PROTEINS: Structure, Function, and Genetics 11:29–34 (1991)

Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method

Andrew Miranker and Martin Karplus
Committee on Higher Degrees in Biophysics and Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT **A new method is proposed for determining energetically favorable positions and orientations for functional groups on the surface of proteins with known three-dimensional structure. From 1,000 to 5,000 copies of a functional group are randomly placed in the site and subjected to simultaneous energy minimization and/or quenched molecular dynamics. The resulting functionality maps of a protein receptor site, which can take account of its flexibility, can be used for the analysis of protein ligand interactions and rational drug design. Application of the method to the sialic acid binding site of the influenza coat protein, hemagglutinin, yields functional group minima that correspond with those of the ligand in a cocrystal structure.**

Key words: drug-design, ligand-binding, hemagglutinin, functional groups, MCSS

INTRODUCTION

The design of ligands that bind strongly to key regions of biologically important molecules (e.g., enzyme active sites, receptor proteins) so as to inhibit or alter their activity is the essence of drug design. Rational approaches to this problem are still limited.[1] In this report, we propose a method based on energy minimization/quenched molecular dynamics for the determination of energetically favorable positions of functional groups in the binding sites of proteins with known three-dimensional structure. The method can aid in the design of molecules that incorporate such functional groups by modification of known ligands or *de novo* construction.

To determine and characterize the local minima of a functional group in the force field of a protein, multiple copies of selected functional groups are first distributed in the binding site of interest. Energy minimization of these copies by molecular mechanics or quenched dynamics yields the distinct local minima. The neighborhood of these minima can then be explored by a grid search or by constrained minimization. An essential element of the approach is the classical time dependent Hartree (TDH) approximation,[2] which can be used to simultaneously minimize or quench many identical groups in the force field of a single protein or of a protein with multiple copies of important residues. The proposed method provides a useful supplement to more standard visual (graphical) approaches[3] and grid search techniques.[4,5]

METHOD

Implementation of the multiple copy simultaneous search method (MCSS) requires a choice of functional groups and a molecular mechanics model for them. Groups must be simple enough to be easily characterized and manipulated (3–6 atoms, few or no dihedral degrees of freedom), yet complex enough to approximate the steric and electrostatic interactions that the functional group would have in binding to the site of interest. An appropriate set is one in which most organic molecules can be described as a collection of such groups.[6] This includes fragments such as acetonitrile, methanol, acetate, methyl ammonium, dimethyl ether, methane, and acetaldehyde.[7] Others have used functionality descriptions with fewer atoms,[8] or a spherical approximation to a multi-atom group.[4] An application of such an atom-based grid search to the docking of molecules by simulated annealing has been described recently.[9] Whereas energetically favorable positions for these smaller fragments are simpler to calculate, such positions are likely to be less useful because they may be sterically unattainable for a larger ligand. Moreover, in the design of ligands it is important to have orientational, as well as positional, information about possible functional group binding sites.

In the present application, parameters for the functional groups were chosen from the polar hydrogen set of CHARMM 20, augmented as necessary by information from *ab initio* calculations. This set treats aliphatic hydrogens as part of the "extended" carbon atom to which they are bonded; polar hydrogens are treated explicitly. A 7.5Å shifted nonbonded cutoff and a constant dielectric of one were used. The exact choice (e.g., distance dependant di-

Received November 28, 1990; revision accepted February 19, 1991.

Address reprint requests to Dr. Martin Karplus, Department of Chemistry, Harvard University, 12 Oxford St., Cambridge, MA 02138.

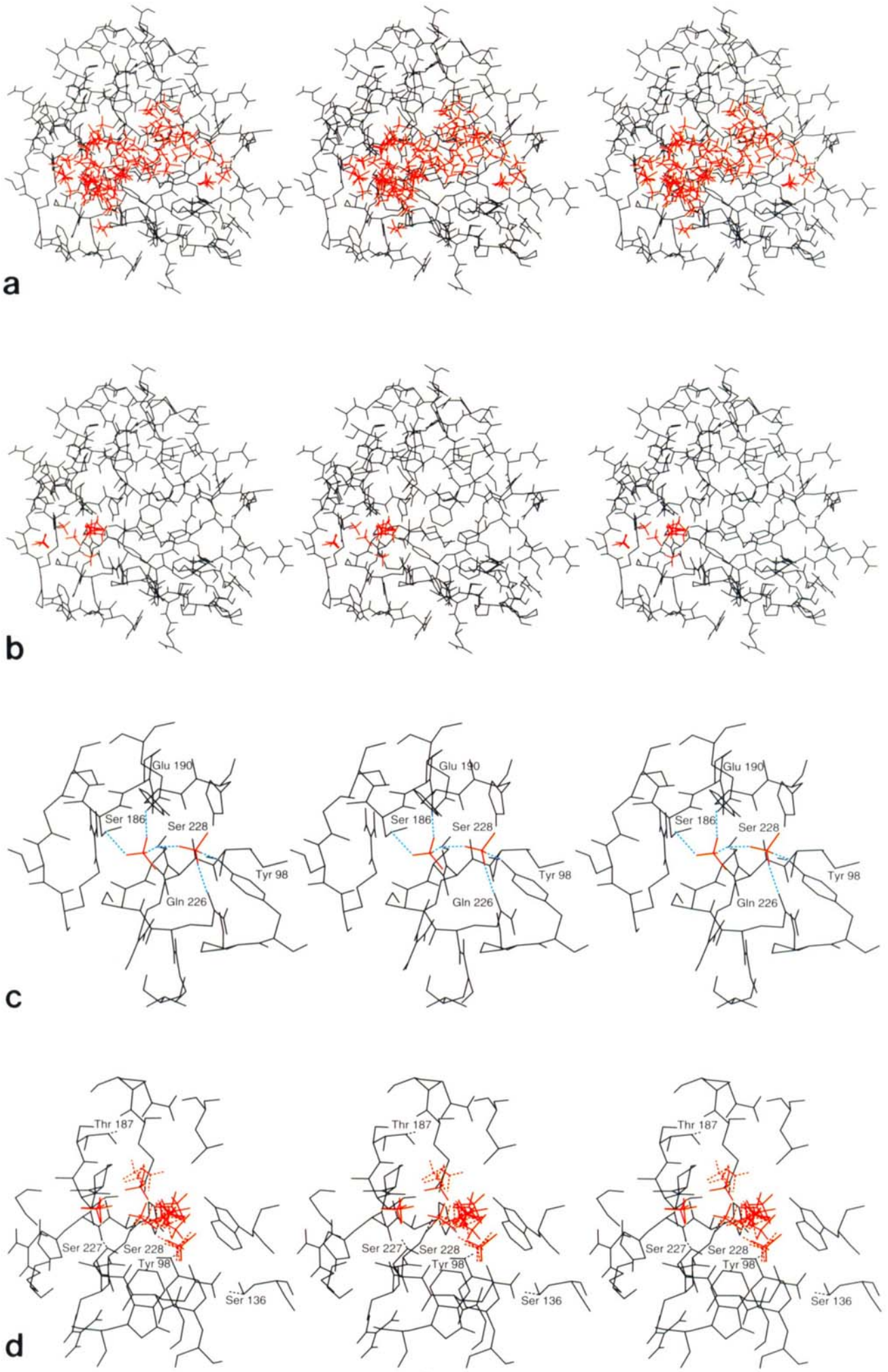
a
b
Glu 190
Ser 186
Ser 228
Tyr 98
Gln 226
c
Thr 187
Ser 227
Ser 228
Tyr 98
Ser 136
d

Fig. 1.

electric vs. a constant dielectric) is expected to have little effect on the positions of the observed minima. A shorter cutoff would speed up the calculations, but might miss longer range electrostatic interactions that could be important.[10] High internal force constants were used to maintain geometry during TDH minimization. Partial charges for atoms not included in CHARMM 20 parameters are estimated from 6–31G* *ab initio* calculations scaled by 0.83. This scaling factor was derived from comparisons with charges fitted to interaction energies.[11]

Determination of the local energy minima in the binding site requires that many starting positions be sampled. This is achieved by distributing 1,000–5,000 groups at random inside a sphere centered on the binding site; only the space not occupied by the protein needs to be considered. If the interaction energy of the group with the protein is more positive than a given cutoff (5.0 kcal/mole was used, but the choice is somewhat arbitrary), the group is discarded. The procedure provides adequate coverage if several thousand copies are accepted for subsequent manipulation. Given the set of starting positions, all the fragments are minimized simultaneously by use of the TDH approximation.[2] In this method, the forces on each fragment consist of its internal forces and those due to the protein. The essential element of the method is that the interactions between the fragments are omitted and the forces on the protein are normalized to those due to a single fragment. In this way, simultaneous minimization or dynamics of any number of functional groups in the field of a single protein can be performed. Also, protein flexibility and/or multiple copies of parts of the protein (e.g., all hydrogens with a flexible torsion) can be treated in a straitforward manner. For the rigid protein used in the present study, the TDH approximation is still employed because it makes very efficient use of the non-bonded atom lists characteristic of molecular mechanics calculations.

Minimization is performed successively on subsets of 100 randomly placed groups. In 1,000-step intervals, the results are examined to eliminate groups converging to the same minimum. Pairs of groups are considered to be converging to a common minimum or to a minimum of another subset if these RMS coordinate separation is less than 0.2Å and if the RMS distance decreased in the last 200 steps. Only one of every pair with this property are kept for subsequent minimization. The use of subsets and the elimination of overlapping pairs reduces the magnitude of the problem, lowers computer memory requirements, and allows minima to be examined prior to completion of the full search. This process is repeated until minimization is complete (RMS gradient of 0.01 kcal/mole/Å); typically, 3,000–6,000 steps of minimization are required. For 5,000 starting positions, depending on the functionality, the entire process requires several hours of CPU on a single processor of a Silicon Graphics Power Series 220. The method is very efficient because the TDH algorithm simultaneously minimizes 100 or more groups in the force field of a single protein.

After minimization, a limited number of the interesting low energy minima are selected for further investigation by a grid search or random sampling within the region of the minimum. A grid search was used at this stage because the search of the neighborhood of a few TDH energy minima (on the order of 10Å^3 per minimum) requires little computer time. This contrasts with the time requirements of a grid search to find minima in the entire binding site, which has an effective volume on the order of 1,000Å.[3] For the latter, relatively large grids have been used (e.g., 0.5 to 1Å), which may overlook multiple minima whose centers of mass are separated by smaller distances. The local grid search used here is done with a spacing of 0.25Å. Our use of multi atom functional groups and the CHARMM molecular mechanics force field necessitated the implementation of a grid search consistent with this model, which is different from that used in the available programs.[4,5] The center of mass of the group is placed at a grid point and rigid body minimization is performed only allowing rotational freedom. Since the energy surface for a group at a particular grid point can have local minima, the global minimum is determined by taking the starting orientation of the probe at each grid point from a set of 100–200 random orientations; i.e., of the random orientations, the one that has the largest interaction energy with the protein is chosen as the starting point for minimization. Depending on the complexity of the functional group and the number of randomly sampled orientations, CPU requirements for this step were on the order of seconds/gridpoint on a SUN 4/110.

RESULTS AND DISCUSSION

To test the MCSS method with a realistic and intrinsically interesting system, a functionality map of the binding site of the influenza virus hemagglutinin was made.[12] Since the refined structures of free and sialic acid liganded hemagglutinin are very similar, conformational changes on ligand binding are small and the fixed protein approximation is appropriate as a first step. An X-PLOR[13] refined struc-

Fig. 1. TDH minimization of methyl ammonium in the binding site of hemagglutinin. Stereo triple projections are shown. Protein atoms are black; methyl ammonium is shown in red. **(a):** 100 of 5,000 random starting positions. **(b):** After TDH minimization, there are 12 minima with E < −39.0 kcal/mol. All the minima make at least two hydrogen bonds with the protein. **(c):** Two minima in which methyl ammonium satisfies all three of its hydrogen bond donors. **(d):** Comparison of minima determined with X-PLOR structure versus minima determined using the same structure but with altered hydrogens. Altered hydrogens and corresponding minima are shown as dotted lines. Hydrogen dihedral changes include: Tyr 98 $\Delta\phi$ = 148°, Ser 136 $\Delta\phi$ = 70°, Thr 187 $\Delta\phi$ = 56°, Ser 227 $\Delta\phi$ = 107°, Ser 228 $\Delta\phi$ = 18°.

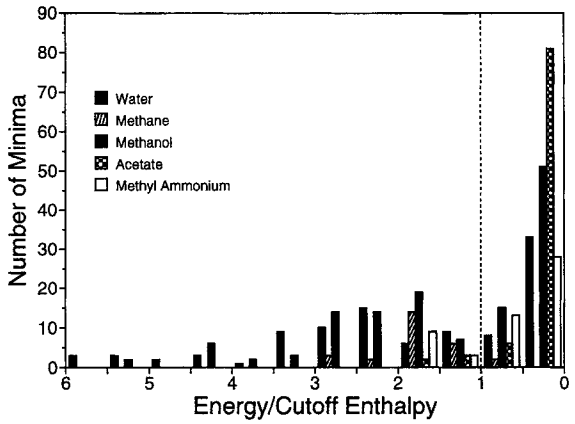

Fig. 2. Histogram of the number of minima vs. the interaction energy with the hemagglutinin binding site. The interaction energies are normalized with the cutoff used to eliminate poorly interacting minima: water $F_{cut} = -5.0$ kcal/mol, methanol $E_{cut} = -5.1$ kcal/mol, methane $E_{cut} = -1.4$ kcal/mol,[19] methyl ammonium $E_{cut} = -37.5$ kcal/mol, acetate E_{cut} −46.5 kcal/mol.[20]

ture for the unliganded protein including polar hydrogen coordinates was used for the calculations.[14]

To start the search process, 1,000–5,000 functional groups (acetate, methanol, methyl ammonium, methane, and water) were distributed inside a 12Å sphere centered on the ζ2 carbon of Trp 153 (Fig. 1a). The sphere radius was chosen to be 2Å greater than the distance from this carbon to the most distant atom of the sialic acid ligand in the co-crystal structure. After minimization, there were between 27 and 133 minima, depending on the functional group, with a broad range of interaction energies (Fig. 2). Since this is too many minima to catalogue effectively and interpret, only groups with interaction energies less than a cutoff were examined. As a convenient cutoff that reflects the importance of the relative binding energy to the protein and the solvent, one-half the solvation enthalpy of the particular group was used (see caption to Fig. 2). For methyl ammonium, this yielded 12 minima from 5,000 starting positions (Fig. 1b). Interpretation of the minima is further simplified because the TDH minima typically form clusters. For example, the 12 methyl ammonium minima form 5 clusters, including a group of 7, a group of 3 and 3 groups each containing 1 minimum (Fig. 1c). The presence of several minima in close proximity suggests that several binding modes may be possible in this region of the binding site.

The ranges of the interaction energies with the protein of the minimized positions can be grouped into three classes corresponding to charged, polar and nonpolar functionalities (see Fig. 2). The strongest interactions were those of the charged groups with the best minima having energies of −68 and −76 kcal/mol for methyl ammonium and acetate respectively. Polar groups, like methanol or water, had the greatest number of minima with the best energies equal to −27 to −29 kcal/mol, respectively. It can be seen that several waters and methanols are in minima with interaction energies five to six times the cutoff value. Most of these are dominated by the contributions from single charged amino acids (Lys, Arg, Asp, Glu), although some of the minima involve strong interactions with several polar groups. Methane minima are distributed over much of the binding site and the lowest minima have only a small interaction energy (−4 kcal/mol).

Visual inspection of a binding site with modern molecular graphics programs is one approach that is frequently used to find optimal positions for functional groups.[3] For a group like methanol, it is relatively simple to find locations where a single hydrogen bond is satisfied, but it is already more complex to find positions and orientations where the methanol acts as both a hydrogen bond donor and acceptor. The present method found 67 minima for methanol with interaction energies less than −5.1 kcal/mol, 21 of which formed two or more hydrogen bonds with the hemagglutinin molecule. A more complex group like methyl ammonium can form three hydrogen bonds with good geometry. Figure 1c shows two minima in which three hydrogen bonds are donated to the protein. Because of the complexity of the interactions at these positions, it would be more difficult to find such minima by visual examination.

For the polar and charged groups, the most energetically favorable minima usually involve hydrogen bonds with the protein. It is important, therefore, to evaluate the sensitivity of the functionality maps to the inherent flexibility of the protein with respect to proton positions. This is particularly true because hydrogen atom placement during crystallographic refinement is done by inspection or by dynamical simulation, unless neutron data are available. One approach is to use the TDH method[2] in the quenched dynamics mode and include polar mobile hydrogens, as well as functional groups. In the present study, the sensitivity of the minima to proton positions was assessed by generating a hemagglutinin structure with alternative hydrogen positions. A 1,000° K dynamics simulation in which only the polar hydrogens were permitted to move, followed by minimization, yielded a binding site with several changes in hydroxyl hydrogen dihedral angles (see caption to Fig. 1). Minima for methyl ammonium and acetate were determined in the altered structure for comparison with the X-PLOR structure results. The minima for acetate showed many small changes in position (< 1Å) due to the altered hydrogen coordinates leaving the gross distribution unchanged. For methyl ammonium (Fig. 1d), which only donates hydrogen bonds, the minima show behavior similar to that of acetate. In addition, several minimum positions are eliminated by the altered

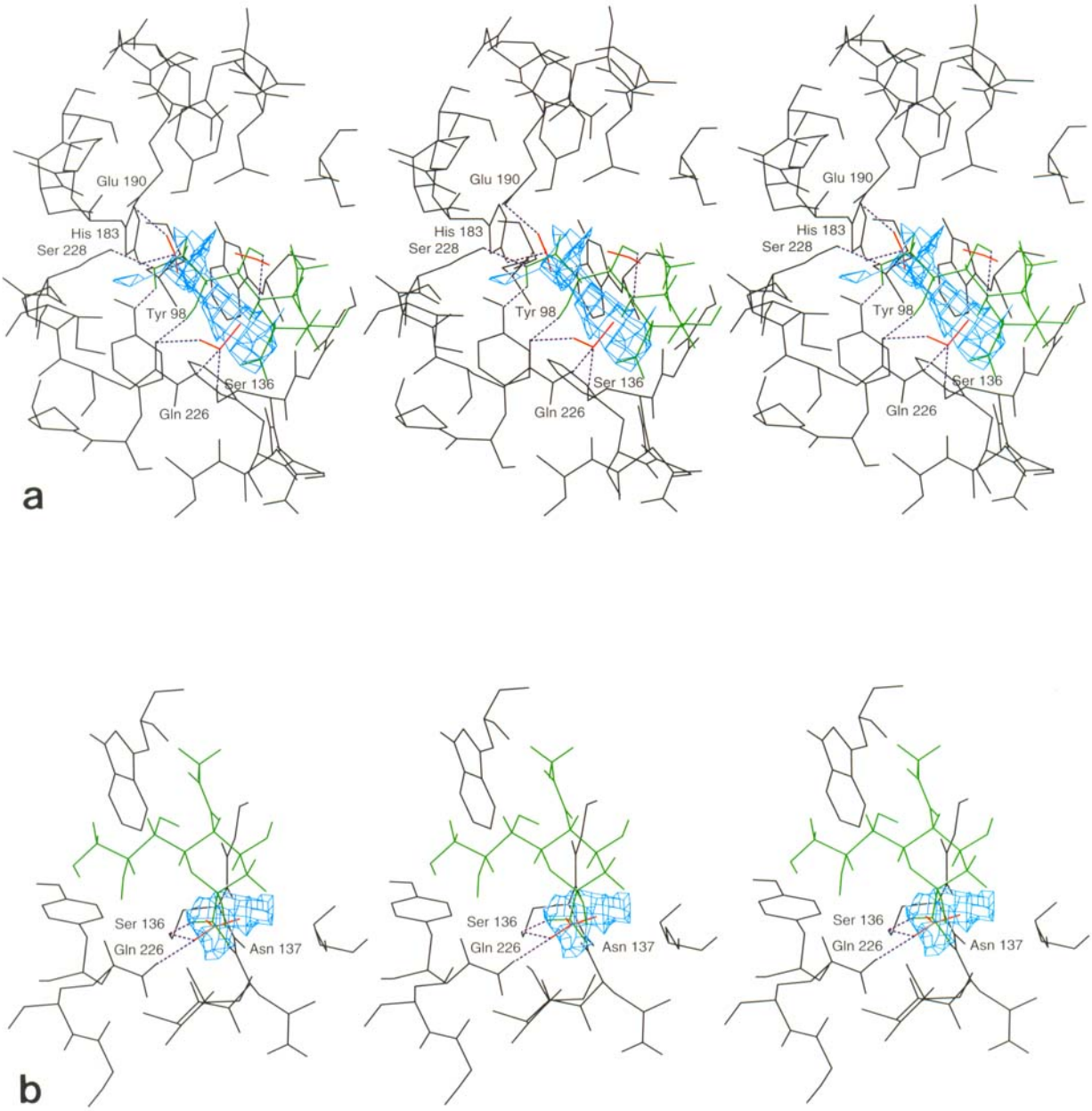

Fig. 3. Comparison of minima and grid search to corresponding groups on sialic acid. Stereo triple projections are shown. Protein is in black, minima in red and sialic acid in green. Relevant hydrogen bonds are shown in violet. **(a):** Methanol in the vicinity of the glycerol side chain. Contour level is E = −10.25 kcal/mol. **(b):** Correlation between acetate TDH minimum and grid search in the neighborhood of sialic acid's carboxylate. Contour level is E = −46.5 kcal/mol.

hydrogens. This includes a minimum position in contact with the hydroxyl oxygen of Tyr 98. These results suggest that modeling ligands on the basis of a single protein conformer may not always be appropriate.[15]

A comparison of the functional group maps was made with the binding interactions found in the sialic acid-hemagglutinin co-crystal structure.[16] Sialic acid can be broken down into functionalities consisting of methanol, acetate, methane, dimethyl ether, and N-methyl acetamide. The methanol functionality map yielded numerous minima in the vicinity of the glycerol side chain of the sialic acid. Since a one-to-one correspondence was not found with the hydroxyl groups on this side chain, a 0.25Å grid search was performed. Contouring at E = −10.2 kcal/mol (Fig. 3a) yields a region that is in the vicinity but does not enclose the glycerol side chain of sialic acid. If instead a contour at E = −5.1 kcal/mol (the standard methanol cutoff) is used, a region

containing virtually all of the glycerol side chain is found. This indicates that there is a broad minimum that allows considerable motional freedom for the side chain after binding. For acetate, there is a minimum having a very large interaction energy (E = −61.6 kcal/mol) with the carboxylate carbon within 0.3Å of the sugar carboxylate (Fig. 3b). A grid search in this area, contoured at E = −46.5 kcal/mol, outlines an energy well that encompasses both the minimum and the crystallographically placed carboxyl group. The change of the oxygen positions in the TDH minimum relative to the sialic acid carboxyl group leads to an additional hydrogen bond to Gln 226, as well as improved geometry for the bifurcated hydrogen bonds to the main chain amide of Asn 137. This orientation for a carboxylate is not achieved by sialic acid since it would require the sugar ring to move by more than 1Å. A modified ligand could be designed by replacing the existing axial sialic acid carboxylate with one linked equatorially.

The lack of complete correspondence of the ligand's functional group positions with the MCSS functionality maps may reflect the inability of sialic acid to achieve optimum positions for its functional groups. This possibility is in accord with the small binding free energy of sialic acid (−3.7 kcal/mol),[17] relative to the value predicted (−11.3 kcal/mol) from its functional groups by the analysis of binding data for a wide variety of ligands.[8] A ligand with tighter binding could be constructed by modifying the sialic acid ligand to take advantage of the functionality positions orientations found by minimization or dynamics, which includes the motion of protein sidechains. Alternatively, a molecule could be designed *de novo* from the functional group information, combined perhaps with results from structural data bases.[18]

ACKNOWLEDGMENTS

We thank Dr. J. Hanson and Dr. D. Wiley for valuable discussions, Dr. W. Weis and Dr. D. Wiley for crystallographic coordinates of hemagglutinin, and Dr. P. Bash and Dr. M. Field for help in the initial stages of the work. The research was supported in part by the National Institute of Health.

REFERENCES

1. Jolles, G., Wooldridge, K.R.H., eds. "Drug Design: Fact or Fantasy?" London: Academic Press, 1984.
2. Elber, R., Karplus, M. Enhanced Sampling in molecular dynamics: Use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. J. Am. Chem. Soc. 112:9161–9175, 1990.
3. Burgen, A.S.U., Roberts, G.C.K., Tute, M.S., eds. "Molecular Graphics and Drug Design." Amsterdam: Elsevier, 1986.
4. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28:849–857, 1985.
5. Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M., Wade, R.C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. J. Med. Chem. 32:1083–1094, 1989.
6. Patai, S., "Patai's Guide to the Chemistry of Functional Groups." New York: John Wiley & Sons, 1989.
7. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 4:187–217, 1983.
8. Andrews, P.R., Craik, D.J., Martin, J.L. Functional group contributions to drug-receptor interactions. J. Med. Chem. 27:1648–1657, 1984.
9. Goodsell, D.S., Olson, A.J. Automated docking of substrates to proteins by simulated annealing. Proteins 8: 195–202, 1990.
10. Tidor, B., Karplus, M. Simulation analysis of the stability mutant R96H of T4 lysozyme. Biochemistry (in press) 1990.
11. Mackerell, A.D. Jr., Karplus, M. Unpublished calculations.
12. Wilson, I.A., Skehel, J.J., Wiley, D.C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature (London) 289:366–373, 1981.
13. Brünger, A.T., Kuriyan, J., Karplus, M. Crystallographic R factor refinement by molecular dynamics. Science 235: 458–460, 1987.
14. Weis, W.I., Brünger, A.T., Skehel, J.J., Wiley, D.C. Refinement of the influenza virus hemagglutinin by simulated annealing. J. Mol. Biol. 212:737–761, 1990.
15. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. J. Med. Chem. 31:722–729, 1988.
16. Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., Wiley, D.C. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature (London) 333:426–431, 1988.
17. Sauter, N.K., Bednarski, M.D., Wurzburg, B.A., Hanson, J.E., Whitesides, G.M., Skehel, J.J., Wiley, D.C. Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: A 500-MHz proton nuclear magnetic resonance study. Biochemistry 28:8388–8396, 1989.
18. Bartlett, P.A., Shea, G.T., Telfer, S.J., Waterman, S. Chapter title. In: "Molecular Recognition: Chemical and Biological Problems," Roberts, S.M., ed. Cambridge: Royal Society of Chemistry, 1989.
19. Ben-Naim, A., Marcus, Y. Solvation thermodynamics of nonionic solutes. J. Chem. Phys. 81:2016–2027, 1984.
20. Gao, J. Theoretical studies of condensed phase chemistry. PhD thesis, Purdue University, 1987.