A Novel Modeling Protocol for Protein Receptors Guided by Bound-Ligand Conformation†

Margaret A. Johnson,* Christer Höög, and B. Mario Pinto*

Departments of Chemistry and of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6

Received October 1, 2002; Revised Manuscript Received December 16, 2002

ABSTRACT: A novel protocol for protein homology modeling is described, in which uncertainty in protein structure is resolved by applying the criterion that the protein must bind to a complementary ligand. A diverse library of protein models is created and then screened by docking with ligands of known conformation. The more accurate protein models form higher-quality docked complexes, and the quality of the fit is used to select the best models. The effectiveness of this technique with both natural and unnatural ligands is demonstrated by modeling the Fv fragment of an antibody and comparing the results to known crystal structures.

Knowledge of protein structure is assuming a central role in research and industry today. Recent advances in genomics have led to the sequencing of thousands of new genes from a variety of different organisms (*1*). An understanding of the functions of these proteins requires an appreciation of their structures. This realization has inspired the field of structural genomics and the development of better and faster methods of protein structure determination (*2*−*4*), with advances in high-throughput NMR[1] spectroscopy and X-ray crystallography of proteins (*4*, *5*).

However, not all proteins are amenable to these experimental techniques, and experiment must be supplemented by theoretical protein structure prediction. Currently, the most accurate method for protein structure prediction is comparative modeling (homology modeling) (*6*−*11*). The accuracy of this technique is limited by the level of similarity between the target protein and the known protein(s) used to construct the model (*10*). Thus, models based on more distantly related proteins, with lower degrees of sequence similarity, are less accurate. In addition, small areas of divergence within two sequences that are similar overall, such as variable loops, are difficult to construct accurately.

We propose a novel protocol in which knowledge of the bound conformation of a ligand may be incorporated into a conventional comparative modeling protocol (*6*−*11*). The ligand conformation will provide additional constraints on the shape and charge distribution of the binding site on the protein, leading to a more accurate model. The bound conformation of a ligand may be determined by NMR spectroscopy using the transferred nuclear Overhauser effect (trNOE) (*12*−*16*). This conformation may then be used to screen a library of protein models by computational docking. The quality of the docked complexes formed may be used to rank the models, with the most accurate models being the ones best able to bind the ligand. The goal of greater accuracy is very important if protein models are to be used in the design of drugs, vaccines, and diagnostic agents.

In our protocol, protein models within the library will share some aspects of their structure, while other aspects will be varied according to the most uncertain or ill-defined areas of the particular protein. One of the greatest challenges in comparative modeling is the treatment of variable loops within a structurally conserved framework (*10*, *17*−*20*). These occur where there are insertions or deletions relative to the known structure, or where there are certain regions with very low levels of sequence similarity. There are many possible, reasonable conformations even for a short loop within a protein, and it is often difficult to distinguish between these possibilities using force field methods. These differences may be quite significant in constructing a model, as variable loops often contribute to binding sites and/or active sites on the protein, and are important in determining specificity and function.

Antibodies are ideal test cases for comparative modeling. Built on a conserved structural framework, the immunoglobulin fold, they nevertheless exhibit a high degree of variability in their surface loops. Six hypervariable loops, or complementarity-determining regions (CDRs), define the binding site; their variation in sequence and structure is responsible for the ability of antibodies (Abs) to recognize an immense variety of different molecules (*21*). Therefore, although a reasonable alignment with the framework of the immunoglobulin fold may often be obtained, the binding site itself presents a considerable challenge to the computational

† This work was supported by the Natural Sciences and Engineering Research Council of Canada (including a postgraduate scholarship to M.A.J.) and the Wenner-Gren Foundation (postdoctoral fellowship to C.H.).

* To whom correspondence should be addressed. M.A.J.: telephone, (604) 291-5650; fax, (604) 291-3765; e-mail, johnsont@sfu.ca. B.M.P.: telephone, (604) 291-4884; fax, (604) 291-5424; e-mail, bpinto@sfu.ca.

[1] Abbreviations: Ab, antibody; mAb, monoclonal antibody; CDR, complementarity-determining region; CVFF, consistent-valence force field; Fab, fragment with antigen binding; Fv, variable fragment; LGA, Lamarckian genetic algorithm; NMR, nuclear magnetic resonance; rmsd, root-mean-square deviation; trNOE, transferred nuclear Overhauser effect; V_H, variable heavy chain; V_L, variable light chain.

10.1021/bi020608f CCC: $25.00
Published on Web 01/28/2003

chemist. Another complication is that the binding site is created by the association of two domains (V_H and V_L), the relative orientation of which varies between different antibodies (*21*). The method described here provides an effective protocol for dealing with these variables and obtaining more accurate models.

As a test case, we chose to model the anti-*Shigella flexneri* Y monoclonal antibody (mAb) SYA/J6 (*22*). The crystal structures of the Fab fragment, native and in complex with two carbohydrate ligands and one peptide ligand, have been determined to high resolution (*23*, *24*). The carbohydrate and peptide ligands use different subsets of specific hydrophobic and hydrogen bonding interactions within the binding site (N. K. Vyas et al., manuscript in preparation). These contacts and the overall shapes of the ligands are powerful tools for screening protein models by computational docking.

MATERIALS AND METHODS

Construction of Antibody Libraries. The framework of the antibody model was constructed using SYBYL 6.6/Composer software (*7*) (Tripos, Inc.). Homologous sequences were found by searching a customized database containing representative structures from the Protein Data Bank (*25*) and all known antibody structures (as of October 2000). Eight and four homologous sequences were used to construct the templates for the light and heavy chains, respectively. Homologous proteins used to construct the template were limited to those determined by X-ray crystallography with a resolution of $\leq$2.5 Å. Higher weights were assigned to more similar sequences and/or higher-resolution structures. For the heavy chain, a sequence with a lower overall level of sequence identity (45%) was included with three others of much greater identity (79−89%) to enforce the choice of a structure with the same canonical class in CDR H3 (*26*). All sequences of homologues chosen for the light chain template were $\geq$90% identical to the SYA/J6 light chain. The level of sequence identity was measured with SYBYL 6.6/Composer (*7*) using the Dayhoff 250 matrix (*27*). Structurally variable regions were defined as the CDR loops by the Kabat definition (*28*) as follows: L1, L24−L34; L2, L50−L56; L3, L89−L97; and H1, H31−H35. To enforce the correct sequence alignment, the variable regions encompassing H2 and H3 were defined as H52A−H55 and H92−H102, respectively. After construction of structurally conserved and structurally variable regions, brief energy minimizations [50 steps, using the conjugate gradient method and the CVFF force field (*29*) within Insight II 2000 (Accelrys, Inc.)] were performed, to gradually relieve strain: all side chains, all side chains and all atoms of variable loops, and, finally, all atoms.

After construction of each chain, the heavy and light chains were superimposed on a known structure to derive a combined model. The structure chosen for superimposition was 1QKZ (*30*), because of its high overall level of sequence identity (91 and 64% identical in the light and heavy chains, respectively), and the same canonical class as SYA/J6 in the CDR L3 and H3 loops. The models of each chain were superimposed on the structure of 1QKZ using framework (non-CDR) residues. The model was then further optimized by brief energy minimizations (50 steps, conjugate gradient method, all CDR atoms, and all side chains; followed by 50 steps, conjugate gradient method, all atoms). The H3 loop which had been added temporarily during construction of the model was then excised to leave the constant framework for the library.

Individual models differing in H3 loop conformation were generated by performing a database search within SYBYL 6.6/Biopolymer (Tripos, Inc.) for residues H92−H102, using the method of Claessens et al. (*31*). Candidate loop conformations were grafted onto the framework with small adjustments of the anchor residues of the loop (anchor melding) (*7*). Side chains of the loop were checked for steric interactions with the framework and were adjusted accordingly. Torsional angles of the CTR (H92−H94) and MDY (H100A−H102) side chains were copied from the donating antibody. The atoms of the loop in each model were adjusted by 100 steps of energy minimization (Powell method), while the rest of the framework was held fixed, followed by 700 steps of minimization (Powell method) for the whole protein. The conformations of some side chains of the light chain loops (L27D, L28, L30, L32, L34, and L91−L94) were adjusted to be similar to the light chain homologues and to the crystal structure. This served to maintain hydrogen bonding networks near the binding site and corrected for errors in side chain positioning introduced during model construction. The models were then energy-minimized for another 400 iterations. During optimizations, the models were assigned Kollman all-atom charges (*32*), and the Tripos force field (*33*) was used. Tripos force field energies were measured within SYBYL 6.6 and refer to the internal potential energy of the resulting protein model, measured after optimization. These energies include the standard components of the force field: bond stretching, angle bending, out-of-plane bending, torsional, van der Waals, and electrostatic components (*33*). The rmsd values with respect to the crystal structure were measured using ProFit version 2.2 (A. C. R. Martin, http://www.bioinf.org.uk/software/profit; *34*). Molecular graphics were performed with Insight II 2000 (Accelrys, Inc.) and SYBYL 6.6 (Tripos, Inc.). All models were checked for quality with PROCHECK (*35*).

Models differing in relative domain rotation were generated by separating the V_H and V_L chains of the original model framework and superimposing each on the two chains of a template antibody, using α-carbon atoms of non-CDR residues. The H3 loop was constructed, using the correct loop conformation from SYA/J6, and the models were optimized in a manner similar to that described for the first series of models. Optimization consisted of brief energy minimizations (50 steps, Powell method, Tripos force field; all atoms, H3 loop; all atoms of H3 loop and side chains of the rest of the protein; all atoms of the protein) followed by longer energy minimization (1700 steps) for the rest of the protein. The relative amount of rotation in each model was measured with respect to the crystal structure within Insight II as follows. The model was superimposed on the crystal structure using α-carbon atoms of the light chain; the additional rotation and translation required to superimpose the heavy chains were then measured (*36*, *37*).

Computational Docking of Antibody Models with Ligands. Computational docking simulations were performed with AutoDock 3.0 (*38*). This program uses rapid search methods in combination with a grid representation of the protein. The scoring function is an empirical binding free energy force

field that includes van der Waals, electrostatic, hydrogen bonding, torsional, and desolvation free energy terms (*38*). Its accuracy has been demonstrated in several studies of known receptors (*38*, *39*). Ligand conformations were extracted from the crystal structures of the corresponding Fab−ligand complexes and were assigned AMBER all atom potentials and charges (*32*) within Insight II 2000 (Accelrys, Inc.). The conformation of β-maltopentaose was extracted from the coordinates of *Bacillus subtilis* α-amylase complexed with maltopentaose (PDB entry 1BAG) (*40*). The protein models were assigned Kollman all atom potentials and charges (*32*) within Sybyl 6.6/Biopolymer (Tripos, Inc.). Grid maps representing the proteins were constructed using 60 × 60 × 60 points, with grid point spacing of 0.375 Å, and centered on the ligand, which was manually positioned within the binding site. Antibody−ligand complexes were generated from this starting point using a Lamarckian genetic algorithm (LGA) (*38*). Parameters for LGA docking were as recommended in the AutoDock manual. For each antibody model−ligand pair, 500 LGA docking runs were performed. The binding modes were clustered using a rmsd cutoff of 2.0 Å with respect to the starting position; the intermolecular energy of the lowest-energy cluster was used for all plots in this study. Intermolecular energy refers to the potential energy of the interaction *between* the protein and ligand [van der Waals, electrostatic, hydrogen bonding, and desolvation free energy components (*38*)]; it does not include the internal energies of the protein and ligand themselves, as these do not change during rigid docking. Correlation coefficients (*r*) were calculated using the VassarStats web site (http://faculty.vassar.edu/lowry/VassarStats.html; *41*).

RESULTS AND DISCUSSION

Construction of Libraries of Antibody Models. The first step in constructing a library of models is to build a framework which will be held constant in all the models while other regions are varied. In this case, the framework consisted of the structurally conserved β-strands of the immunoglobulin fold and, in addition, the first five hypervariable loops of the antibody, which can be modeled accurately based on known canonical loop structures (*42*−*45*). Diversity in this library could then be incorporated by varying the conformation of the third hypervariable loop of the heavy chain (CDR H3), a loop which varies greatly in sequence and structure, and which is often the most important in binding ligands (*26*, *46*). The CDR H3 loop of the antibody SYA/J6 is relatively long, and therefore presents a useful challenge.

The framework for the library, consisting of the structurally conserved regions and five of the six hypervariable loops (Figure 1a), was constructed using current comparative modeling procedures. Comparison with the crystal structure showed that this portion was modeled accurately (rmsd = 0.90 Å, all backbone atoms). A library of models was then generated by grafting various possible loop conformations onto this framework. A database loop search was performed to identify possible conformations for residues H92−H102 (CDR H3 and the three preceding residues), resulting in 11 different conformations which could reasonably fit into the framework (Table 1 and Figure 1b,c). The correct loop conformation from the crystal structure (*24*) was added to this group of possibilities. Each loop was added to the

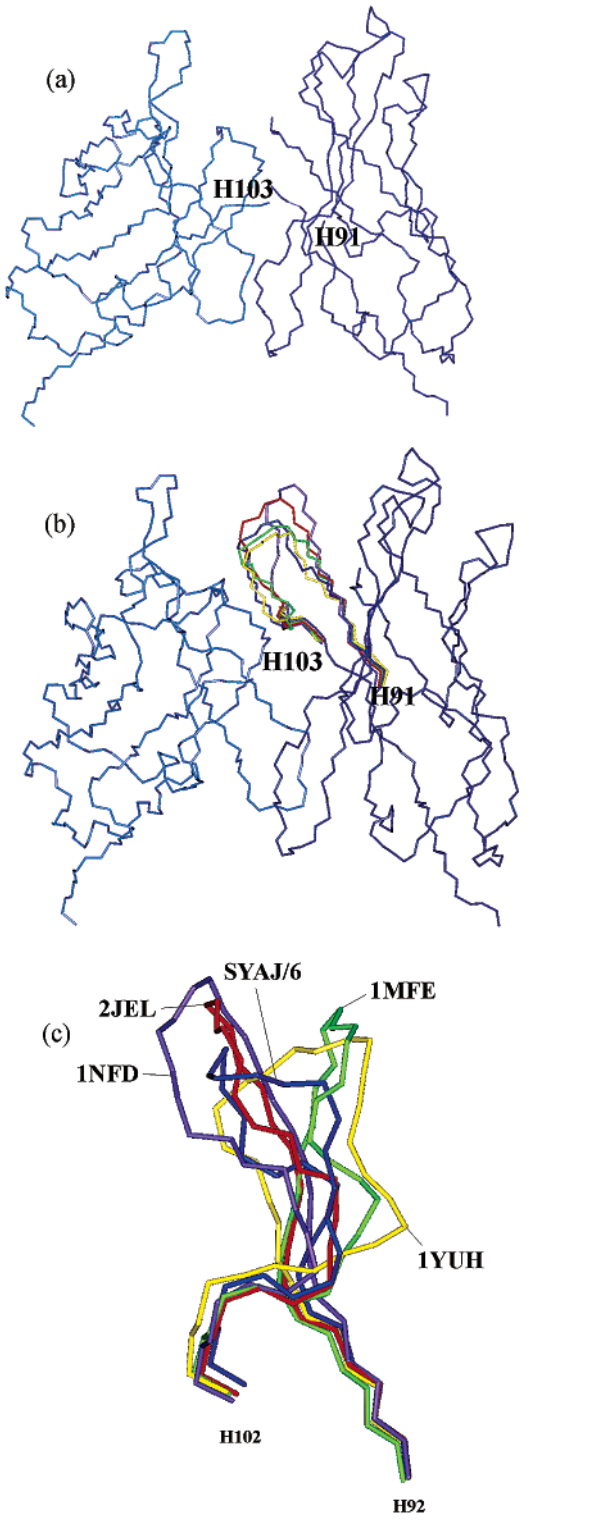


FIGURE 1: (a) Framework which was constant for all models in the first library, consisting of the structurally conserved regions and five of the six hypervariable loops (heavy chain, dark blue; light chain, light blue). The varied region, including the CDR H3 loop, consisted of residues H92−H102. (b) Framework and five representative loop conformations. The correct (crystallographically determined) conformation of the CDR H3 loop (dark blue) and four other loop conformations from models in the library are shown, in the positions where they were grafted into the framework. (c) Detailed view of the loop conformations in panel b. The loops are labeled with the PDB entry of the antibody from which the loop conformation was taken.

Table 1: Loops from the Protein Database Which Provided Different Possible Conformations for the CDR H3 Loop

Source protein	Loop sequence in source protein	Sequence homology[a] (%)
SYA/J6	CTRGGAVGAMDY	100
1JHL	CTRDDNYGAMDY	88
1IGT	CARHGGYYAMDY	81
1NFD	CTRAGRFDHFDY	81
1MFE	CTRGGHGYYGDY	77
1C5D	CTREDGWNYFDY	74
1YUH	CARYAYCRPMDY	74
1AR1	CVRHEYYYAMDY	74
1FSK	CTRGARDTWFAY	73
7FAB	CARNLIAGGIDV	71
2FBJ	CARLHYYGYNAY	66
2JEL	CARVMGEQYFDV	62

[a] Sequence homology, measured in Sybyl 6.6/Composer, using the Dayhoff 250 matrix (*27*), for the database loop vs the SYA/J6 loop.

Table 2: Accuracy of the Models with H3 Loop Conformational Variation

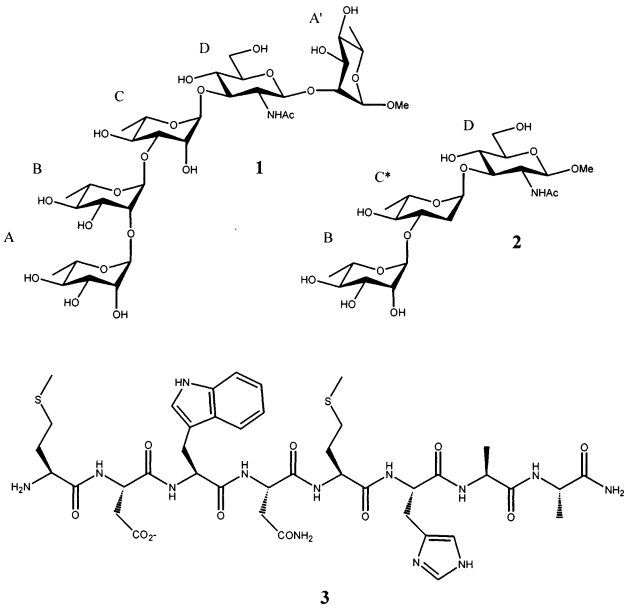
source protein	rmsd of loop[a]	rank by rmsd of loop	sequence homology[b] (%)	rank as in Composer (by sequence homology and fit)[c]
SYA/J6	0.69	1	100	1
1IGT	1.62	2	81	3
2JEL	1.72	3	62	12
1FSK	1.81	4	73	9
1C5D	2.16	5	74	6
1JHL	2.20	6	88	2
2FBJ	2.42	7	66	11
1AR1	2.43	8	74	8
7FAB	2.55	9	71	10
1YUH	2.63	10	74	7
1MFE	2.88	11	77	5
1NFD	3.01	12	81	4

[a] The rmsd of Cα atom positions of the CDR H3 loop, with respect to those in the X-ray crystal structure, after superimposition of each model on the crystal structure using all other Cα atom positions. [b] Sequence homology, measured in Sybyl 6.6/Composer, using the Dayhoff matrix (*27*), for the database loop vs the SYA/J6 loop. [c] The rms fit between the database loop anchor residues (the two residues preceding and one following the loop; N, Cα, and C atoms) and those in the framework is used to distinguish database loops which have the same level of sequence homology; i.e., the loop with the highest level of sequence homology is ranked the highest, and in the case of equal levels of sequence homology, the loop with the lower rms fit is chosen first.

framework to produce a model. The accuracy of each model was measured by calculating the root-mean-square deviation in position (rmsd) of the Cα atoms of the loop residues, with respect to their positions in the antibody crystal structure. The rmsd value provides an estimate of how greatly each model deviates from the true structure. For example, the correct model, in which the correct conformation of the H3 loop had been grafted into the constant framework of the library, has an rmsd of the loop Cα atoms of 0.69 Å, with respect to those of the crystal structure (Table 2); the corresponding value for the least accurate model is 3.01 Å.

Screening of Models by Docking with Ligands. A selection criterion must be applied to choose between different possible, reasonable loop conformations. We propose that

Scheme 1: Ligands Used for Screening the Library of Antibody Models



the ability of a protein model with a particular loop conformation to dock with a complementary ligand is a useful criterion. As a test, our library of antibody models was screened by computational docking with known ligands. The pentasaccharide L-Rha-α-(1→2)-L-Rha-α-(1→3)-L-Rha-α-(1→3)-D-GlcNAc-β-(1→3)-L-Rha-OMe (ABCDA′, **1**, Scheme 1) is a natural ligand in that it corresponds to a fragment of the *S. flexneri* Y O-antigen polysaccharide, the natural target of the antibody. The trisaccharide L-Rha-α-(1→3)-2-deoxy-L-Rha-α-(1→3)-D-GlcNAc-OMe (BC*D, **2**, Scheme 1) is an analogue that contains a site of deoxygenation in the central rhamnose ring, but binds with a higher affinity than the pentasaccharide **1** to SYA/J6 (*24*). The peptide ligand MDWNMHAA (**3**, Scheme 1) was discovered by phage display library screening with SYA/J6 and is not related to the natural antigen, but is specifically recognized by the antibody (*47*).

The ligands were docked with the antibody models in the library using AutoDock 3.0 (*38*), as described in Materials and Methods. The docking calculations were performed with the ligand constrained in the known conformation, while its orientation within the binding site was varied. Predicted binding modes for the pentasaccharide ligand, obtained from these experiments, are shown in Figure 2. The binding mode is predicted correctly for the best model (Figure 2a); the crystal structure of the complex is shown for comparison. The corresponding complex with the peptide ligand is also predicted correctly (Figure 3a). The correct pattern of hydrogen bonding and hydrophobic interactions is predicted in each case, although this pattern is quite different for each ligand. No water molecules were included with the models, as the prediction of their positions is computationally intensive (*48*). Hydrogen bonds mediated by water molecules are therefore lost in the modeled complexes. However, the overall positioning of the ligands is maintained by other driving forces incorporated into the force field: direct protein−ligand hydrogen bonds and hydrophobic interactions. These interactions are reflected in the calculated intermolecular potential energy (energy of interaction be-

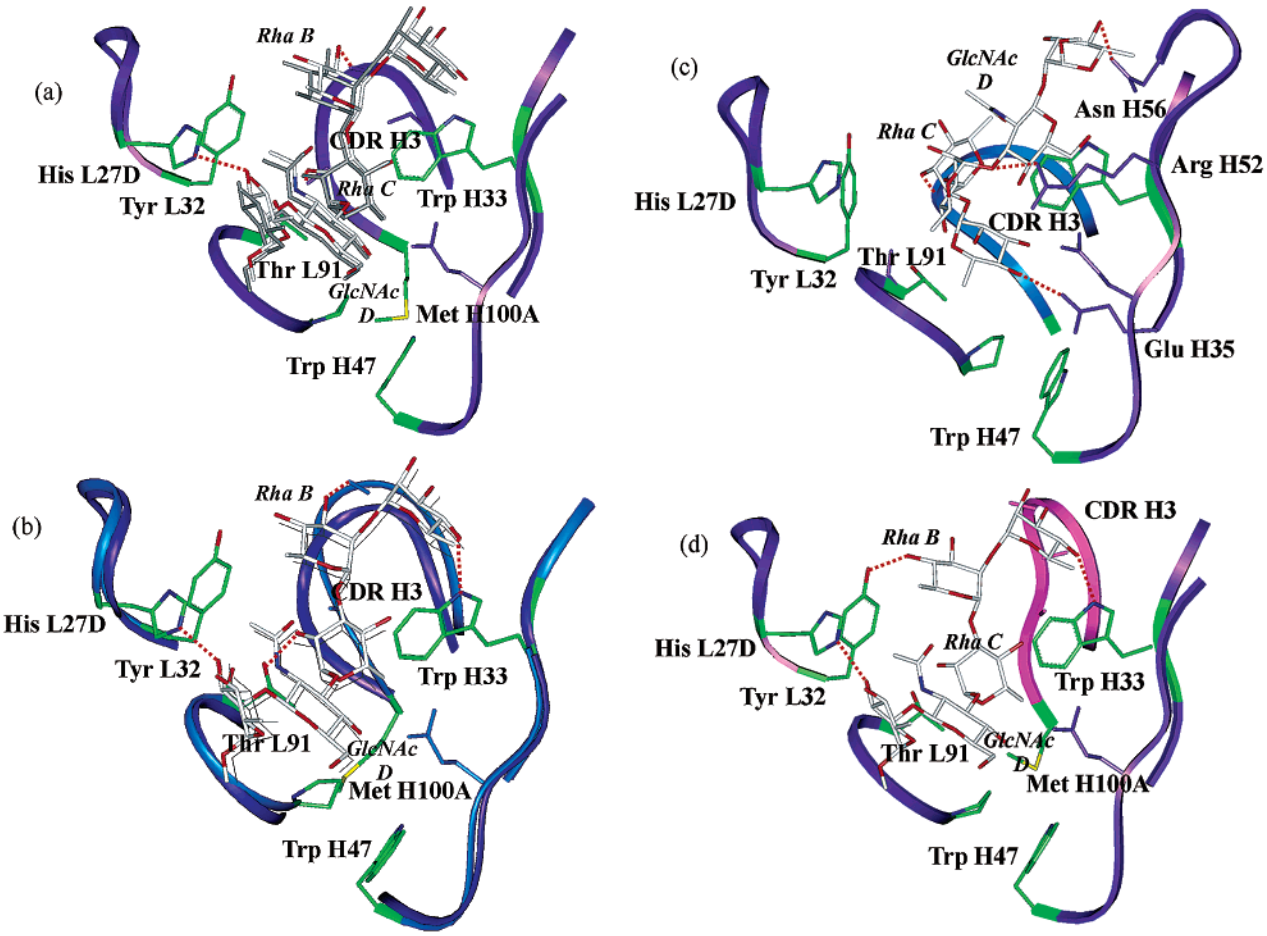


FIGURE 2: Predicted binding modes of the pentasaccharide ligand (**1**). Residues lining the combining site are shown in ribbon format, while atoms of the side chain are also shown for selected residues making specific contacts. Hydrogen bonds are shown as red dashed lines, while hydrophobic residues are colored green and represent patches of hydrophobic contact with the ligand. (a) AutoDock-predicted complex of the pentasaccharide ligand (**1**) with the correct model, consisting of the conserved antibody framework with the correct H3 loop conformation. The X-ray crystallographic position of **1** is shown for comparison (colored gray). (b) Predicted complex of **1** and the third-best model (loop rmsd = 1.72 Å). Many of the important interactions are predicted correctly, as the change in the conformation of CDR H3 is relatively small. The binding site of the model is represented as a ribbon (blue) with selected side chains in green. The ribbon structure of the same residues in the crystal structure is also shown (purple); the crystallographically determined position of **1** is shown for comparison (gray lines). (c) Predicted complex of **1** and the 11th-ranked model (loop rmsd = 2.88 Å). The very different conformation of CDR H3, highlighted in blue, results in a much smaller binding site that can accommodate only two residues of the pentasaccharide. None of the correct hydrogen bonds can be formed, although other hydrogen bonds are possible in this modeled complex. The Rha B and C and GlcNAc D methyl groups are exposed to the solvent, and the hydrophobic pocket formed by His L27D and Tyr L32 cannot be occupied by the ligand, because of the large change in the shape of the binding site. (d) Predicted complex of **1** with the 12th-ranked model (loop rmsd = 3.01 Å). The longer, narrower binding site causes the loss of hydrogen bonds to the Rha B and C residues and the exposure of the Rha C hydrophobic face to solvent.

tween protein and ligand), which is highly favorable (−11.90 kcal/mol for the complex between the pentasaccharide and best model).

As the loop conformation in the models deviates from its true conformation, some of these favorable interactions are necessarily lost. For the models quite close to the true loop conformation, this may result in the loss of only a few interactions. Interactions between the ligand and other CDR loops may be maintained, despite the change in the shape of CDR H3. An example is shown in Figure 2b for the pentasaccharide ligand and third-best antibody model (rmsd of loop = 1.72 Å). Several of the important hydrogen bonds, namely, Rha A′ O3−His L27D Nϵ1, Rha B O3−Ala H97 O, and Rha C O2−Thr L91 Oγ (*24*), are maintained in the predicted complex. The key hydrophobic interaction between the GlcNAc D *N*-acetyl methyl group and the pocket formed by Tyr L32 and His L27D (*24*) is also maintained. The change in CDR H3 conformation affects only a few specific interactions; for example, the Rha C O2−Gly H99 O and GlcNAc D HN−Thr L91 O hydrogen bonds are lost. Although there is a slight increase in the Rha C C6−Met H100A Cγ van der Waals contact distance, compensating hydrophobic contacts occur between the ligand and the backbone of the changed loop (Figure 2b).

The less accurate models have CDR H3 loop conformations that are very different from the true structure (rmsd of loop = 2−3 Å). Large changes in the overall shape of the binding site, and the loss of more of the intermolecular interactions, are observed. For example, as shown in Figure 2c, the H3 loop conformation from the 11th-ranked model (rmsd = 2.88 Å) results in a much smaller binding site, so that only two residues of the pentasaccharide can be accommodated properly, and none of the correct hydrogen bonds can be formed. The rest of the pentasaccharide lies toward the edge of the binding site and along the outer surface of the protein, and the methyl groups of Rha B and

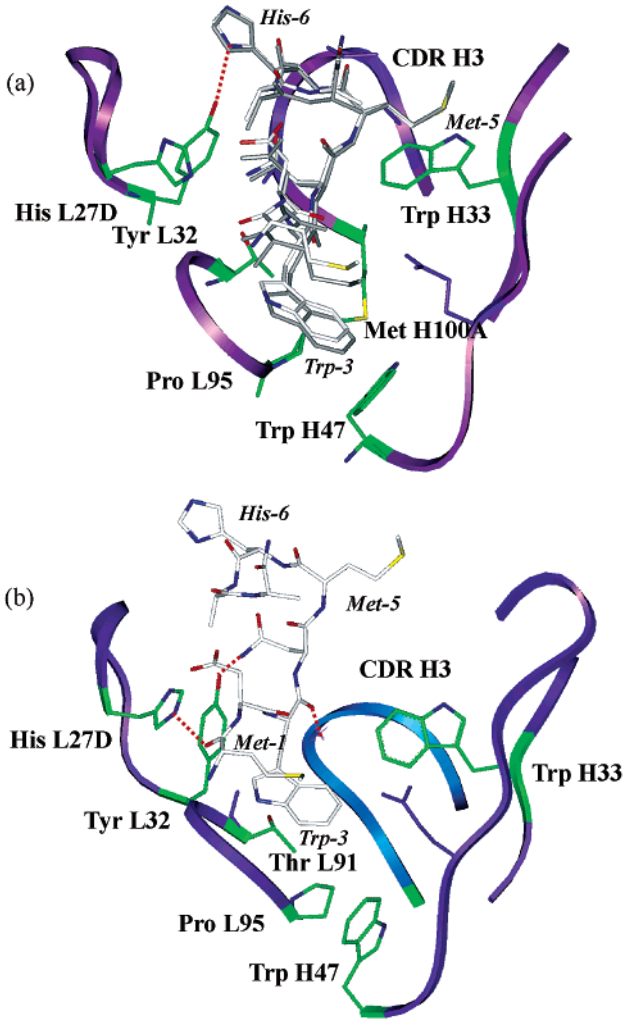


FIGURE 3: Predicted binding modes of the peptide ligand (**3**). (a) AutoDock-predicted complex of the peptide ligand with the correct model. The X-ray crystallographic position of **3** is shown for comparison (colored gray). Areas of the binding site with important hydrophobic contacts to the ligand are colored green. (b) AutoDock-predicted complex of **3** with the 11th-ranked model (loop rmsd = 2.88 Å). The small size of the binding site prevents entry of the ligand, and only Met 1 and Trp 3 are accommodated within the site; the rest of the peptide is partially or completely exposed to solvent.

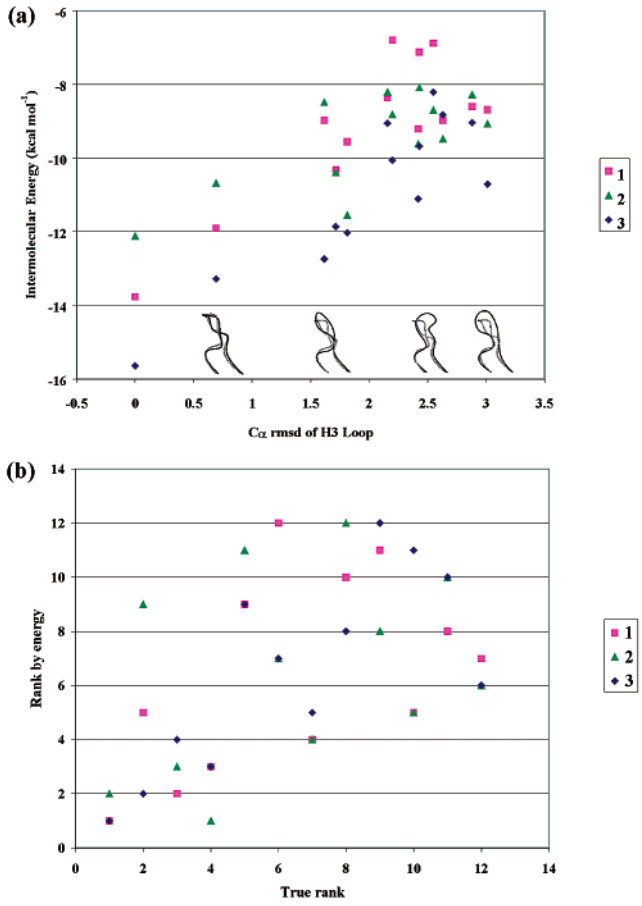


FIGURE 4: (a) Calculated intermolecular energies from AutoDock 3.0 for each model in the library, in complex with pentasaccharide (**1**), trisaccharide (**2**), and peptide (**3**) ligands, plotted against the accuracy of each model, as measured by the rms deviation of Cα atom positions of the CDR H3 loop with respect to their positions in the crystal structure. The Cα traces of loops from the library (black) with different rmsd values, with respect to the crystal structure (gray), are plotted on the graph for a visual comparison of the loops. (b) Ranking of the models based on the intermolecular energies of their complexes with the ligands, plotted against the true ranking based on the rmsd of Cα atom positions.

C and GlcNAc D are forced to be exposed to water. A much less favorable intermolecular energy (−8.60 kcal/mol) reflects that this is not a realistic binding mode.

Another possible, but incorrect, loop conformation results in a much longer, narrower binding site in the 12th-ranked model (rmsd = 3.01 Å; Figure 2d). In this case, the ligand still maintains some of the correct interactions with the other CDRs; two hydrogen bonds, Rha A′ O3−His L27D Nϵ1 and Rha B O4−Tyr L32 O, are possible, and the hydrophobic contacts of the GlcNAc D and Rha C methyl groups are maintained. However, the ligand cannot contact the backbone of CDR H3, and three hydrogen bonds are lost. In addition, the hydrophobic face of Rha C becomes partially exposed to solvent. Again, this is reflected in a much less favorable intermolecular energy (−8.69 kcal/mol).

The changes in the H3 loop conformation have similar effects on the docked complexes with the peptide ligand. Thus, the predicted complex with the correct model is very similar to the crystallographic structure of the complex (Figure 3a). Similar structures are obtained with the second- to fourth-ranked models, as the contacts between the peptide and CDR H3 are not extensive. With increasing deviation from the true loop conformation, the modeled complexes become less plausible; for example, in the 11th-ranked model, the small binding site can accommodate only two of the eight peptide residues, while the rest are placed partially or completely outside the binding site (Figure 3b).

A general trend is apparent, as shown in Figure 4a. The calculated intermolecular energies increase (become less favorable) as the models become more inaccurate. For both the pentasaccharide and peptide ligands, the correct model has the most favorable intermolecular energy of all models in the library. The second to fourth models, with rmsd values of <2 Å, have somewhat higher energies, while for the less accurate models, the energies are even higher. If the energy of the docked complexes is used as the criterion to rank the models in the library, a good correlation is observed between this ranking and the true ranking based on the rmsd of the H3 loop atoms (Figure 4b). The *r* values provide a numerical measure of this correlation; these values are 0.52, 0.36, and

0.76 for **1**−**3**, respectively (pentasaccharide, trisaccharide, and peptide, respectively). As a control, the models were screened with an unrelated ligand, maltopentaose, and no such correlation was observed ($r = 0.035$; Figure S1 of the Supporting Information).

Not surprisingly, the complexes of the ligands with the Ab crystal structure are lower in energy by 1−2 kcal/mol than those of the best model. This most likely reflects errors resulting from the comparative modeling procedure, such as differences in side chain positioning or in backbone conformation. These errors appear to have a greater effect on the positioning of the smaller trisaccharide ligand. The correct binding mode was identified, as for the other ligands, but additional incorrect binding modes were also generated. We conclude, therefore, that the larger ligands have greater discriminating power due to their characteristic shapes. If the ligand is large, there is little room available in the binding site to place it in the wrong orientation, and such binding modes are correctly evaluated by the program as being much less favorable.

Comparison with Existing Methods. The method of screening for complementarity with ligands produces a ranking of the protein models that is more accurate than that obtained using other criteria. For example, if loop conformations are generated by a database search, the level of sequence similarity between the target loop and the database loop may be used as a criterion. This may be combined with the rms fit of the anchor positions of the loop, as in Composer (*7*). However, this procedure does not necessarily lead to an accurate ranking. While a high level of sequence homology is a good indication of structural similarity in regions of regular secondary structure, this may not necessarily be true in loop or turn regions of proteins, especially in the highly variable H3 loop of antibodies. Conversely, similar loop structures may be adopted by several different sequences, as in the examples of other canonical loops of the antibody binding site (*42*−*45*). Therefore, we compared the ranking of loops in our library by ligand docking to the ranking obtained by other criteria. In Table 2 and Figure 5a, the sequence similarity between each of 12 loops in the protein database, which provided the loop conformations for the 12 models, and the sequence of the loop in mAb SYA/J6 is given. There is no apparent correlation between the level of sequence similarity in the source loop and the actual similarity of the loop conformation to the crystal structure conformation observed for SYA/J6. For example, the third most similar loop (rmsd = 1.72 Å) is actually the least similar in sequence, and would have been ranked last out of all the models by the Composer criteria. Conversely, the loop which would have been chosen by considering sequence homology and fit is actually one of the worse models (sixth in closeness to the crystal structure, rmsd = 2.20 Å). Use of the criterion of sequence homology results in only a weak correlation between the predicted ranking and the true ranking (based on accuracy) of the models in the library (Figure 5c). The *r* value for these data is 0.126, reflecting a very weak correlation.

Energy criteria may also be used to select the best loop conformation for a given model. The force field energies calculated for the models in the library are shown in Figure 5b (the force field energy refers to the internal potential energy of the protein model itself, as estimated by the Tripos force field; see also Materials and Methods). There is no apparent correlation between the calculated energy of the model and the closeness of the loop to the true structure. All of the conformations are reasonable and fit plausibly into the framework of the model. Again, there is no observable correlation between the predicted ranking of the models and the true ranking based on accuracy (Figure 5c); the *r* value of 0.035 reflects that this is not a useful predictive criterion. This is not unexpected, as the potential energy landscape of

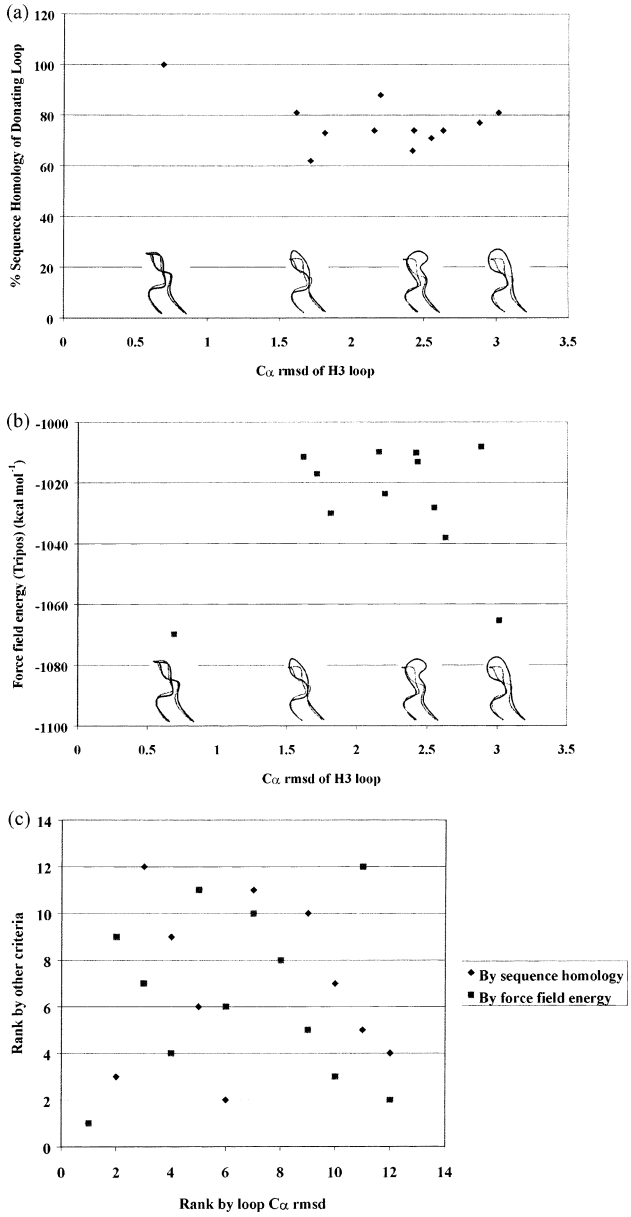


FIGURE 5: Correlations obtained with other criteria. (a) Accuracy of the models in the first library (CDR H3 loop variation), as measured by the rmsd of loop Cα atom positions with respect to their positions in the crystal structure, vs the sequence homology between the loop in the database (source of the conformation for each model) and the SYA/J6 loop. The Cα traces of loops from the library (black) with different rmsd values, with respect to the crystal structure (gray), are plotted on the graph for a visual comparison. (b) Accuracy of the models in the first library (CDR H3 loop variation), vs the potential energy of each model, calculated with the Tripos force field. (c) Ranking of the models based on the sequence homology and rms fit of the loop anchor residues, as in the Composer program, and ranking of the models based on the force field energy, vs the true rank, based on the loop Cα rmsd.

Table 3: A Series of Models Varying in Domain Orientation

source protein	rotation relative to crystal structure (deg)	translation relative to crystal structure (Å)	Trp H33 CZ3−Thr L91 O distance (Å)	sequence identity, heavy chain (%)[a]	sequence identity, light chain (%)[a]
crystal structure, SYA/J6	–	–	7.67	–	–
SYA/J6	0.99	0.45	7.99	–	–
1JHL	5.50	1.71	7.99	50	55
1MRE	7.02	0.94	8.03	50	94
1SM3	8.09	1.75	6.72	89	38
1FLR	9.53	0.14	6.63	81	92
1GGI	12.71	1.69	7.40	43	67
1GGC	20.65	1.63	5.37	43	67

[a] Percentage of exact amino acid identity.

a protein is highly complex, and there are many possible loop conformations that are correct by potential energy criteria but do not correspond to the native structure. Current force fields are often unable to discriminate between a native loop conformation and various other possible, reasonable loop conformations (refs *20* and *49* and references therein).

In contrast, effective discrimination between the correct and incorrect models is achieved by screening with ligands, using our protocol. For each ligand, a significant correlation is observed between the predicted ranking of the models and the true ranking based on accuracy (Figure 4b). Therefore, the quality of fit with a ligand is a useful criterion that could be applied to identify models with accurate loop conformations. We propose that this method could be applied effectively, when a bound-ligand conformation has been determined by trNOE NMR spectroscopy, to derive information about the shape of the binding site of the protein receptor, and to produce accurate protein models.

Orientation of Domains. Another difficult problem in the comparative modeling of antibodies is that of choosing the correct orientation of the V_H and V_L domains (*49*, *50*). The amount of rotation and translation relating these domains may vary considerably between different Abs, and even between native and bound forms of the same Ab. A summary of these quaternary structural changes for Abs whose bound and free structures are known was reported by Stanfield et al. (*36*). These authors also showed that the rotations did not occur about any particular axis of the antibody, and that the axis of rotation varied between different Abs (*36*). This variation may be relevant to mechanisms of antigen recognition. For example, changes in the V_H-V_L interface may accompany antigen binding and may serve to optimize interactions with the antigen (*37*). The variation in quaternary structure also presents a challenge in comparative modeling: how should the two domains be brought together as accurately as possible, as their relative orientation may considerably affect the shape of the binding site? For construction of a model, a framework may be chosen by using an average template based on several Abs (*50*). The V_H and V_L chains are constructed separately, superimposed on this template, and then subjected to further optimization (*50*). Alternatively, the Ab with the sequence most identical to that of the target Ab may be chosen as a template (*49*). We chose to investigate whether the correct domain orientations could be identified by docking with known ligands, in a manner analogous to the selection of correct loop conformations. Therefore, a series of seven models varying in relative domain orientation was constructed (Table 3). Each model had the correct H3 loop conformation, obtained by grafting the loop conformation from the crystal structure onto the model framework. However, the framework itself was varied by superimposing the V_H and V_L domains on different template antibodies (Figure 6a). The amount of rotation was measured with respect to the crystal structure, as described in Materials and Methods.

These models were screened by computational docking with the pentasaccharide and peptide ligands (**1** and **3**). The resulting changes in the shape of the combining site were found to be less severe than those produced by changes in the H3 loop conformation but, nevertheless, affected the intermolecular energies of the complexes. The most significant effect was a variation in the distance between the sides of the site (Table 3); the rotation caused changes in the distance between the light chain loops and the heavy chain loops, which comprise the binding site. As for the first library, the quality of docked complexes generally decreased with increasing deviation of the models from the crystal structure. The trend is shown in Figure 7, where the intermolecular energy of the complexes is plotted versus accuracy, measured by the amount of rotation relative to the crystal structure. For small rotational angles, minor repositioning of the ligand could occur to maintain a favorable, accurate docked complex. Thus, binding modes that were nearly correct (and complexes of lowest energy) were obtained with the three best antibody models, with rotation of the heavy chain of $\leq 7°$ (Figure 6b). With increased rotation, the positions of residues at the narrowest portion of the binding groove, for example, Trp H33, Arg H52, and Thr L91, were such that the groove became too narrow to accommodate the ligands correctly (Table 3). In the models with a heavy chain rotation of $8-10°$, incorrect, higher-energy complexes resulted, with the ligand displaced from its correct position or forced out of the binding site (Figure 6c, pink). One exception to the general trend of increasing intermolecular energy with increasing deviation from the crystal structure was observed at a rotational angle of 13°; at this point, the width of the groove became great enough that ligands could again enter the site. Repositioning of the ligands compensated, to some extent, for the other changes in side chain positions of the antibody, and lower-energy complexes were again obtained. Some of these complexes had errors such as missing hydrogen bonds and hydrophobic contacts, but the binding modes were closer to the true structure than those found for the $8-10°$ models. Finally, when the rotation was increased even further to 21°, the changes in side chain positions became more significant and minor movement of the ligand could no longer compensate for the error; unrealistic complexes were obtained for the

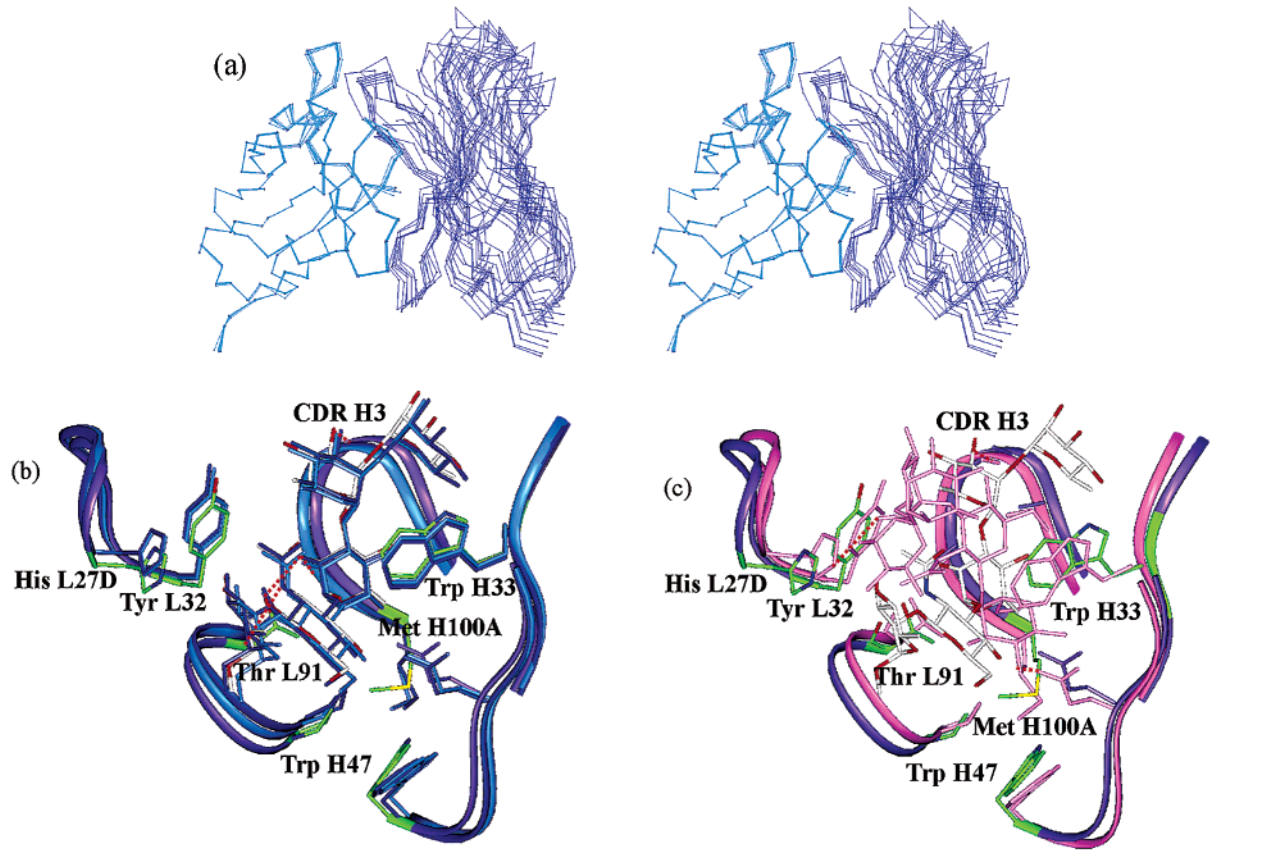


FIGURE 6: (a) Stereoview of the seven models with varying domain orientation. The models were superimposed using the Cα atoms of residues L2−L108 of the light chain. A Cα trace of the models is shown with the light chain in light blue and the heavy chain in dark blue. (b) Docked complexes calculated for the pentasaccharide (**1**) in the binding sites of the three most accurate antibody models. The correct model (framework based on the X-ray structure of SYA/J6; 0.99° rotation) is shown in ribbon representation (purple), and selected side chains are shown (green). The two models with the least domain rotation relative to the crystal structure are also shown (dark blue, 5.5° rotation; light blue, 7.02° rotation). Hydrogen bonds are indicated as dashed lines. The models were superimposed using the Cα atoms of all CDR residues. Similar binding modes are found for these three models. (c) Docked complex calculated for one of the less accurate antibody models, with 9.53° rotation (magenta ribbon, pink side chains). The ligand (pink) is displaced due to narrowing of the binding site. The binding mode predicted for the correct model is shown for comparison (purple for the binding site, green for side chains, and white for pentasaccharide).

least accurate antibody model. Similar trends were observed for both pentasaccharide and peptide ligands (Table 3 and Figure 7).

In this series of models, the quality of docked complexes correlated well with accuracy. It is noteworthy that the quality of the models did not correlate with the level of sequence identity between the template and target antibodies. For example, the sequence of Ab 4-4-20 (PDB entry 1FLR) (*51*) is 92 and 81% identical to that of SYA/J6 in the light and heavy chains, respectively, greater than any of the other Abs used as templates for the rotation series (Table 3). However, because of the positions of certain side chains, this amount of rotation led to the narrowest binding site, and incorrect binding modes were obtained with this model. Therefore, the level of sequence identity within the subset of residues involved in the V_H-V_L interface may be a better guide than the overall level of sequence identity, for the choice of a template antibody. Although the CDR residues comprise only ~25% of the interface (*21*), they may have significant effects on its structure (*37*); therefore, similarity in these regions should also be considered. Again, we note that when there is uncertainty, a series of models representing various possibilities may be constructed and evaluated on the basis of their complementarity to known ligands. We find that the highest-quality complexes, those with the best complementarity and most favorable binding energy, generally represent the highest-quality protein models, in terms of closeness to the true protein structure. These models would also be the

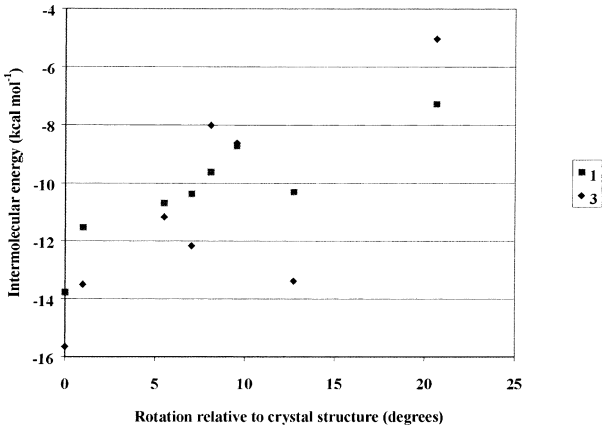


FIGURE 7: Calculated intermolecular energies of the complexes of the pentasaccharide (**1**) and peptide (**3**) ligands with each model in the second library, in which the models have the correct CDR H3 loop conformation but vary in the relative orientation of the V_H and V_L domains, plotted vs accuracy of the models, estimated by the amount of rotation relative to the crystal structure.

most useful in analyzing factors responsible for ligand recognition, and in the design of new ligands (antigens or inhibitors).

General Applicability of the Method. The quality of fit with a known ligand is a useful criterion that could be applied to identify the most accurate models out of a series. We propose that this method could be applied effectively, when a bound-ligand conformation has been determined by transferred nuclear Overhauser effect (trNOE) NMR spectroscopy, to derive information about the shape of the binding site of the protein receptor, and to produce accurate protein models. The transferred nuclear Overhauser effect (*12*−*16*) has been shown to be useful in investigating bound ligand conformations (*52*−*56*). TrNOE NMR-derived conformations have even been combined with protein crystal structures (*57*−*61*) or with protein models (*62*, *63*) to generate models of the mode of interaction. However, to the best of our knowledge, this study is the first example of *the use of the known conformation of a ligand as a screening tool for protein models—in effect, to indirectly derive information about the protein structure.* This protocol provides a method of incorporating more experimental data to discriminate between possibilities of similar likelihood at the level of theory used. The method works because of the complementarity of fit between the protein binding site and the ligand.

As in any application of computational methods, there are certain possibilities for error which must be considered. In preliminary studies, we found that incorrect positioning of side chains within the binding site sometimes prevented proper positioning of the bound ligand if there was a close interaction with that area of the binding site. Therefore, side chain conformations should be modeled carefully, with torsional angles being taken from the homologous known proteins wherever possible; several possibilities should be evaluated in the case of substitutions for which no structure is known. Modern methods for the choice of the most likely side chain conformations could also be effective (*64*). Our protocol should be considered not as a way of obtaining final answers, but as a way of rapidly identifying accurate protein models that could then be subjected to further optimization.

It should also be noted at this point that in our test case, antibody SYA/J6, there are no significant differences between the free and ligand-bound Fab structures (*24*). The structure of the antibody, including the binding site, is very similar in the free state and when bound to any of the ligands **1**−**3**. This "lock-and-key" type of interaction is observed for many protein−ligand systems (*65*), but there are also many examples for which rearrangements of the protein accompany binding (*65*, *66*). In such cases, our method would necessarily only produce an accurate model of the bound structure. Although the bound structure is only part of the "story" of any protein−ligand interaction system, it is likely to be of greater interest than the free structure in terms of understanding the basis of binding, examination of specific intermolecular interactions, and design of other ligands and inhibitors.

CONCLUSIONS

The complementary fit of protein and ligand provides a quick screening process in protein modeling. A library of many different possible models may be screened to identify one or a few of the most likely possibilities, which give rise to the best shape of the binding site. Loop conformations for a library may be generated by several different methods; here, a database search has been used, but simulated annealing or conformational searching is also useful (refs *18*−*20* and references therein). Our method provides a fast route to the identification of correct or nearly correct loop conformations. Beginning the modeling process by identifying one or a few of these would allow more computational effort to be focused on refining and optimizing these nearly correct models.

Our protocol is not limited to high-affinity ligands, as those described above have affinity constants of 10^5−10^6 M^{-1} (*24*). The protocol is not limited to "natural" ligands and could be applied using enzyme inhibitors, haptens, or even larger molecules such as proteins. As shown in our study, medium- to larger-sized ligands, such as peptides or oligosaccharides, appeared to have greater discriminating power with respect to this particular binding site. The ligand size that is most appropriate will likely be determined by the size of the binding site on the protein; for example, a cavity-shaped binding site on an enzyme that acts on small substrates, or the combining site of an anti-hapten antibody, would be complementary to smaller ligands.

The power of our method is that the uncertainty inherent in current protein modeling techniques is not avoided, nor is excess computer time spent trying to resolve it. Rather, it is incorporated directly into the library to be evaluated. Thus, diversity may be incorporated into the library according to the specific problem. Although the modeling of loop conformations is one of the most common challenges (*10*, *17*−*20*), there are many other variables which could be incorporated into the library, including side chain torsional angles, the relative orientation of domains or individual secondary structure elements, or even different sequence alignments, in the case of more distantly related proteins.

ACKNOWLEDGMENT

We are grateful to F. A. Quiocho and N. K. Vyas for providing coordinates of the crystal structures of the SYA/J6 Fab and its complexes prior to publication. In addition, we thank R. J. Batchelor for helpful discussions regarding the transformations of protein domains.

SUPPORTING INFORMATION AVAILABLE

Plot of the intermolecular energies calculated for the unrelated ligand, maltopentaose, binding to the models in the first library, and ranking of the models based on these energies. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., et al. (1998) New goals for the U.S. human genome project: 1998−2003, *Science 282*, 682−689.
2. Burley, S. K., Almo, S. C., Bonanno, J. B., Capel, M., Chance, M. R., Gaasterland, T., et al. (1999) Structural genomics: beyond the human genome project, *Nat. Genet. 23*, 151−157.
3. Maggio, E. T., and Ramnarayan, K. (2001) Recent developments in computational proteomics, *Trends Biotechnol. 19*, 266−272.
4. Teichmann, S. A., Murzin, A. G., and Chothia, C. (2001) Determination of protein function, evolution and interactions by structural genomics, *Curr. Opin. Struct. Biol. 11*, 354−363.

5. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., et al. (2000) Structural proteomics of an archaeon, *Nat. Struct. Biol. 7*, 903−909.
6. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. (1987) Knowledge-based prediction of protein structures and the design of novel molecules, *Nature 326*, 347−352.
7. Blundell, T., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., et al. (1988) Knowledge-based protein modelling and design, *Eur. J. Biochem. 172*, 513−520.
8. Šali, A., Overington, J. P., Johnson, M. S., and Blundell, T. L. (1990) From comparisons of protein sequences and structures to protein modelling and design, *Trends Biochem. Sci. 15*, 235−240.
9. Sánchez, R., and Šali, A. (1997) Advances in comparative protein-structure modelling, *Curr. Opin. Struct. Biol. 7*, 206−214.
10. Sánchez, R., Pieper, U., Melo, F., Eswar, N., Martí-Renom, M., Madhusudhan, M. S., et al. (2000) Protein structure modeling for structural genomics, *Nat. Struct. Biol. 7* (Suppl.), 986−990.
11. Sánchez, R., and Šali, A. (1999) Comparative protein structure modeling in genomics, *J. Comput. Phys. 151*, 388−401.
12. Balaram, P., Bothner-By, A. A., and Dadok, J. (1972) Negative nuclear Overhauser effects as probes of macromolecular structure, *J. Am. Chem. Soc. 94*, 4015−4016.
13. Balaram, P., Bothner-By, A. A., and Breslow, E. (1972) Localization of tyrosine at the binding site of neurophysin II by negative nuclear Overhauser effects, *J. Am. Chem. Soc. 94*, 4017−4018.
14. Albrand, J. P., Birdsall, B., Feeney, J., Roberts, G. C. K., and Burgen, A. S. V. (1979) The use of transferred nuclear Overhauser effects in the study of the conformations of small molecules bound to proteins, *Int. J. Biol. Macromol. 1*, 37−41.
15. Clore, G. M., and Gronenborn, A. M. (1982) Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins, *J. Magn. Reson. 48*, 402−417.
16. Clore, G. M., and Gronenborn, A. M. (1983) Theory of the time dependent transferred nuclear Overhauser effect: applications to structural analysis of ligand-protein complexes in solution, *J. Magn. Reson. 53*, 423−442.
17. Al-Lazikani, B., Jung, J., Xiang, Z., and Honig, B. (2001) Protein structure prediction, *Curr. Opin. Chem. Biol. 5*, 51−56.
18. Vásquez, M., Némethy, G., and Scheraga, H. A. (1994) Conformational energy calculations on polypeptides and proteins, *Chem. Rev. 94*, 2183−2239.
19. Ring, C. S., and Cohen, F. E. (1994) Conformational sampling of loop structures using genetic algorithms, *Isr. J. Chem. 34*, 245−252.
20. Fiser, A., Do, R. K. G., and Šali, A. (2000) Modeling of loops in protein structures, *Protein Sci. 9*, 1753−1773.
21. Padlan, E. A. (1994) Anatomy of the antibody molecule, *Mol. Immunol. 31*, 169−217.
22. Carlin, N. I. A., Gidney, M. A. J., Lindberg, A. A., and Bundle, D. R. (1986) Characterization of *Shigella flexneri*-specific murine monoclonal antibodies by chemically defined glycoconjugates, *J. Immunol. 137*, 2361−2366.
23. Vyas, M. N., Vyas, N. K., Meikle, P. J., Sinnott, B., Pinto, B. M., Bundle, D. R., and Quiocho, F. A. (1993) Preliminary crystallographic analysis of a Fab specific for the O-antigen of *Shigella flexneri* cell surface lipopolysaccharide with and without bound saccharides, *J. Mol. Biol. 231*, 133−136.
24. Vyas, N. K., Vyas, M. N., Chervenak, M. C., Johnson, M. A., Pinto, B. M., Bundle, D. R., and Quiocho, F. A. (2002) Molecular recognition of oligosaccharide epitopes by a monoclonal Fab specific for *Shigella flexneri* Y lipopolysaccharide: X-ray structures and thermodynamics, *Biochemistry 41*, 13575−13586.
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235−242.
26. Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A. M. (1998) Conformations of the third hypervariable region in the VH domain of immunoglobulins, *J. Mol. Biol. 275*, 269−294.
27. Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1979) *Atlas of Protein Sequence and Structure 5* (Suppl. 3), 345.
28. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S., and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, 5th ed., NIH Publication 91-3242, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD.
29. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., and Hagler, A. T. (1988) Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system, *Proteins 4*, 31−47.
30. Derrick, J. P., Maiden, M. C. J., and Feavers, I. M. (1999) Crystal structure of an Fab fragment in complex with a meningococcal serosubtype antigen and a protein G domain, *J. Mol. Biol. 293*, 81−91.
31. Claessens, M., Van Cutsem, E., Lasters, I., and Wodak, S. (1989) Modelling the polypeptide backbone with "spare parts" from known protein structures, *Protein Eng. 2*, 335−345.
32. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
33. Clark, M., Cramer, R. D., and Van Opdenbosch, N. (1989) Validation of the general purpose Tripos 5.2 force field, *J. Comput. Chem. 10*, 982−1012.
34. McLachlan, A. D. (1982) Rapid comparison of protein structures, *Acta Crystallogr. A38*, 871−873.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
36. Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T., and Wilson, I. A. (1993) Major antigen-induced domain rearrangements in an antibody, *Structure 1*, 83−93.
37. Colman, P. M. (1988) Structure of antibody−antigen complexes: implications for immune recognition, *Adv. Immunol. 43*, 99−132.
38. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem. 19*, 1639−1662.
39. Rao, M. S., and Olson, A. J. (1999) Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach, *Proteins 34*, 173−183.
40. Fujimoto, Z., Takase, K., Doui, N., Momma, M., Matsumoto, T., and Mizuno, H. (1998) Crystal structure of a catalytic-site mutant α-amylase from *Bacillus subtilis* complexed with maltopentaose, *J. Mol. Biol. 277*, 393−407.
41. Lowry, R. (1999−2002) *Concepts and applications of inferential statistics* (Web textbook) (http://faculty.vassar.edu/lowry/webtext.html).
42. Chothia, C., and Lesk, A. M. (1987) Canonical structures for the hypervariable regions of immunoglobulins, *J. Mol. Biol. 196*, 901−917.
43. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., et al. (1989) Conformations of immunoglobulin hypervariable regions, *Nature 342*, 877−883.
44. Martin, A. C., and Thornton, J. M. (1996) Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies, *J. Mol. Biol. 263*, 800−815.
45. Al-Lazikani, B. A., Lesk, A. M., and Chothia, C. (1997) Standard conformations for the canonical structures of immunoglobulins, *J. Mol. Biol. 273*, 927−948.
46. Vargas-Madrazo, E., Lara-Ochoa, F., and Almagro, J. C. (1995) Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition, *J. Mol. Biol. 254*, 497−504.
47. Harris, S. L., Craig, L., Mehroke, J. S., Rashed, M., Zwick, M. B., Kenar, K., et al. (1997) Exploring the basis of peptide-carbohydrate crossreactivity: Evidence for discrimination by peptides between closely related anti-carbohydrate antibodies, *Proc. Natl. Acad. Sci. U.S.A. 94*, 2454−2459.
48. Makarov, V. A., Andrews, B. K., Smith, P. E., and Pettitt, B. M. (2000) Residence times of water molecules in the hydration sites of myoglobin, *Biophys. J. 79*, 2966−2974.
49. Thornton, J. M. (1991) Modelling antibody combining sites: a review, *Ciba Found. Symp. 159*, 55−71.
50. Webster, D. M., and Rees, A. R. (1995) Molecular modeling of antibody combining sites, *Methods Mol. Biol. 51*, 17−49.
51. Whitlow, M., Howard, A. J., Wood, J. F., Voss, E. W., and Hardman, K. D. (1995) 1.85 Å structure of anti-fluorescein 4-4-20 Fab, *Protein Eng. 8*, 749−761.
52. Campbell, A. P., Wong, W. Y., Irvin, R. T., and Sykes, B. D. (2000) Interaction of a bacterially expressed peptide from the receptor binding domain of *Pseudomonas aeruginosa* pili strain

PAK with a cross-reactive antibody: conformation of the bound peptide, *Biochemistry 39*, 14847−14864.
53. LaPlante, S. R., Aubry, N., Bonneau, P. R., Cameron, D. R., Lagacé, L., Massariol, M. J., et al. (1998) Human cytomegalovirus protease complexes its substrate recognition sequences in an extended peptide conformation, *Biochemistry 37*, 9793−9801.
54. Scheffler, K., Brisson, J.-R., Weisemann, R., Magnani, J. L., Wong, W. T., Ernst, B., and Peters, T. (1997) Application of homonuclear 3D NMR experiments and 1D analogs to study the conformation of sialyl Lewis^x bound to E-selectin, *J. Biomol. NMR 9*, 423−436.
55. Sykes, B. D. (1993) Determination of the conformations of bound peptides using NMR-transferred nOe techniques, *Curr. Opin. Biotechnol. 4*, 392−396.
56. Campbell, A. P., and Sykes, B. D. (1993) The two-dimensional transferred nuclear Overhauser effect: theory and practice, *Annu. Rev. Biophys. Biomol. Struct. 22*, 99−122.
57. Weimar, T., Stoffer, B., Svensson, B., and Pinto, B. M. (2000) Complexes of glucoamylase with maltoside heteroanalogues: bound ligand conformations by use of transferred NOE experiments and molecular modeling, *Biochemistry 39*, 300−306.
58. Glover, N. R., and Tracey, A. S. (1999) Nuclear magnetic resonance and restrained molecular dynamics studies of the interaction of an epidermal growth factor-derived peptide with protein tyrosine phosphatase 1B, *Biochemistry 38*, 5256−5271.
59. Maurer, M. C., Trosset, J.-Y., Lester, C. C., DiBella, E. E., and Scheraga, H. A. (1999) New general approach for determining the solution structure of a ligand bound weakly to a receptor: structure of a fibrinogen Aα-like peptide bound to thrombin (S195A) obtained using NOE distance constraints and an ECEPP/3 flexible docking program, *Proteins 34*, 29−48.
60. Casset, F., Hamelryck, T., Loris, R., Brisson, J.-R., Tellier, C., Dao-Thi, M. H., et al. (1995) NMR, molecular modeling, and crystallographic studies of lentil lectin-sucrose interaction, *J. Biol. Chem. 270*, 25619−25628.
61. Asensio, J. L., Cañada, F. J., and Jiménez-Barbero, J. (1995) Studies of the bound conformations of methyl α-lactoside and methyl β-allolactoside to ricin B chain using transferred NOE experiments in the laboratory and rotating frames, assisted by molecular mechanics and dynamics calculations, *Eur. J. Biochem. 233*, 618−630.
62. Pitner, J. B., Beyer, W. F., Venetta, T. M., Nycz, C., Mitchell, M. J., Harris, S. L., et al. (2000) Bivalency and epitope specificity of a high-affinity IgG3 monoclonal antibody to the *Streptococcus* Group A carbohydrate antigen. Molecular modeling of a Fv fragment, *Carbohydr. Res. 324*, 17−29.
63. Kleinjung, J., Petit, M.-C., Orlewski, P., Mamalaki, A., Tzartos, S., Tsikaris, V., et al. (2000) The third-dimensional structure of the complex between an Fv antibody fragment and an analogue of the main immunogenic region of the acetylcholine receptor: a combined two-dimensional NMR, homology, and molecular modeling approach, *Biopolymers 53*, 113−128.
64. De Maeyer, M., Desmet, J., and Lasters, I. (1997) All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of side chains by dead-end elimination, *Folding Des. 2*, 53−66.
65. Sundberg, E. J., and Mariuzza, R. A. (2002) Molecular recognition in antibody−antigen complexes, *Adv. Protein Chem. 61*, 119−160.
66. Steitz, T. A., Harrison, R., Weber, I. T., and Leahy, M. (1983) Ligand-induced conformational changes in proteins, *Ciba Found. Symp. 93*, 25−46.

BI020608F