H1	1.02	2.76
H2	0.08	0.11
H3	0.87	1.97

Fig. 3. **a, b:** Model structures A and B, respectively. The loops are color coded as follows: L1, red; L2, magenta; L3, orange; H1, blue; H2, cyan; H3, green.

ent sheets within the VL and VH domains, respectively. It has been suggested by Chothia et al. (1989) that antibodies have only a few main chain conformations or canonical structures for some of the hypervariable loops

Table 2. *Potential energies of the loops and solvent-accessible surface areas of the loops*

Structure	L1	L2	L3	H1	H2	H3
Potential energies (kcal/mol)						
HyHel5-L	-30.6	-28.3	-54.0			
HyHel10-H				-46.1	-33.1	-54.9
Model A	-39.6	-37.5	-30.3	-60.6	-20.5	-35.2
Model B	-39.7	-38.2	-37.6	-46.1	-34.4	-34.6
Solvent-accessible surface areas (Å ²)						
HyHel5-L	478.9	390.8	378.7			
HyHel10-H				473.1	338.8	260.8
Model A	567.8	342.1	422.7	571.6	367.7	332.5
Model B	568.1	362.0	452.5	588.3	365.0	282.1

despite differences in sequence. Evaluation of the loops in the ANO2 model shows that a majority of these key structural features and interactions are present in the model of ANO2. Figures 4 and 5 show each loop oriented to maximize the view of the turn involved. Figure 6 is a stereo view showing the loops oriented as they are in the antibody structure.

The crystal structure of the Fab fragment of the murine ANO2 complexed with its hapten has been solved at 2.9 Å resolution using a novel molecular replacement method and the 2.5-Å resolution crystal structure of HyHel5 as a search model (Brünger et al., 1991). Preliminary coordinates of the crystal structure were obtained from the authors. Comparison to the model will be based on this structure as well as their discussion. For model structure A, the RMSD with the X-ray structure is 2.49 Å and 2.07 Å for the heavy and light chains, respectively. For model structure B, the RMSD is 2.51 Å for the heavy chain and 2.00 Å for the light chain. Table 3a,b shows the RMSD for the two model structures with the X-ray structure on an atom-by-atom basis for the residues in the loops.

In the light chain L1 connects strands in two different β -sheets and straddles the VL domain. HyHel5, REI (Epp et al., 1975), and McPC603 (Satow et al., 1986) have large hydrophobic residues (Val, Ile, and Leu, respectively) in the fourth position that pack into the cavity formed by the loop. The main chain N of this residue also hydrogen bonds to nearby Gly-68. A similar conformation is found in the L1 loop of ANO2. The loops are virtually identical in the two model structures (see Table 1). This loop consists of two turns as shown in Figure 4a for model structure A. The hydrophobic side chain of Val-29 points into the cavity formed by the loop between the β -sheets and is buried. The main chain N of Val-29 makes a hydrogen bond with Gly-68. This loop is consistent with the class I canonical structure described by Chothia and Lesk for the L1 loop. Analysis of the X-ray structure by Brünger et al. (1991) also places this loop in canonical class 1 with HyHel5 and J539, having a tight turn of type I (Richardson, 1981). The hydrophobic side chain of Val-29 pointing into the cavity can be observed in the X-ray structure as well. The only difference between the model structures and the X-ray structure is a kink in the backbone at Tyr-30 that displaces the side chain somewhat. However, the surrounding residues have a low RMSD to the X-ray structure (see Table 3).

Loop L2 is similar in all other known structures (Chothia & Lesk, 1987). Similarities include a three-residue hairpin turn as well as conservation of the framework region into which the loop packs. Loop L2 is shown in Figure 4b for model structure A. This loop is also virtually identical in both model structures as well as the X-ray structure.

Loop L3 is longer than the L3 loop of any previously resolved Fab structures. The crystallographers suggest that

Fig. 4. Hypervariable loops oriented to maximize the view of the turn involved. a: L1. b: L2. c: L3.

the L3 loop of ANO2 may represent a new class of L3 loops. This loop contains two consecutive prolines that highly constrain the loop, and it is slightly different in the two model structures both of which are shown in Figure 4c. However, in both models, this loop maintains many of the key structural features found in other anti-

bodies. It is a medium-sized loop with H-bonds formed by the inward pointing main chain polar atoms (i.e., Ser-91). This type of interaction was noted in REI and McPC603 by Tramontano et al. (1989). REI and McPC603 have a *cis*-proline, which prevents the formation of a tight four-residue turn. This loop in ANO2 is more unusual in having

Fig. 5. Hypervariable loops oriented to maximize the view of the turn involved. a: H1. b: H2. c: H3.

two sequential prolines, which also place residues Tyr-93–Pro-94–Pro-95 in an extended conformation and prevent the formation of a tight four-residue turn. The backbone of residues 95–97 in the X-ray structure are very similar

to those in the models, whereas residues 91–94 are very different (see Table 3). However, this loop in the X-ray structure should not be compared to the model due to the disordered residues (Tyr-93 and Pro-94) whose conforma-

Fig. 6. Hypervariable loops oriented as they are in the antibody structure. **a:** Model structure A. **b:** Model structure B. The loops are color coded as in Figure 3. (See also Kinemage 1.)

tion in the X-ray structure is uncertain (Brünger et al., 1991). These residues have *B*-factors greater than $55 \text{ \AA}^2$.

In the heavy chain hypervariable loops, the conserved conformations are also found. Loop H1 connects strands in two β -sheets of the VH domain and is similar to L1. This loop is also longer in ANO2 than in previously solved Fab structures. The hydrophobic side chain of Ile-29 packs into the cavity between the two sheets and there is an H-bond between the main chain nitrogens of Asp-32 and Tyr-33 and the carbonyl oxygens of Ser-28 and Ile-29. Gly-26 forms a sharp turn. Figure 5a shows that residues 26–30 are different in the model structures. However, the overall patterns are consistent with those found in other antibodies (Chothia & Lesk, 1987). The X-ray structure shows similar packing of the Ile-29 side chain, which is buried deeply in the framework. The sequence and length of this loop places it in canonical class I.

Loop H2 characterizes a tight turn with typical internal loop H-bonding patterns. The loops are identical in the model structures. The carbonyl oxygen of Ser-57 is H-bonded to the main chain N of Ser-53 (Fig. 5b). The same H-bonding pattern is found in the X-ray structure.

Loop H3 has a distorted two-residue hairpin turn and does not conform to usual β -turns. The size, sequence, and conformation of this loop varies the most widely in all antibodies. In ANO2 very few of the side chains are the same as those in HyHel5, HyHel10, or McPC603; only Arg-98 and Trp-100 are conserved. In HyHel5 Phe-102 is found in a non-allowed conformation. The equivalent

residue is Gly-100 in ANO2, which is also in an unusual conformation in our models. The Arg-98 would normally form a salt bridge with an Asp in position 102 in HyHel10 and McPC603. However, this residue is an Ala in ANO2, so salt bridge formation is not possible. The presence of Pro-101 limits conformational space. Figure 5c shows the best conformation obtained for this loop. This loop in the X-ray structure also has a distorted β -hairpin. However, the backbone conformations are very different as evidenced by the RMSDs in Table 3. This loop is not discussed in any detail their paper. However, the authors (Brünger et al., 1991) indicate that the placement of the hapten required rebuilding of the H3 loop. This rebuilding could account for the differences in the backbone structure.

The model structures shown in Figure 3 show a number of exposed tyrosine residues. This is also consistent with the X-ray structure and NMR structure (Theriault et al., 1991), which reveal 8–10 exposed residues in the combining site. The X-ray structure, which included the DNP-hapten, indicates that the DNP is sandwiched between Trp-90 on the light chain and Trp-100 on the heavy chain. This is confirmed by the NMR studies. Our models show similar stacking of these Trp residues. However, Arg-98 is sandwiched between the tryptophans in both models. It is not clear whether this residue is misplaced or if the arginine sits in the pocket in the absence of the DNP. It is also not clear if the rebuilding of the H3 loop in the X-ray structure included moving the arginine.

Table 3a. Root mean square deviations (RMSDs) between the model structure A and the X-ray structure on an atom-by-atom basis for residues in the loops

(continued)

Table 3a. *Continued*

Residue	N	CA	CB	OG	C	O	CG	CD1	CD2	CE1	CE2	CZ	OH	CD	CG1	CG2	OG1	RMSD
Ser-91	2.9	3.4	4.0	1.9	3.5	3.6												3.3
Ser-92	3.6	3.8	4.7	4.1	3.3	4.9												4.1
Tyr-93	2.2	3.0	3.7		2.4	2.2	6.0	6.9	7.7	9.2	9.9	10.5	12.9					7.3
Pro-94	2.6	2.1	2.2		2.7	4.9	1.8							3.1				2.9
Pro-95	1.6	0.5	1.4		0.6	0.7	2.5							2.4				1.6
Ile-96	0.8	0.9	0.8		0.8	1.0								4.2	2.5	2.5		2.1
Thr-97	0.7	0.6	0.6		0.6	0.7										0.6	0.5	0.6
	RMSD—back			RMSD—side			RMSD—all											
Total L3	2.6			5.4			4.2											

^a RMSD—back, includes atoms N, HN, CA, C, and O only; RMSD—side, includes all side chain atoms; RMSD—all, includes all atoms.

Conclusion

In this study we used homology modeling and conformational sampling to model the three-dimensional structure of ANO2. Evaluation of the modeled loops show that a majority of the canonical features and interactions studied by Chothia and Lesk are present. The structure is very similar to the X-ray structure for all loops except H3. Thus, conformational sampling is a useful tool for modeling antibody hypervariable loops. The program CONGEN is available from Dr. R.E. Bruccoleri.

Materials and methods

The ANO2 heavy and light chain sequences of the Fab fragment were obtained from Leahy et al. (1988). X-ray coordinates of the high resolution Fab crystal structures were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977). The sequences were compared using the program EUGENE (Molecular Biology Resource, Baylor College of Medicine, Houston, Texas). Figures were obtained from the INSIGHT graphics package (Biosym Technologies). All other calculations were per-

Table 3b. *Root mean square deviations (RMSDs) between the model structure B and the X-ray structure on an atom-by-atom basis for residues in the loops*

Residue	N	CA	C	O	CB	CG	CD1	CD2	CE1	CE2	CZ	OH	OG	CG1	CG2	CD	OG1	OD1	OD2	RMSD
Gly-26	1.0	1.2	1.5	3.2																1.9
Tyr-27	1.6	1.8	1.9	3.1	3.0	5.1	5.7	7.1	8.0	9.0	9.3	11.5								6.5
Ser-28	2.2	2.4	1.9	2.6	3.9								4.4							3.0
Ile-29	2.6	3.4	2.4	3.2	6.1									7.3	7.3	8.5				5.6
Thr-30	1.1	1.8	2.2	3.6	2.0										2.5		1.4			2.2
Ser-31	1.8	3.0	2.3	4.0	4.4								5.8							3.8
Asp-32	1.1	1.2	0.9	2.9	2.3	4.9												6.3	6.2	3.8
Tyr-33	1.6	1.0	1.2	2.3	0.6	0.7	0.6	1.1	0.8	1.3	1.1	1.3								1.2
Ala-34	1.4	1.4	1.0	1.0	1.9															1.9
		RMSD—back ^a			RMSD—side			RMSD—all												
Total H1		2.2			5.3			4.0												
Residue	N	CA	C	O	CB	CG	CD1	CD2	CE1	CE2	CZ	OH	SD	CE	OG	RMSD				
Gly-50	0.6	0.7	0.7	0.8												0.7				
Tyr-51	0.7	0.7	0.7	0.8	0.5	0.6	0.5	0.9	0.7	1.1	0.9	1.1				0.8				
Met-52	0.9	0.7	0.7	0.7	0.6	1.1							2.2	2.8		1.4				
Ser-53	0.7	1.0	1.8	2.2	1.3										1.8	1.6				
Tyr-54	2.2	2.9	3.3	5.3	3.3	3.5	3.5	4.8	3.7	5.1	4.0	4.3				3.9				
Ser-55	2.3	2.5	2.1	1.8	2.5										3.3	2.5				
Gly-56	2.2	1.8	1.9	2.5												2.1				
		RMSD—back			RMSD—side			RMSD—all												
Total H2		1.9			2.7			2.3												

(continued)

Table 3b. Continued

Residue	N	CA	CB	CG	CD	NE	CZ	NH1	NH2	C	O	CD1	CD2	NE1	CE2	CE3	CZ2	CZ3	CH2	CE1	OH	RMSD
Arg-98	3.1	4.7	6.5	8.9	8.9	11.2	11.3	11.7	11.3	3.9	3.9											8.4
Gly-99	4.2	5.3								5.1	6.6											5.4
Trp-100	4.0	4.1	4.0	4.2						4.0	4.8	5.4	4.5	5.3	5.1	5.6	6.3	7.5	7.6			5.3
Pro-101	3.7	3.8	4.9	4.9	4.3					3.9	3.5											4.2
Leu-102	5.7	6.4	9.1	11.5						4.4	3.8	14.0	12.4									9.1
Ala-103	3.5	1.9	3.1							1.8	3.7											2.9
Tyr-104	0.6	0.9	1.5	1.7			2.2			0.6	1.0	1.3	3.0		3.4					1.2	2.6	1.9
RMSD—back RMSD—side RMSD—all																						
Total H3	4.0		7.2		6.0																	
Residue	N	CA	CB	OG	C	O	CG1	CG2	CG	CD1	CD2	CE1	CE2	CZ	OH	RMSD						
Ser-26	0.5	0.5	0.6	2.9	0.8	1.7										1.5						
Ser-27	0.3	0.5	0.8	2.8	1.0	3.0										1.8						
Ser-28	0.8	0.6	1.5	3.7	0.5	0.5										1.7						
Val-29	0.9	1.5	1.3		2.0	2.2	1.6	0.7								1.5						
Tyr-30	2.6	3.3	4.4		2.5	2.8			5.4	8.0	3.9	9.0	4.8	7.4	8.5	5.7						
Tyr-31	1.7	0.9	0.6		0.6	1.0			0.9	0.8	1.5	1.2	1.8	1.6	2.0	1.3						
RMSD—back RMSD—side RMSD—all																						
Total L1	1.6		4.1		3.1																	
Residue	N	CA	CB	OG1	CG2	C	O	OG	CG	OD1	ND2	CD1	CD2	RMSD								
Thr-50	0.9	0.4	0.5	1.0	0.8	0.3	0.4							0.7								
Ser-51	0.4	0.6	0.8			0.7	1.2	0.4						0.7								
Asn-52	0.5	0.7	0.7			0.6	0.6		0.8	0.8	0.8			0.7								
Leu-53	0.6	0.7	0.7			0.7	1.1		1.0			1.0	1.0	0.9								
Ala-54	0.6	0.7	0.8			1.0	1.3							0.9								
Ser-55	0.9	1.2	2.2			0.8	1.5	2.4						1.6								
Gly-56	0.9	1.2				0.9	0.6							0.9								
RMSD—back RMSD—side RMSD—all																						
Total L2	0.8		1.1		0.9																	
Residue	N	CA	CB	OG	C	O	CG	CD1	CD2	CE1	CE2	CZ	OH	CD	CG1	CG2	OG1	RMSD				
Ser-91	2.7	3.1	4.0	2.1	3.3	3.4												3.1				
Ser-92	3.5	3.9	4.4	4.3	3.9	5.9												4.4				
Tyr-93	2.2	2.5	2.3		2.2	2.5	4.0	5.1	5.5	7.0	7.3	7.9	9.9					5.5				
Pro-94	2.4	2.1	2.5		2.1	4.2	2.4							3.2				2.8				
Pro-95	1.2	1.5	3.2		1.5	1.3	3.9							2.9				2.4				
Ile-96	1.7	1.5	1.6		1.1	0.7								3.8	3.1	2.1		2.2				
Thr-97	1.6	1.4	1.8		1.1	1.3										2.1	2.0	1.6				
RMSD—back RMSD—side RMSD—all																						
Total L3	2.6		4.5		3.6																	

^a RMSD—back, includes atoms N, HN, CA, C, and O only; RMSD—side, includes all side chain atoms; RMSD—all, includes all atoms.

formed using CONGEN (Brucoleri & Karplus, 1987). Solvent-accessible surface areas were calculated with CONGEN using a subroutine previously written by Lee and Richards (1971).

Acknowledgments

The authors thank Dr. Robert Fox for providing the preliminary coordinates of their X-ray structure before they were available through the Brookhaven Protein Data Bank. We also thank Dr. Jiri Novotny at Bristol-Myers Squibb Pharmaceutical Re-

search Institute (BMSPRI) for his insightful discussions and review of the manuscript. S.S. thanks the Department of Chemistry, Princeton University, and the BMSPRI for their hospitality and Professor Harden McConnell for urging him to model ANO2. S.S. also acknowledges grant support from the Office of Naval Research and a Biomedical Research Support Grant.

References

- Anchin, J.M., Subramaniam, S., & Linthicum, D.S. (1992). Binding of the neuroleptic drug haloperidol to a monoclonal antibody: Refine-

- ment of the binding site using canonical structures. *J. Mol. Recog.* 4, 7–15.
- Anglister, J., Frey, T., & McConnell, H.M. (1984). Magnetic resonance of a monoclonal anti-spin-label antibody. *Biochemistry* 23, 1138–1142.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
- Bruccoleri, R.E., Haber, E., & Novotny, J. (1988). Structure of the antibody hypervariable loops reproduced by a conformational search algorithm. *Nature* 335, 564–568; also *Nature* 336, 266.
- Bruccoleri, R.E. & Karplus, M. (1985). Chain closure with bond angle variations. *Macromolecules* 18, 2767–2773.
- Bruccoleri, R.E. & Karplus, M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26, 127–196.
- Brünger, A.T., Leahy, D.J., Hynes, T.R., & Fox, R.O. (1991). 2.9 Å resolution anti-dinitrophenyl-spin-label monoclonal antibody Fab fragment with bound hapten. *J. Mol. Biol.* 221, 239–256.
- Chothia, C. & Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobins. *J. Mol. Biol.* 196, 901–917.
- Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, V., & Poljak, R. (1986). The predicted structure of immunoglobulin DL3 and its comparison with the crystal structure. *Science* 233, 755–758.
- Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M., & Poljak, R.J. (1989). Conformations of immunoglobulin hypervariable loops. *Nature* 342, 877–883.
- Davies, D.R., Padlan, E.A., & Sheriff, S. (1990). Antibody/antigen complexes. *Annu. Rev. Biochem.* 59, 439–473.
- de la Paz, P., Sutton, B.J., Datsley, M.J., & Rees, A.R. (1986). Modeling of the combined sites of anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.* 5, 415–425.
- Dower, S.K., Wain-Hobson, S., Gettins, P., Givol, D., Jackson, B.J., Perkins, S.J., Sunderland, C.A., Sutton, J.E., Wright, C.E., & Dwek, R. (1977). The combining site of the dinitrophenyl binding immunoglobulin A myeloma protein MOPC315. *Biochem. J.* 165, 207.
- Epp, O., Latham, E., Schiffer, M., Huber, R., & Palm, W. (1975). The molecular structure of a dimer composed of the variable portions of the Bence Jones protein Rei refined at 2.0 Å resolution. *Biochemistry* 14, 4943–4952.
- Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., & Levinthal, C. (1986). Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of McPC603 from many randomly generated loop conformations. *Proteins Struct. Funct. Genet.* 1, 342–362.
- Gö, N. & Scheraga, H. (1970). Ring closure and local conformational deformation of chain molecules. *Macromolecules* 3, 178–187.
- Kabat, E.A. & Wu, T.T. (1972). Construction of a three dimensional model of the polypeptide backbone of the variable region of kappa immunoglobulin light chains. *Proc. Natl. Acad. Sci USA* 69, 960–964.
- Kabat, E.A., Wu, T.T., & Bilofsky, H. (1977). Unusual distributions of amino acids in complementarity determining (hypervariable) segments of heavy and light chains of immunoglobins and their possible role in specificity of antibody combining sites. *J. Biol. Chem.* 252, 6609–6616.
- Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., & Gottesman, K.S. (1987). *Sequences of Proteins of Immunological Interest*, 4th Ed. National Institutes of Health, Bethesda, Maryland.
- Kussie, P.H., Anchin, J.M., Subramaniam, S., Glasel, J.A., & Linthicum, D.S. (1991). Analysis of the binding site architecture of monoclonal antibodies to morphine by using competitive ligand binding and molecular modeling. *J. Immunol.* 146, 4248–4257.
- Leahy, D.J., Rule, G.S., Whittaker, M.M., & McConnell, H.M. (1988). Sequences of 12 monoclonal anti-dinitrophenyl spin-label antibodies for NMR studies. *Proc. Natl. Acad. Sci USA* 85, 3661–3665.
- Lee, B.K. & Richards F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.
- Martin, A.C.R., Cheetham, V.C., & Rees, A.R. (1989). Modeling antibody variable loops. A combined algorithm. *Proc. Natl. Acad. Sci. USA* 86, 9268–9272.
- Nell, L.J., McCammon, J.A., & Subramaniam, S. (1992). Anti-insulin antibody structure and conformation. I. Molecular modeling and mechanics of an insulin antibody. *Biopolymers* 32, 11–21.
- Novotny, J., Bruccoleri, R.E., Newell, J., Murphy, D., Haber, E., & Karplus, M. (1983). Molecular anatomy of the antibody binding site. *J. Biol. Chem.* 258, 14433–14437.
- Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J., & Davies, D.R. (1989). Structure of an antibody-antigen complex: Crystal structure of HyHel-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. USA* 86, 5938–5942.
- Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34, 167–339.
- Satow, Y., Cohen, G.H., Padlan, E.A., & Davies, D.R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* 190, 593–604.
- Shenkin, P.S., Yarmush, D.L., Fine, R.M., Wang, H., & Levinthal, C. (1987). Predicting the antibody hypervariable loop conformation I. Ensembles of random conformations for ringlike structures. *Biopolymers* 26, 2053–2085.
- Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., & Davies, D.R. (1987). The three dimensional structure of an antibody-antigen complex. *Proc. Natl. Acad. Sci. USA* 84, 8075–8079.
- Snow, M.E., & Amzel, L.M. (1986). Calculation of three dimensional changes in protein structure due to amino-acid substitutions: The variable region of immunoglobulins. *Proteins* 1, 267–279.
- Stanford, J.M. & Wu, T.T. (1981). A predictive method for determining the possible three dimensional foldings of immunoglobulin backbones around antibody combining sites. *J. Theor. Biol.* 88, 421–439.
- Theriault, T.P., Leahy, D.J., Levitt, M., McConnell, H.M., & Rule, G.S. (1991). Structure and kinetic studies of the Fab fragment of a monoclonal anti-spin label antibody by nuclear magnetic resonance. *J. Mol. Biol.* 221, 257–270.
- Tramontano, A., Chothia, C., & Lesk, A.M. (1989). Structural determinants of the conformations of medium-sized loops in proteins. *Proteins Struct. Funct. Genet.* 6, 382–394.
- Wu, T.T. & Kabat, E.A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132, 211–250.