Computer-aided molecular modeling of the binding site architecture for eight monoclonal antibodies that bind a high potency guanidinium sweetener

Jerry M. Anchin,* Chhabinath Mandal,* Chris Culberson,† Shankar Subramaniam,‡ and D. Scott Linthicum*

**Department of Veterinary Pathobiology, Texas A&M University, College Station, TX*

†Merck Sharp and Dohme Research Laboratories, West Point, PA

‡Beckman Institute, University of Illinois, Urbana, IL

Computer-aided molecular modeling of the antibody binding site of eight different monoclonal antibodies (mAb) that bind the intense sweetener ligand (N-(p-cyanophenyl)-N'-diphenylmethyl) guanidine acetic acid was completed using canonical loop structures and framework regions from known immunoglobulins as "parent structures" for the molecular scaffoldings. The models of the fragment variable (Fv) region of the mAb were analyzed for the presence and location of residues predicted to be involved in ligand binding. Several binding site tryptophan residues in these models were located in positions that support previous flurospectroscopic observations of the mAb-ligand complexation. Computer-aided renderings of the electrostatic potential at the van der Waals surface of the Fv region were compared and found to be consistent with the ligand binding specificity profiles for the different mAb. The Fv model of mAb NC6.8 was consistent with the binding site features determined in the Fab structure recently solved by X-ray diffraction techniques. These Fv models should provide an adequate basis for site-directed mutagenesis experiments in order to characterize interactive motifs in the mAb binding site.

Keywords: monoclonal antibody, molecular modeling, sweeteners, ligands, canonical structures

Color Plates for this article are on pages 289–290.

Address reprint requests to Dr. Linthicum at the Department of Veterinary Pathobiology, Texas A&M University, College Station, TX 77843, USA.
Received 22 November 1993; accepted 28 March 1994

INTRODUCTION

Although there are more than 3,000 immunoglobulin (Ig) variable (V) region sequences known,[1] the only readily available experimental methods for solving the three-dimensional structure of antibody binding sites have been X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.[2] The structures for more than 40 different Fab structures have been solved by X-ray diffraction.[3]

In some instances, arduous attempts to crystallize and solve the structure of a specific antibody have failed for a variety of reasons. Over the past several years, many investigators have turned to techniques of computer-aided molecular modeling (CAMM) to generate three-dimensional structures of the antigen combining site.[4–6] There has been a steady increase in the sophistication of the computational approaches for the generation and modeling of the antibody binding site architecture.[7]

The Ig Fv region consists of two flattened β-barrel-like structures that are comprised of the light (L) and heavy (H) chain V-region domains. The antigen combining site of the Fv portion molecule is created by the ensemble of six hypervariable loops or *complementarity determining regions* (CDR): three from the light chain (L1, L2, L3) and three from the heavy chain (H1, H2, H3). The CDR are located at the "top" of the Fv region.[8]

In many proteins, loop segments that emerge from the protein core have been referred to as *random coil structures*, and are thought to possess no apparent structural regularity. However, Sibanda and Thornton[9] found that the main chains of some short loop conformations were highly constrained by their end-loop positions and orientations, and therefore appear to adopt a limited set of "preferred" loop conformations. With respect to antibodies, Padlan and Davies[10] observed that the amino acid sequence variation in antibody

CDR loops had little effect on the main chain loop conformations. Kabat et al.[11] suggested that the conservation of specific residues within the antibody loops were structurally important to the overall conformation of the CDR.

An important analysis by Chothia and Lesk[12] of the empirical structures of known Fab and L-chain dimers (myeloma proteins) revealed that there was a relationship between certain amino acids and the three-dimensional structure of the binding sites. They were able to identify several "key" residues that, through their packing, hydrogen bonding patterns, or the ability to assume unusual ϕ, ψ, or ω angles, were primarily responsible for the main-chain conformations of the CDR. The term *canonical structures* was coined to describe the commonly occurring main-chain conformations found in the CDR. Despite the diversity of CDR sequences, the main-chain canonical features appear to be conserved in these loop regions, at least in those structures solved by X-ray cystallography to date.

Using a canonical structure scaffolding approach, CAMM studies[4,13] have generated accurate Fv models of the main-chain conformations for five of the six CDRs (L1, L2, L3, H1, H2). Accurate models of CDR H3 are difficult to create because this loop exhibits a great degree of structural variability in known structures. To address this issue, the conformational search algorithm CONGEN[14] has been used to model the structure of CDR H3. This algorithm permits a torsion-space sampling of conformations of short polypeptide segments. Bruccoleri et al.[5] used CONGEN to construct all the CDR loops of the anti-phosphocholine antibody McPC603 and the anti-lysozyme antibody HyHEL-5. In addition to conformational sampling, CONGEN permits the construction of longer chain closed loops from shorter chain loops.[15] Comparison of the model to crystal structure of McPC603 showed a rmsd of 2.4 Å for all loop atoms, and a 1.7-Å rmsd for the alpha carbon backbone atoms. Similar results were obtained for the model of antibody HyHEL-5.

In the present study, we modeled the binding sites (Fv region) of eight different monoclonal antibodies (mAb) that were produced against a high potency guanidinium sweetener. The N,N′,N″-trisubstituted guanidine ligand is one of the sweetest compounds known to man, and is 200,000 times sweeter than sucrose as assessed by a threshold taste test.[16] We have previously reported the V-region sequences and ligand affinities for the mAb in this study.[17] We have also reported spectroscopic observations of the sweetener-antibody complexation, and have speculated as to the identity of key-binding site tryptophans that are involved in ligand recognition.[18] In addition, the sweetener ligand-Fab crystallized complex for one of these mAb (clone NC6.8) has recently been solved at a resolution of 2.2 Å.[19] The Fv molecular models presented here support previous experimental observations and provide insight into the molecular determinants that participate in sweetener-antibody complexation. These studies may also shed light on putative receptor recognition motifs for high potency sweeteners.

EXPERIMENTAL

MAbs to the ligand (N-(*p*-cyanophenyl)-N′-diphenylmethyl) guanidine acetic acid (Figure 1) were derived from hybridomas produced in our laboratory. The nucleotide and deduced amino acid sequences of the eight mAb clones designated NC2.3, NC6.7, NC6.8, NC10.1, NC10.10, NC10.11, NC10.12, and NC10.14 are reported elsewhere.[17] The CDR amino acid residues for L and H-chain are presented in Tables 1 and 2.

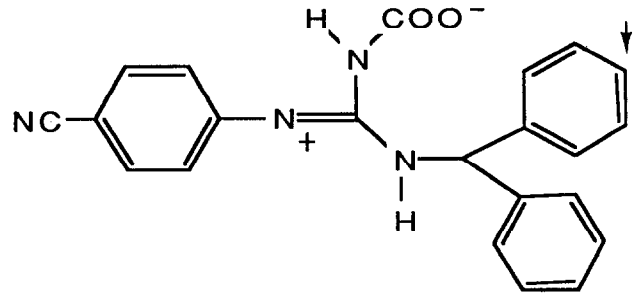

*Figure 1. The guanidinium sweetener, N-(*p*-cyanophenyl)-(N′-diphenylmethyl)guanidineacetic acid, is more than 200,000 times sweeter than sucrose. The pKa for the aryl nitrogen is 9.2 and therefore carries a positive charge at neutral pH, making this ligand a zwitterion. The arrow indicates the area to which the ligand was covalently linked to a protein carrier for the purposes of producing the monoclonal antibodies examined in this study.*

A number of different computer algorithms were used in this investigation to model and graphically display the FV regions of the mAbs. The programs ABalign and ABbuild (from the ABGEN algorithm, for *antibody generation*), written in our laboratory, were used to select "parent structures" and generate first-order models for each mAb. Details of the ABGEN algorithm (and programs ABalign and ABbuild) are reported elsewhere.[20] CHARMM[21] was used for molecular mechanics calculations. CONGEN[14] was used to generate low-energy conformations for CDR loops, especially for loops that lacked an exact match (in terms of the number of residues) with an empirical parent structure (as identified by the ABalign program). The algorithms NCView[22] and ACCESS[23] were used to generate electrostatic surface renderings and calculate solvent accessible surface areas for residues of the mAb models, respectively.

ABalign

The program ABalign aligns the amino acid sequences of the mAb to be modeled to a database of Ig FV sequences, derived from the Ig V-region sequences compiled by Kabat et al.[1] This Fv database contains the beginning and end sequence number for each L and H chain V-region, and the frequency of occurrence of all known residues in each position. The numbering of all residue positions in this study is according to the scheme of Kabat et al.[1] The amino acid sequences of the ten parent Fab structures are also used to generate a similar Fv database, complete with the R factors and atomic resolution. The ten known parent Fab structures used in this study were obtained from the Brookhaven Protein Database (BPDB).[24] Table 3 lists the name, BPDB code, species source, antigen specificity, Ig class, and atomic resolution (Å) for each structure used. The FR overlaps of alpha carbons for eight of the Fv structures (6FAB and 3FAB not included) are shown in Color Plate 1. The high degree of FR

Table 1. Light chain CDR amino acid sequences

	CDR L:1																
	24	25	26	27	a	b	c	d	e	f	28	29	30	31	32	33	34
NC2.3	T	A	S	S	S	*	*	*	*	*	V	S	S	S	Y	K	H
NC6.7	K	A	S	Q	*	*	*	*	*	*	N	V	G	I	K	V	A
NC6.8	R	P	S	Q	S	L	V	H	S	*	N	G	N	T	Y	L	H
NC10.1	T	A	S	S	S	*	*	*	*	*	V	N	S	N	Y	K	H
NC10.10	R	S	S	Q	S	L	V	R	S	*	N	G	N	A	Y	L	H
NC10.11	R	S	S	Q	S	L	V	H	S	*	N	G	N	T	Y	L	H
NC10.12	S	S	T	G	A	V	*	*	*	*	T	T	S	N	Y	A	I
NC10.14	S	S	T	G	A	V	*	*	*	*	T	T	S	N	Y	A	H

	CDR L:2							CDR L:3								
	50	51	52	53	54	55	56	89	90	91	92	93	94	95	96	97
NC2.3	S	T	S	T	L	A	S	H	Q	Y	H	R	S	P	P	T
NC6.7	S	A	S	Y	R	Y	S	Q	Q	Y	N	S	Y	P	Y	T
NC6.8	R	V	S	N	L	A	S	S	Q	G	T	H	V	P	Y	T
NC10.1	R	T	S	N	R	F	A	Q	Q	G	N	F	I	P	F	T
NC10.10	K	V	S	N	R	F	S	S	Q	S	T	H	V	P	W	T
NC10.11	R	V	S	N	R	F	S	S	Q	G	T	H	I	P	Y	T
NC10.12	G	T	N	N	R	V	P	A	L	W	Y	S	N	H	S	V
NC10.14	G	T	N	N	R	V	P	A	L	W	Y	S	N	H	W	V

Table 2. Heavy chain CDR amino acid sequences

mAb clone	CDR H:1							CDR H:2																	
	31	32	33	34	35	a	b	50	51	52	a	53	54	55	56	57	58	59	60	61	62	63	64	65
NC2.3	N	Y	W	V	H	*	*	Y	I	I	P	R	T	D	Y	T	E	Y	S	K	N	F	K	D
NC6.7	D	Y	E	I	H	*	*	V	I	H	P	G	S	G	G	I	V	Y	N	Q	K	F	K	V
NC6.8	F	Y	W	I	E	*	*	E	I	L	P	G	S	G	R	T	N	Y	R	E	K	F	K	G
NC10.1	S	G	Y	Y	W	N	*	Y	I	R	*	Y	D	G	D	S	N	F	N	P	S	L	K	N
NC10.10	S	Y	Y	M	S	*	*	A	I	N	S	N	G	G	T	T	Y	Y	P	D	T	M	K	G
NC10.11	Y	Y	W	I	E	*	*	E	I	L	P	G	T	I	R	S	N	Y	N	E	N	F	K	D
NC10.12	R	Y	W	M	S	*	*	E	I	N	P	D	S	S	T	I	N	Y	T	P	S	L	K	D
NC10.14	T	S	G	M	G	V	G	D	I	W	*	W	N	D	K	K	Y	Y	N	P	S	L	K	S

	CDR H:3													
	95	96	97	98	99	100	a	b	c	d	j	k	101	102
NC2.3	W	G	G	N	H	E	G	*	*	*	*	L	A	Y
NC6.7	Q	G	*	*	*	*	*	*	*	*	*	*	D	Y
NC6.8	G	Y	S	S	*	*	*	*	*	*	*	M	D	Y
NC10.1	V	L	Y	G	*	*	*	*	*	*	*	S	D	Y
NC10.10	E	V	Y	D	Y	Y	T	*	*	*	*	L	D	Y
NC10.11	A	Y	S	N	*	*	*	*	*	*	*	F	D	Y
NC10.12	L	I	T	T	A	T	W	Y	*	*	Y	F	D	Y
NC10.14	R	T	F	S	Y	Y	Y	G	S	S	F	Y	Y	F

Table 3. Parent Ig structures selected for modeling scaffolding

Name	BPDB Code	Source	Antigen	Class	Resolution (Å)
D1.3	1D13	murine	lysozyme	G_1, κ	2.5
36-71	6FAB	"	phenylarsonate	G_1, κ	1.9
HyHel-10	3HFM	"	lysozyme	G_1, κ	3.0
4-4-20	4FAB	"	fluorescein	G_{2a}, κ	2.7
R19.9	2F19	"	benzenearsonate	G_{2b}, κ	2.8
HyHel-5	2HFL	"	lysozyme	G_1, κ	2.5
McPC603	2MCP	"	phosphorylcholine	A, κ	3.1
J539	2FBJ	"	(1-6) D-galactan	A, κ	2.7
Newm	3FAB	human	vitamin K_1 OH	G_1, λ	2.0
KOL	2FB4	"	—	G_1, λ	1.9

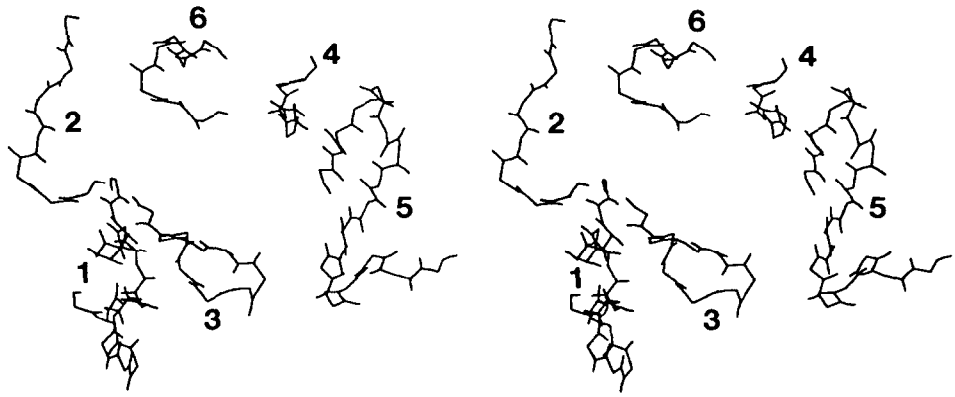

Figure 2. En face *view of the mAb binding site showing the six CDR loop segments of the Fv. Three are from the L-chain (1 = L1, 2 = L2, and 3 = L3), and three are from the H-chain (4 = H1, 5 = H2, and 6 = H3). This same orientation and view is used for the electrostatic surface renderings shown in Color Plates 2 and 3.*

overlap and canonical CDR structures among the known Fv regions is the basis for the homology scaffolding modeling approach.

ABalign selects the best "match" known structure to be used as the parent scaffolding (CDR and FR) upon which the mAb model is constructed. The program can select independently CDR and FR segments from different known structures, depending upon the homology match (based on loop size and sequence). ABalign separates the FR and CDR of the model mAb (based on the templates previously defined), and identifies conserved amino acids, in specific locations, in each of these regions (e.g., L:10). The framework comparisons are executed in the order L:1 to L:23, L:35 to L:57, L:58 to L:88, L:98 to L:106. To ensure that the residues at the end of the L chains are present and properly reported, residues L:103 to L:108 are matched. The H chain is subsequently examined, in the order H:1 to H:35, H:36 to H:46, H:59 to H:93 and H:103 to H:112. The output from this alignment and matching comparison is called the *matching index,* and represents the total number of residues that match between the FR of a known BPDP parent structure and the mAb model. ABalign selects the best scoring structure to use for the FR for the L and H-chains.

ABbuild

The program ABbuild is designed to model the matched CDR and FR structures identified by ABalign. Briefly, to generate the first-order mAb model, the program performs the following functions:

(1) Superimposition of the selected BPDB structures.
(2) CDR loop "grafting" from donor structures and "annealing" of the loop ends to the parent FR scaffolding.
(3) Insertions and deletions of specific residues.
(4) Execution of specific amino acid replacements.
(5) Rotation of side chains for replacements to remove any improper contacts with nearby atoms.
(6) Generation of a Brookhaven format coordinate (.pdb) file, which is used in conjunction with the QUANTA software package for further refinement of the model.

CHARMM

The algorithm CHARMM is contained in the QUANTA modeling suite developed by Molecular Simulations Inc. (Waltham MA). The force fields used in the molecular mechanics have been previously reported.[21] In our work, we

used CHARMM to carry out constrained energy minimizations on loop residues. A dihedral constraint value of 200 kcal/mol/radian was used during the energy minimization of the mAb models; CHARMM constraints were applied to the invariant cysteine residues at L:23, L:88, H:22, and H:92. These cysteine residues form the disulfide bonds that are contained in all the mAb structures; CHARMM constraints were applied to keep these disulfide bonds intact. The steepest decent minimization algorithm was first used, followed by conjugate gradient and Adopted-Basis Newton–Raphson.

CONGEN

The method of grid sampling as implemented in the program CONGEN is described by Bruccoleri et al.[25] We executed the search procedure as follows:

(1) Generation of stereochemically acceptable conformations for all loop structures that avoided significant steric overlaps with the rest of the structure.
(2) Ranking of acceptable conformations generated according to their potential energies.
(3) Selection of the lowest energy conformation.

Because the time required for conformational sampling grows exponentially with the number of residues, CONGEN was used on only five contiguous amino acids of the loop at one time. The CONGEN algorithm was employed in all cases of residue insertion or deletion, and in situations where entire loops were grafted to the parent FR scaffolding.

Surface Renderings (NCView) and Calculations (ACCESS)

Electrostatic surface renderings of the models were generated using the NCView algorithm. NCView uses the van der Waals (VDW) surface derived from the radii values compiled by Bondi;[26] the surface contours were produced using the algorithm of Purvis and Culberson.[22] The electrostatic potential (EP) was generated for each surface point using the point charge approximation. The Coulombic charges for each amino acid were obtained from a list of united-atom atomic charges for standard amino acids. The color renderings represent a gradient of the calculated point charges, with blue, green, and cyan indicating the gradient of cationic values, and yellow, orange, and red indicating anionic regions.

Solvent accessible surface areas (SASA) were calculated according to the program ACCESS by Lee and Richards,[23] which uses a spherical probe of water (r = 1.4 Å), and calculates the contact area accessible (Å^2) on the VDW surface of each atom that can be contacted by the probe.

RESULTS

We describe here the modeling for mAb NC6.7 as an example of the ABalign and ABbuild procedures for the initial model. Based on the ABalign selection criterion, antibody HyHEL-5 was selected for use as the scaffolding for the H-chain FR, as well as CDR H1 and H2. ABalign searched for an exact size match for CDR H3, because HyHEL-5 required three deletions to match the loop length for mAb NC6.7. Examination of the ten BPDB structures showed that all structures required deletions in order to match mAb NC6.7. HyHEL-10 required the fewest deletions (one residue), and was therefore selected for the CDR H3 scaffolding.

Table 4 lists the ABalign matches for the eight mAb models. The number of residue replacements required for the Fv models are listed, as well as any necessary grafting, insertions or deletions in the CDR segments. For example, the L-chain model of mAb NC10.10 required the fewest number of replacements [n = 9], while the L-chain model of mAb NC10.14 required the most [n = 66]. Most of the mAb models required few CDR grafts, insertions, or deletions. NC10.12 and NC10.14 were exceptional in that they required major changes, such as grafts, insertions, and deletions (see Table 4); this is probably due to the fact that they are lambda-containing mAbs and the database of known

Table 4 Summary of additions, deletions, replacements, and parent structures used for modeling

	Light Chain					Heavy Chain				
MAb	Replacements[a]	Frame	CDR L:1	CDR L:2	CDR L:3	Replacements[a]	Frame	CDR H:1	CDR H:2	CDR H:3
NC2.3	42	R19.9	R19.9 (1a)[b]	R19.9	R19.9	38	R19.9	R19.9	R19.9	McPC603 (1d)[b]
NC6.7	44	HyHEl-10	HyHEL-10	HyHEL-10	HyHEL-10	34	HyHEL-5	HyHEL-5	HyHEL-5	HyHEL-10 (1d)[b]
NC6.8	12	4-4-20	4-4-20	4-4-20	4-4-20	22	HyHEL-5	HyHEL-5	HyHEL-5	HyHEL-5
NC10.1	41	R19.9	R19.9 (1a)[b]	R19.9	R19.9	48	D1.3	D1.3 (1a)[b]	D1.3	HyHEL-5
NC10.10	9	4-4-20	4-4-20	4-4-20	4-4-20	42	J539	J539	J539	J539 (1a)[b]
NC10.11	12	4-4-20	4-4-20	4-4-20	4-4-20	25	HyHEL-5	HyHEL-5	HyHEL-5	HyHEL-5
NC10.12	65	KOL	NEW	KOL	McPC603	17	J539	J539	J539	3671
NC10.14	66	KOL	NEW	KOL	McPC603	54	D1.3	D1.3(2a)[b]	D1.3	KOL (1d)[b]

[a]Replacements are the number of residues in the parent structures that were changed to those in the modeled structures.
[b]The number of deletions (d) and additions (a) needed in the CDR are noted.

structures for lambda is small (only three structures have been solved).

The complete first-order Fv model of NC6.7 was assembled from appropriate L and H-chain scaffolding using the ABbuild program. The Cartesian coordinates for mAb HyHEL-5 and HyHEL-10 were used by ABbuild, and the molecules were superimposed by matching the FR residues. The rmsd for the main-chain atoms of the two structures was 0.65 Å. Superimposition was performed on CDR H3 from HyHEL-10, and on CDR H3 from HyHEL-5 by comparing atoms H:92 to H:93 and H:103 to H:104 (rmsd = 0.54 Å). ABGEN placed CDR H:3 (residues H:94 to H:102) from HyHEL-10 onto the new composite model for mAb NC6.7, and annealed the ends of the loop onto the framework of HyHEL-5. Amino acid H:98 was cut from CDR H3 in HyHEL-10 to make it the same length as CDR H3 for mAb NC6.7, and the ends were annealed. Next, ABbuild replaced the amino acids of HyHEL-5 and HyHEL-10 with those of mAb NC6.7 (44 replacements in the L-chain and 34 replacements in the H-chain—mostly in the CDR). Rotation of each replaced residue side chain was carried out to relieve any improper steric contacts with neighboring atoms (76 residues rotated). A file was generated from the output of ABbuild that contained the new Cartesian coordinates for the first-order Fv model of mAb NC6.7. The other mAb Fv models were completed in a similar manner.

CHARMM energy minimizations were carried out on all eight mAb models. The minimization protocol consisted of 100 steps each of steepest descents, conjugate gradient, and Adopted-Basis Newton-Raphson minimizations. Dihedral constraints and CHARMM patches were applied during all energy minimization procedures. Table 5 lists the potential energy values (*in vacuo*) for the models. Values are listed for ABbuild generated models (prior to molecular mechanics), CHARMM energy minimized models and models that were further refined using CONGEN. Some of the primary models generated by ABGEN had extremely large potential energies (NC2.3 = 4.05×10^9 kcal/mol), while some had fairly low potential energies (NC6.7 = 1709 kcal/mol); these values were most likely due to improper steric contacts prior to molecular mechanics calculations. Table 5 also shows that the initial potential energies were greatly reduced by using the energy minimization algorithms. All models that required loop grafting, insertions, or deletions had lower potential energy values after CONGEN was used to regenerate the loops. The final potential energies of the models ranged from −7198 to −9120 kcal/mol, which represents values that are similar to those obtained from the antibody structures contained in the BPDB.

Figure 3 shows that alpha carbon backbone overlaps (stereo view) of the six CDR from the eight Fv models. Most of the new loop structures could be easily superimposed onto the parent scaffolding, and a high degree of overlap is apparent. Modeling of the H3 loop, however, required "grafting" in four of the eight models. Table 6 lists the rmsd of superimposition for the alpha carbons of the parent BPDB structures (Table 4) and the completed models that required only residue replacements (no insertions, deletions or grafting). The rmsd of superimposition was derived for the FR of L and H chain (L:1 to L:25, L:34 to L:47, L:58 to L:87, and L:99 to L:106; H:1 to H:25, H:36 to H:49, H:66 to H:93, and H:103 to H:113) and all α-carbon atoms of the

Table 5. Potential energy for the eight mAb models

MAb	Potential Energy (kcal/mol)		
	ABGEN[a]	CHARMM	CONGEN
NC2.3	4.05×10^9	−5614	−8212
NC6.7	1709	−8941	−9120
NC6.8	3405	ND	−8986
NC10.1	2.26×10^7	−5352	−7964
NC10.10	1.45×10^7	ND	−8864
NC10.11	7979	ND	−8770
NC10.12	2.13×10^8	−8633	−9383
NC10.14	3.56×10^9	−5670	−7198

[a]Initial ABGEN model structure prior to molecular mechanics using CHARMM and CONGEN; values are calculated *in vaccuo*

models (L and H-chains combined). Molecular mechanics refinement did not significantly distort the overall scaffolding backbone from the positions of the parent structures (NC6.7 L chain rmsd 0.22 Å for all C-α, and 0.20 Å for FR C-α).

Color Plates 2 and 3 show the NCView *en face* renderings of the eight mAb models. Figure 3 shows the *en face* orientation of the six CDR for comparison with the NCView surface renderings. These graphical representations provide information concerning the EP at the VDW surface of the Fv models. The surfaces are color coded, with blue representing the basic or positively charged regions, and red representing the acid or negatively charged regions. Other colors represent the EP gradient between these two extremes.

Inspection of the NCView surface for the FV model of mAb NC6.8 reveals several charged residues near the binding pocket surface. A deep pocket is visible at the center of the model (Color Plate 2). At the edge of the central pocket is an area of positive charge. These data suggest that residues H:35 GLU and H:50 GLU (with approximately 10% of their total surface area exposed) and ARG at position H:56 (with 40% of its total surface area exposed) are key residues in forming a salt bridge with the acetic acid moiety of the ligand, while the GLU at positions H:35 and H:50 participate in H-bond formation with the proton on the aryl nitrogen (pK = 9.2) of the guanidine.

A similar analysis of mAb NC10.11 was performed using the NCView generated model. Visual inspection of the Fv model (Color Plate 3) reveals charged surface exposed residues similar in nature to mAb NC6.8. Examination of the NCView surface shows a negatively charged binding pocket with a positively charged area nearby. Figure 4 shows the percent SASA for charged amino acids (LYS, ARG, GLU, ASP, and HIS) in the CDR of mAb NC10.11. The charged residues near the binding pocket are H:35E, H:50E, and H:56R. ARG H:56 is highly solvent exposed in this model. It is predicted that these amino acids contact the ligand in the same manner as mAb NC6.8 (i.e., a salt bridge between the acetic acid moiety of the ligand and H:56R).

MAb NC10.12 and NC10.14 have uncharged, nonpolar binding residues in the binding site as compared to the other NC mAbs in this collection (Color Plate 3). There are few or no charged residues at or near the ligand binding pocket of

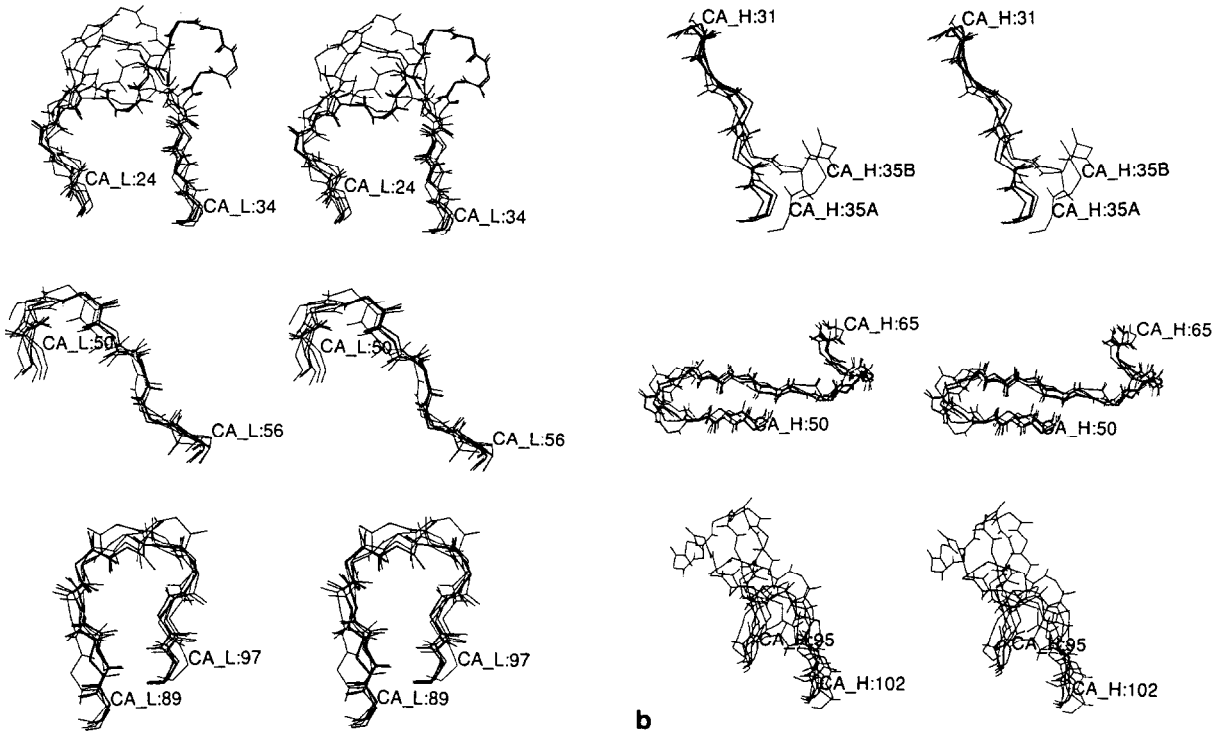

Figure 3. The alpha carbon atoms for the (a) three light chain CDR loops (L1, L2, L3) and (b) three heavy chain loops (H1, H2, H3) are overlapped in stereo for each of the eight Fv models. The alpha carbon atoms at the beginning and end of each loop are labelled. Two Fv models contain lambda light chains; the other four are kappa chains (note the major differences for the L1 conformations). The size and conformation of CDR H3 for each mAb are quite varied.

Table 6. Comparison of alpha carbons between the model Fv and the parent structure

Light Chain				Heavy Chain			
		rmsd (Å)[a]				rmsd (Å)[a]	
Number of residue replacements	mAb	All[b]	FR[b]	Number of residue replacements	mAb	All[b]	FR[b]
42	NC6.7	0.223	0.206	22	NC6.8	0.188	0.189
12	NC6.8	0.202	0.207	25	NC10.11	0.145	0.138
11	NC10.8	0.193	0.188				
9	NC10.10	0.180	0.169				
12	NC10.11	0.158	0.147				

[a]Expressed as root mean square deviation (rmsd) between parent and final Fv model structure; comparison is made for only those models that did not require insertions or deletions in the parent scaffolding (i.e., the same number of residues in CDR and FR throughout the model structure).
[b](All) denotes all α-carbon FR and CDR atoms. (FR) denotes the α-carbon atoms in framework regions only.

these mAbs; most are at the periphery of the FV face. SASA analysis for charged residues in the CDR of the model for NC10.14 (Figure 4) confirms the visual inspection. The majority of the charged residues are basic and are located in CDR H2. These are distal to the pocket, while others have poor solvent exposure. It is predicted that mAb NC10.14 has predominately hydrophobic interactions with the ligand, and that electrostatic interactions probably play a minor role in the complexation. This prediction, which is based on the model, is supported by the ligand binding specificity exhibited by this mAb, in which changes in the diphenyl ring substituent of the ligand produce a dramatic change in antibody binding (Anchin and Linthicum, unpublished observations). X-ray diffraction data for Fab crystals of NC10.14 complexed with ligand have been collected and the structure solution is currently underway (Guddat, et al., unpublished observations).

We examined several of these Fv models for the presence and position of V-region tryptophan residues. Previous spectroscopic analyses of the mAb-ligand complexation reactions indicated that tryptophan residues participated in ligand recognition as evidenced by intrinsic tryptophan fluorescence quenching.[18] In the mAb NC10.1 there are several V-region tryptophans (H:35, H:36, H:47, H:103,

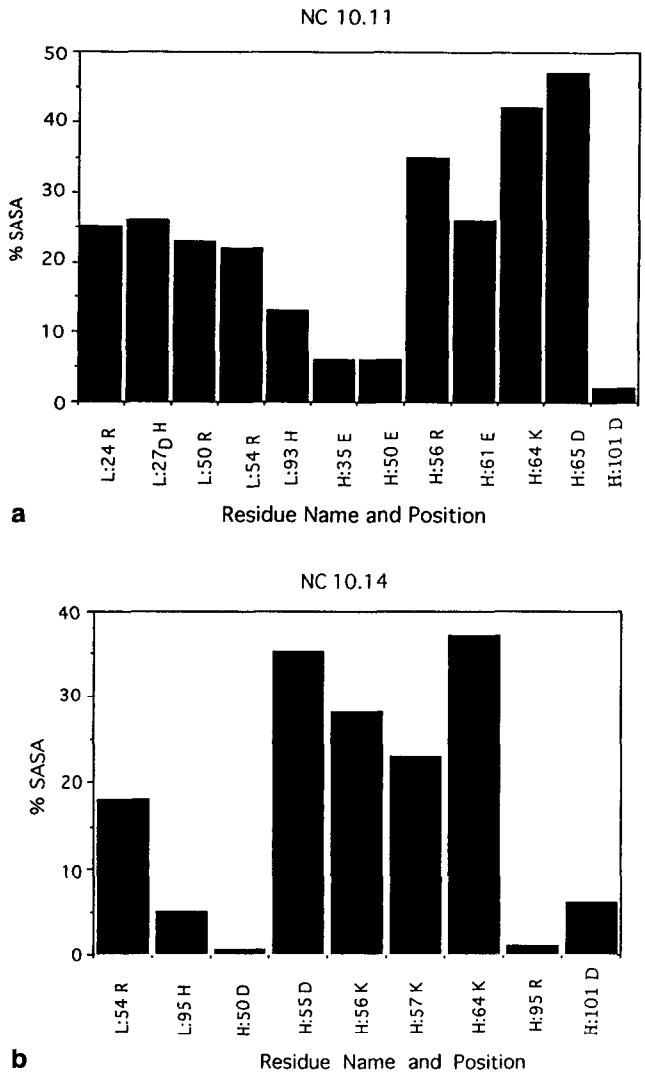

Figure 4. The solvent accessibility surface area (SASA) of charged residues in the L-chain and H-chain CDR are calculated for the models using the algorithm ACCESS. The values shown for mAb models (a) NC10.11 and (b) NC10.14 are expressed as the percentage of the total van der Waals surface area for each CDR residue that can be contacted by the 1.4-Å spherical surface probe. These data can be compared to the NCView surface renderings shown in Color Plates 2 and 3. Based on the model here and the X-ray diffraction data of NC6.8,[19] we predict that two key residues in the binding site for NC10.11 are H:56R, which forms a salt bridge with the acetic acid moiety of the ligand, and H:50E, which forms a hydrogen bond with the aryl nitrogen (pK = 9.2) of the ligand. The SASA for exposed, charged residues of the NC10.14 model shows that most are in CDR H2; many of these are distal to the ligand-binding pocket.

and L:35), but three of these (H:36, H:103, and L:35) are structural residues and are found in 98–99% of known antibody molecules. Visual examination of the Fv model for NC10.1 reveals that H:35 is a likely candidate for ligand recognition. It is located in CDR H1 and is partially buried in the putative ligand binding site (Color Plate 4). The SASA for H:35 in the Fv model of NC10.10 was calculated to be 26 Å^2, which represents about 10% of the available SASA for tryptophan. In the Fv model for NC10.10, the key V-region tryptophan (L:96W) is located in CDR L3 near the putative ligand binding site (Color Plate 5) and had a solvent exposure of 37 Å^2. Based on these models, we predict that the ligand complexes with these tryptophans and these specific residues are responsible for the ligand-induced fluorescence quenching.

DISCUSSION

The molecular complexation between antibody and antigen is dictated by interactions between a precise set of amino acid residues located in the antibody variable region and complementary sites on antigen.[3,27,28] X-ray diffraction studies of Fab crystals derived from myeloma proteins and mAb have revealed that the "binding site" of an antibody is created by six CDR peptide loops located at the N-terminal variable region of the heavy and light chain polypeptides.[11] The CDR loops are shaped by the antiparallel folding of the β-sandwich structures of each polypeptide chain. Certain types of amino acids in the CDR segments, such as those capable of π–π interactions (aromatics) or ionic (so called "salt bridge") interactions (acidic and basic residues), provide highly favorable motifs for complexation with complementary sites of the antigen and the propensity for such amino acids to be located in antibody binding sites has been previously noted.[29,30,31]

Several excellent studies of comparative molecular anatomy of antibody binding sites and successful CAMM of Fab molecules have been reported in the past few years.[6,13,32] Chothia and Lesk[12] were the first to carefully analyze the X-ray derived atomic structures of Fab fragments, and identified the relatively few residues that, through their primary packing, hydrogen bonding, or ability to assume unusual ϕ, ψ, or ω conformations, were primarily responsible for the backbone conformations of the CDR segments. These commonly occurring backbone conformations of the CDR have been called *canonical structures.*

It was the working concept of *canonical structures* that permitted CAMM of the Fv structures for several mAb before the X-ray diffraction studies were completed.[4,6] Techniques for accurate modeling of Fv regions has undergone further refinement by Bruccoleri and coworkers[5] using CONGEN, a conformational search algorithm,[14] to calculate low-energy loop configurations for models of mAb McPC603 and HyHEL-5; predicted structures for both mAbs were found to agree with the known crystal structures. CONGEN has proved to be especially useful in the modeling of CDR H3, for which there are no canonical structures because this loop is created by VDJ splice assembly of the antibody genes, a process which tends to be highly varied in terms of the number of residues placed in this region of the polypeptide chain.

Accurate prediction of Fv structures prior to empirical determinations can provide useful models for site directed mutagenesis experiments involving the antigen binding site and "humanization" of specific FR residues. Holm et al.[33] examined the binding interactions that occur between a mAb (Ox1) and its antigen (2-phenyloxazolone) using CAMM. Mutated variants of this mAb were found to bind the antigen

with even greater affinity. They observed that the modelled antigen-binding cleft of the Ox1 mAb was sterically complementary to the ligand, and certain mutations were found to increase the complementary electrostatic interactions. Using canonical structures and a combined algorithm,[6] Jackson et al.[34] were able to model the Fv region of a bovine antitestosterone monoclonal antibody, and predicted some of the key features of antibody–antigen interaction prior to site-directed mutagenesis experiments. Examination of the Fv model revealed that CDR H3 and a single heavy chain FR residue (H:47W) associated to form a hydrophobic cavity large enough to accommodate a single molecule of testosterone. The model was consistent with a hydrogen bond formation between a tyrosine in CDR H:3 (H:97Y) and the hydroxyl group on the D-ring of testosterone. Their predictions, based on the theoretical model, were confirmed by site-directed mutagenesis in which removal of the H-bonding donor (Y97F) resulted in a small, but measurable, loss in affinity for the steroid ligand.

The Fv models presented herein have proved useful in the analyses of experimental data involving ligand binding and specificity for this collection of mAbs. In a previous study,[18] we examined the steady state quenching of intrinsic tryptophan fluorescence for five of the anti-guanidine mAbs modeled. The quenching of intrinsic fluorescence is caused by intermolecular complexation between the mAb and the sweetener ligand. For NC10.10, the emission intensity was quenched in the presence of ligand and the emission maxima was shifted to the blue region by 4 nm. The observed hypsochromic shift suggests that the ligand is interacting with a solvent exposed tryptophan. As the concentration of the mAb-ligand complex in solution increases, the overall emission of the complex is dominated by the remaining shielded tryptophans located in the interior of protein matrix, thereby causing a "blue shift" in the emission maximum. In sharp contrast, mAbs NC2.3, NC6.8, and NC10.1 did not show hypsochromic shifts upon complexation with the ligand; the absence of a spectral shift suggests the tryptophans present in the CDR of these mAbs are not highly exposed to solvent.

The location of the tryptophans in the Fv models presented here support the previous experimental data for ligand induced fluorescence quenching. All of the mAbs in this study have tryptophans at H:36, H:103, and L:35, but these specific tryptophan residues appear in 98–99% of known murine V-region antibody sequences, and are probably not responsible for the different spectral properties of each mAb. Examination of the Fv model for NC10.10 suggests that L:96W (Color Plate 4) in CDR L3 is involved in ligand complexation and is responsible for the hypsochromic shift, due to its solvent exposure. Residues at this position (L:96) are known to contact antigen in several other mAbs as determined by X-ray diffraction of the crystallized Fab-antigen complexes.[3]

In contrast to the mAb NC10.10, mAbs NC6.8 and NC10.1 show monotonic ligand-induced quenching, and the V-region tryptophans located at H:33, H:35, and H:47 (see Table 2) are, as in other known Fab structures, not highly exposed to solvent. Any of these three residues could be involved in the ligand complexation and tryptophan quenching. H:47 W is probably not involved in these reactions, because this residue tends to be a structurally conserved residue, as it is found in 92% of all known V-region sequences. Examination of the Fv models generated in the study here, suggest that H:33 and H:35 (Color Plate 5), for NC6.8 and NC10.1, respectively, are the most likely to be involved in the ligand complexation.

X-ray diffraction studies of mAb NC6.8 Fab crystals complexed with the guanidine sweetener ligand have recently been completed.[19] Inspection of the ligand-Fab NC6.8 complex (2.2 Å resolution) reveals VDW contacts between the ligand and H:33W. This finding supports the theoretical Fv model and prediction of the participation of H:33W in ligand complexation as presented here. We report elsewhere detailed structural comparisons between the Fv model of NC6.8 and the X-ray determined structure.[35] X-ray diffraction studies for the crystals for several other NC mAbs are currently underway. The Fv models presented in this report should provide an excellent foundation for further development of accurate techniques for CAMM of antibody binding sites and provide models for site-directed mutagenesis experiments. The ABGEN algorithm is available from the authors upon written request (academic or other license agreement).

ACKNOWLEDGMENTS

This work was supported by Research Grant R01-GM46535 from the National Institutes of Health. We thank V. Lakamsani and G. Beedubail for technical assistance.

REFERENCES

1 Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. Sequences of proteins of immunological interest (Fifth Edition). *US Dept. HHS* 1991, NIH Publication 91-3242
2 Wright, P.E., Dyson, H.J., Lerner, R.A., Riechmann, L., and Tsang, P. Antigen-antibody interactions: An NMR approach. *Biochem. Pharmacol.* 1990, **40,** 83–88
3 Davies, D.R., Padlan, E.A., and Sheriff, S. Antibody-antigen complexes. *Ann. Rev. Biochem.* 1990, **59,** 439–473
4 Chothia, V., Lesk, A., Levitt, M., Amit, A., Mariuzza, R., Phillips, S., and Poljak, R. The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science* 1986, **233,** 755–758
5 Bruccoleri, R., Haber, E., and Novotny, J. Structure of antibody hypervariable loops reproduced by a conformational search alogrithm. *Nature* 1988, **355,** 564–568
6 Martin, A., Cheetham, J., and Rees, A. Modeling antibody hypervariable loops. A combined algorithm. *Proc. Nat. Acad. Sci. USA* 1989, **86,** 9268–9277
7 Thornton, J.M. Modeling antibody combining sites. *Catalytic antibodies* 1991, Chichester (Ciba Found. Symp. 159) Wiley, pp. 55–71
8 Davies, D.R., and Metzger, H. Structural basis of antibody function. *Ann. Rev. Immunol.* 1983, **1,** 87–117
9 Sibanda, B.L., and Thornton, J.M., b-hairpin families in globular proteins. *Nature* 1985, **316,** 170–174
10 Padlan, E.A. and Davies, D.R. Variability of three-dimensional structure in immunoglobins. *Proc. Nat. Acad. Sci. USA* 1975, **72,** 819–823

11 Kabat, E.A., Wu, T.T., and Bilofsky, H. Unusual distribution of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J. Biol. Chem.* 1977, **252,** 6609–6616
12 Chothia, C. and Lesk, A.M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* 1987, **196,** 901–917
13 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., and Poljak, R.J. Conformations of immunoglobulin hypervariable regions. *Nature* 1989, **342,** 877–883
14 Bruccoleri, R.E. and Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 1987, **46,** 137–168
15 Go, N. and Sheraga, H.A. Ring closure and local conformational deformation of chain molecules. *Macromolecules* 1970, **3,** 178
16 Muller, G.W., Walters, D.E., and DuBois, G.E. N,N′-Disubstituted guanidine high-potency sweeteners. *J. Med. Chem* 1992, **35,** 740–743
17 Anchin, J.M. and Linthicum, D.S. Variable region sequence and characterization of monoclonal antibodies to a N,N′,N″-trisubstituted guanidine high potency sweetener. *Molec. Immunol.* 1993, **16,** 1463–1471
18 Droupadi, P.R., Anchin, J.M., Meyers, E.A., and Linthicum, D.S. Spectrofluorimetric study of the intermolecular complexation of monoclonal antibodies with the high potency sweetener N-(p-cyanophenyl)-N′-diphenylmethyl) guanidine acetic acid. *J. Molec. Recog.* 1993, **5,** 173–179
19 Guddat, L., Shan, L., Anchin, J., Linthicum, D.S., and Edmundson, A.B. Local and transmitted conformational changes on complexation of an anti-sweetener Fab. *J. Molec. Biol.* 1993, **236,** 247–274
20 Mandal, C., Anchin, J., Viswanathan, M., Subramanian, S., and Linthicum, D.S. ABGEN: An automated algorithm for generation of antibody molecular models. *Biotechniques* 1994, submitted
21 Brooks, B., Bruccoleri, R., Olafson, H., States, D., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy minization and dynamics calculations. *J. Comput. Chem.* 1983, **4,** 187–217
22 Purvis, G.D. and Culberson, J.C. On the graphical display of molecular electrostatic force-fields and gradients of the density. *J. Molec. Graphics* 1986, **4,** 88–92
23 Lee, B. and Richards, F. The interpretation of proteins structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400
24 Bernstein, F.C., Koetzle, T.F., Williams, E.J.B., Meyer, E.F. Jr., Kennard, O., Shimanouchi, T., and Tasumi, M. The Protein Data Bank: A computer-based archival file for molecular structures. *J. Molec. Biol.* 1977, **112,** 535–542
25 Novotny, J., Bruccoleri, R.E., and Haber, E. Computer analysis of mutations that affect antibody specificity. *Proteins* 1990, **7,** 93–98
26 Bondi, A. Van Der Waals volumes and radii. *J. Phys. Chem.* 1964, **68,** 441–451
27 Novotny, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E., and Karplus, M. Molecular anatomy of the antibody binding site. *J. Biol. Chem.* 1983, **258,** 14433–14437
28 Novotny, J., Bruccoleri, R.E., and Saul, F.A. On the attribution of binding energy in antigen–antibody complexes of McPC603, D1.3, and HyHEL-5. *Biochem.* 1989, **28,** 4735–4749
29 Padlan, E.A. On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins* 1990, **7,** 112–124
30 Mian, I., Bradwell, A., and Olson, A. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* 1991, **217,** 133–151
31 Tramontano, A. and Lesk, A.M. Common features of the conformations of antigen-binding loops in immunoglobulins and application to modeling loop conformations. *Proteins* 1992, **13,** 231–245
32 de la Paz, P., Sutton, B., Darsley, M., and Rees, A. Modeling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO* 1986, **5,** 415–425
33 Holm, L., Laaksonen, L., Kaartinen, M., Terri, T., and Knowles, J. Molecular modeling studies of antigen binding to oxazolone-specific antibodies: the Ox1 idiotypic IgG and its mature variant with increased affinity to 2-phenyloxazolone. *Prot. Engineer.* 1990, **5,** 403–409
34 Jackson, T., Morris, B.A., Martin, A.C.R., Lewis, D.F.V., and Sanders, P.G. Molecular modeling and site-directed mutagenesis on a bovine anti-testosterone monoclonal antibody. *Prot. Engineer.* 1992, **5,** 343–350
35 Viswanathan, M., Anchin, J.A., Droupadi, P.R., Mandal, C., Linthicum, D.S., and Subramaniam, S. Structural predictions of the binding site architecture for monoclonal antibody NC6.8 using ligand binding, spectroscopy and computer-aided molecular modeling. *Biophys. J.* 1994, submitted