Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability To Form More Than Two Hydrogen Bonds

Rebecca C. Wade† and Peter J. Goodford*

The Laboratory of Molecular Biophysics, The Rex Richards Building, University of Oxford, Oxford OX1 3QU, England

Received November 27, 1991

The specificity of interactions between biological macromolecules and their ligands may be partially attributed to the directional properties of hydrogen bonds. We have now extended the GRID method (Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849. Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J. Med. Chem.* **1989**, *32*, 1083). of determining energetically favorable ligand binding sites on molecules of known structure, in order to improve the treatment of groups which can make multiple hydrogen bonds. In this method, the interaction energy between a probe (a small chemical group that may be part of a larger ligand) and a target molecule is calculated using an energy function which includes a hydrogen bond term which is dependent on the length of the hydrogen bond, its orientation at the hydrogen-bonding atoms, and their chemical character. The methods described in the preceding paper (Wade, R. C.; Clark, K. J.; Goodford, P. J. *J. Med. Chem.*, preceding paper in this issue) for probes capable of making two hydrogen bonds are here extended to the following probes which have the ability to make more than two hydrogen bonds: ammonium $-NH_3^+$, amine $-NH_2$, sp^3-hybridized hydroxyl, and water. Use of the improved GRID procedure is demonstrated by the determination of the conformation of an amino acid side chain at the subunit interface in hemoglobin and of the location of water binding sites in human lysozyme.

Introduction

In designing a therapeutic agent based on the three-dimensional structure of a target receptor molecule, it is necessary to be able to predict where it will bind to the target and with what specificity. Hydrogen bonds are a major determinant of the specificity of binding to biological macromolecules.[1,2] This is because of their short-range nature; dependence on chemical atom type; and directional properties (see refs 3 and 4 for more detailed discussions). We have now extended the treatment of hydrogen bonds in the GRID method[3-6] of determining energetically favorable ligand binding sites on macromolecules. The GRID method uses an empirical energy function that includes a hydrogen bond term (E_{hb}) as well as Lennard-Jones (E_{lj}) and electrostatic (E_{el}) terms:

$$E = \sum E_{lj} + \sum E_{el} + \sum E_{hb} \qquad (1)$$

in order to calculate the interaction energy between a probe, representing a functional group, and a target molecule. The atoms of the target molecule are held fixed while the calculations are performed, although movements of hydrogen atoms and lone pairs of electrons are taken into account. Energy maps can thus be generated for the interaction of each probe with the appropriate conformation or conformations of the target molecule. The hydrogen bond term, E_{hb}, is given by

$$E_{hb} = Er \times Et \times Ep \quad \text{for } 0 < Et < 1 \text{ and } 0 < Ep < 1 \qquad (2)$$

Here, Er is a function of the separation r of the hydrogen-bonding atoms and is given in kilocalories/mole, and Et and Ep are dimensionless functions of the angles t and p made by the hydrogen bond at the target and probe atoms respectively (see Figure 1).

The E_{lj} and E_{el} terms and the Er and Et components of E_{hb} have been described in previous papers.[3,5] The Ep component for probes that are capable of making two hydrogen bonds is described in the preceding paper.[4] Here, we discuss the Ep component for probes that are capable of making three or four hydrogen bonds.

At each probe position at which the interaction energy is calculated, the orientation of the probe that results in the most energetically favorable interaction must be found. This can be done by assuming, as for probes with the ability to make only two hydrogen bonds, that there is an asymmetric arrangement of hydrogen bonds around the probe, with the strongest hydrogen bond formed being linear and the weaker hydrogen bonds showing some nonlinearity. The energy of each of these weaker hydrogen bonds is modified by a function Ep which models the experimentally observed geometries of hydrogen bonds to such probes. Due to the increased number of permutations for these probes compared to probes capable of making only two hydrogen bonds, the computation and selection of the hydrogen bonds formed is more complicated. Some of these probes also have more than one ideal,

† Present address: European Molecular Biology Laboratory, Meyerhofstr. 1, 6900 Heidelberg, Germany.

(1) Pauling, L. *The Nature of the Chemical Bond*, Cornell University Press: New York, 3rd ed.; 1960, Chapter 12.
(2) Fersht, A. R. The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **1987**, *12*, 301-304.
(3) 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.* **1989**, *32*, 1083-1094.
(4) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favourable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups with the Ability To Form Two Hydrogen Bonds. *J. Med. Chem.*, previous paper in this issue.
(5) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.
(6) Inquiries regarding the availability of program GRID should be directed to the authors.

0022-2623/93/1836-0148$04.00/0

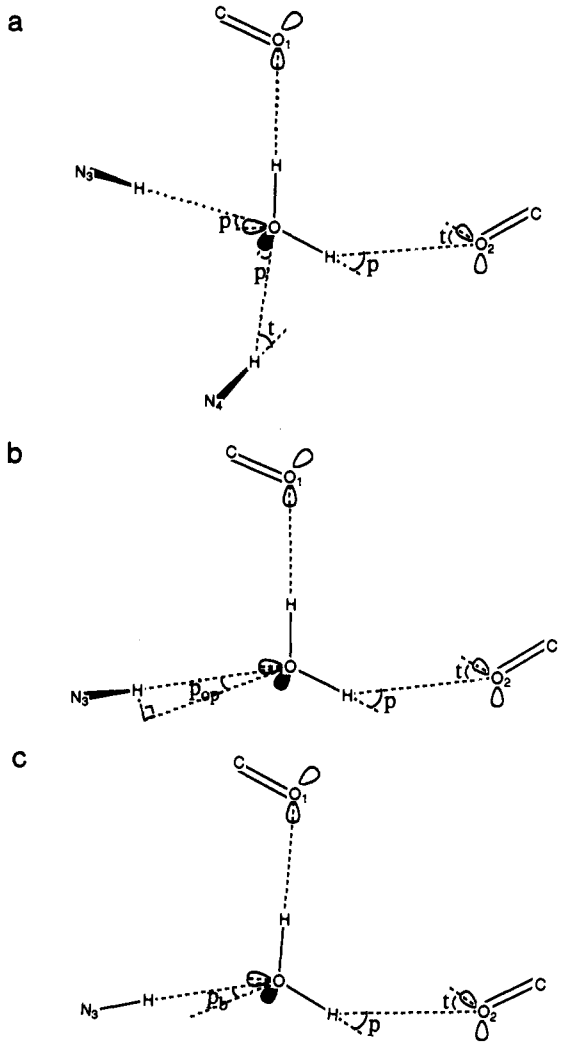

Figure 1. Diagrams showing different hydrogen bond geometries for a water probe interacting with different model targets to which it can donate two hydrogen bonds and accept one (b and c) or two (a) hydrogen bonds. (a) The probe accepts two hydrogen bonds and has a tetrahedral hydrogen bond coordination geometry. (b) The probe accepts one hydrogen bond and has a tetrahedral hydrogen bond coordination geometry. In the illustrated case, the two donated hydrogen bonds are stronger than the accepted hydrogen bond so the plane of the lone pair orbitals can be deduced and *Ep* can be determined from angle p_{op}. See text. (c) The probe accepts one hydrogen bond and has a planar triangular hydrogen bond geometry. In the illustrated case, the linear donated hydrogen bond for which $p = 0°$ is the strongest hydrogen bond. The location of the bisector of the probe's lone pairs of electrons can be deduced and *Ep* can be determined from angle p_b. See text. The angles made by the hydrogen bonds at the probe and the target atoms are defined as follows: *t* is the angle by which a hydrogen bond deviates from its optimum geometry at a target atom, i.e. from alignment along the direction of one of the target atom's hydrogen atoms or lone-pair orbitals or the bisector of the lone-pair orbitals. *p* is the angle by which a hydrogen bond deviates from its optimum geometry at the probe, i.e. from alignment along the direction of one of the probe's hydrogen atoms or lone-pair orbitals. p_{op} is the angle subtended at the probe by the hydrogen bond that it accepts and the plane of its lone-pair orbitals when only one hydrogen bond is accepted. p_b is the angle between the hydrogen bond accepted by the probe and the bisector of the probe's lone-pair orbitals when only one hydrogen bond is accepted.

optimum hydrogen bond coordination geometry, and different functional forms of *Ep* may be appropriate for each of these. In order to determine the most favorable hydrogen bond energy, the actual arrangement of the probe and the surrounding target atoms must be compared to all of these ideal hydrogen bond coordination geometries.

The treatment of ammonium $-NH_3^+$, amine $-NH_2$, sp^3-hybridized hydroxyl, and water probes is described in this paper. Examples of the calculation of binding sites in proteins for probes which are capable of making three or four hydrogen bonds are given. The ability to predict water binding sites is demonstrated, indicating that the GRID program may be of assistance in the assignment of water positions in biological macromolecules whose structures have been solved by X-ray crystallography or NMR.

Method of Determining the Hydrogen Bonds Made by the Probe

The Derivation of Function *Ep*. Function *Ep* for probes which are able to make more than two hydrogen bonds was derived from experimental data in the same way as for probes which can only make two hydrogen bonds.[4] For an sp^3-hybridized amine ($-NH_2$) or ammonium ($-NH_3^+$) probe, the same expression for *Ep* was found to be suitable:

$$Ep = \cos^2 p \quad 0° < p < 90°$$
$$= 0 \quad p > 90° \qquad (3)$$

As the orientation of the hydrogen bonds to these probes is primarily governed by the positions of their hydrogen atoms, their optimum hydrogen bond coordination geometry is tetrahedral.[7-9] Deviations of the actual hydrogen bond arrangement from tetrahedral geometry result in the modification of the energy of the hydrogen bonds by the factor *Ep*.

For an sp^3-hybridized hydroxyl group, a tetrahedral hydrogen bond coordination geometry is favored for all combinations of hydrogen bonds except that, when one hydrogen bond is donated and only one hydrogen bond is accepted, the accepted hydrogen bond may lie near to the bisector of the hydroxyl group's lone pairs of electrons so that the angle subtended by the two hydrogen bonds at the hydroxyl group is close to 125°. In the latter case, $Ep = \cos^2 p_b$ is calculated for the weaker hydrogen bond with p_b corresponding to the angle of deviation from the ideal sp^2 hydrogen bond coordination geometry. For a hydroxyl probe with a tetrahedral hydrogen bond coordination geometry, $Ep = \cos^2 p$ is calculated for the weaker hydrogen bonds with *p* given by the angle of deviation from the tetrahedral arrangement.

The experimentally observed hydrogen bond geometry of the sp^3-hybridized amine ($-NH_2$), ammonium ($-NH_3^+$), and hydroxyl groups was considered in detail during the formulation of *Et* in order to model the dependence of hydrogen bonds on their orientation at the target atoms as described in ref 3. However, the detailed geometry of water molecules was not dealt with at that time because the water molecules of the target were then modeled as isotropic neutral spheres capable of making four hydrogen bonds in any direction. The hydrogen bond coordination geometry of water molecules is now taken into account when they are considered as probes, and therefore, the experimentally observed hydrogen bond coordination

(7) Ramakrishnan, C.; Prasad, N. Study of Hydrogen Bonds in Amino Acids and Peptides. *Int. J. Protein Res.* 1971, *III*, 209–229.
(8) Taylor, R.; Kennard, O. Hydrogen Bond Geometry in Organic Crystals. *Acc. Chem. Res.* 1984, *17*, 320–326.
(9) Taylor, R.; Kennard, O.; Versichel, W. The Geometry of the N—H···O=C Hydrogen bond. 3. Hydrogen Bond Distances and Angles. *Acta Crystallogr.* 1984, *B40*, 280–288.

geometries of water molecules are discussed here together with the derivation of a model for a water probe.

The Model of the Water Probe and Its Hydrogen Bond Coordination Geometry. A water molecule is able to donate two hydrogen bonds and accept two hydrogen bonds. Experiments show that, on hydrogen bonding, the internal H–O–H bond angle tends to widen from its water vapor value of 104.5° by an average of about 2.5°.[10] In crystal structures, the water molecule has been observed to have both tetrahedral and triangular hydrogen bond coordination geometries with angle H–O–H = 106.9 ± 0.6° and 109.0 ± 0.5°, respectively.[11] In the tetrahedral geometry, one or two hydrogen bonds may be accepted and these usually lie close to the directions of the lone-pair orbitals. In the triangular geometry, only one hydrogen bond is accepted and this lies close to the bisector of the lone-pair orbitals. For the purpose of determining the orientation of a water probe which optimizes the arrangement of its hydrogen bonds and maximizes the attractive interactions with the target, a water probe is modeled in program GRID under the assumption that the optimum angle subtended by two donated hydrogen bonds is 109° irrespective of whether its hydrogen bond coordination geometry is tetrahedral or triangular.

Hydrogen Bonds Donated by the Water Probe. Water appears to donate hydrogen bonds more often than it accepts them. In small molecule crystal hydrates, water molecules nearly always donate two hydrogen bonds.[10,12-14] In proteins, bound water molecules are more frequently observed donating hydrogen bonds than accepting them.[15,16] This is, inter alia, because there are more oxygens than nitrogens in proteins and these can generally accept more hydrogen bonds with fewer geometrical constraints than nitrogens. In the peptide bond, the carbonyl oxygen can stick out further into the solvent than the amide nitrogen and has a greater hydrogen-bonding capacity. It has been proposed that water may also have an innate tendency to donate hydrogen bonds.[15] This may arise because the hydrogen-bonding capacity of water itself can be satisfied by the donation of two hydrogen bonds and the acceptance of only one hydrogen bond in the direction of the bisector of its lone-pair electrons.

Linear hydrogen bonds donated along the direction of the water O–H bond are favored. Bending of these hydrogen bonds is generally isotropic, tending to follow a Gaussian distribution for angle $p < 20°$.[10,13] In program GRID, the same Ep function is used to calculate the energy of hydrogen bonds donated by the water probe as for other probes ($Ep = \cos^2 p$).

Hydrogen Bonds Accepted by the Water Probe. Water molecules show a tendency to accept hydrogen bonds in which the donor hydrogen is close to the plane of the water molecule's lone-pair orbitals. The angle subtended at the water oxygen atom by the donor hydrogen and this plane, p_{op} (see Figure 1), is generally less than 20°.[10,17] The hydrogen bond distribution shows less angular dependence between the lone-pair orbitals than outside them. The orientation toward the lone-pair directions is more strictly observed if two hydrogen bonds are accepted than if only one is.[10,17]

In small molecule crystal hydrates, water has been classified into the following categories according to the characteristics of its accepted hydrogen bonds:[10,12,18] (1) Tetrahedral hydrogen bond geometry with two hydrogen bonds accepted roughly along the directions of the two lone-pair orbitals. (2) Tetrahedral hydrogen bond geometry with one hydrogen bond accepted roughly along the direction of one lone-pair orbital. (3) Triangular hydrogen bond geometry with one hydrogen bond accepted roughly along the bisector of the lone-pair orbitals. (4) Bipyramidal hydrogen bond geometry with three hydrogen bonds accepted. (5) No hydrogen bonds accepted.

The bipyramidal geometry is uncommon[10,11] and has not been modeled in program GRID. A triangular hydrogen bond geometry is probably less common than a tetrahedral hydrogen bond geometry in biological systems, but it occurs with sufficient frequency to be taken into account by program GRID. The triangular coordination geometry has been observed most frequently in small molecule crystal hydrates. For example, in a study[10] of small molecule crystal hydrates observed by neutron diffraction, 31 water molecules were identified which were coordinated by hydrogen-bonding atoms and had no coordinating cations. Of these, 8 had triangular geometry, 10 had tetrahedral geometry with one accepted hydrogen bond, and 13 accepted two hydrogen bonds. None of the water molecules observed in this study had a bipyramidal hydrogen bond geometry.

In program GRID, the water probe is modeled as a sphere with a hydrogen bond coordination geometry which is dependent on the hydrogen-bonding capacity of the target molecule and may be classified as either tetrahedral or triangular. For a water probe which accepts two hydrogen bonds with a tetrahedral hydrogen bond geometry, the energy of the hydrogen bonds is calculated with $Ep = \cos^2 p$ (see Figure 1a). For a water probe which accepts only one hydrogen bond with a tetrahedral hydrogen bond geometry, the expression for Ep is dependent on the relative strengths of the donated and accepted hydrogen bonds. If the plane of the lone-pair orbitals can be defined (i.e. the two strongest hydrogen bonds made by the water probe are donated), $Ep = \cos^2 p_{op}$ for an accepted hydrogen bond which lies within the lone-pair orbitals, where angle p_{op} is the angle subtended at the probe by the donor hydrogen atom and the plane of the probe's lone pair orbitals (see Figure 1b). Thus, the hydrogen bond is constrained to the plane of the lone-pair orbitals but has a uniform probability of occurrence in that plane between the lone-pair orbitals. If the accepted hydrogen bond lies outside the lone-pair orbitals, then again $Ep = \cos^2 p$. If the plane of the lone-pair orbitals cannot be defined (i.e.

(10) Chiari, G.; Ferraris, G. The Water Molecule in Crystalline Hydrates Studied by Neutron Diffraction. *Acta Crystallogr.* 1982, *B38*, 2331–2341.
(11) Falk, M.; Knop, O. *Water: A Comprehensive Treatise*; Franks, F., Ed.; Plenum Press: New York, 1973; 2, p 55.
(12) Ferraris, G.; Franchini-Angela, M. Survey of the Geometry and Environment of Water Molecules in Crystalline Hydrates Studied by Neutron Diffraction. *Acta Crystallogr.* 1972, *B28*, 3572–3583.
(13) Pedersen, B. The Geometry of Hydrogen Bonds from Donor Water Molecules. *Acta Crystallogr.* 1974, *B30*, 289–291.
(14) Kroon, J.; Kanters, J. A.; van Duijnevelt-Van de Rijdt, J. G. C. M.; van Duijneveldt, F. B.; Vliegenthart, J. A. O–H···O hydrogen bonds in molecular crystals. A statistical and quantum-chemical analysis. *J. Mol. Struct.* 1975, *24*, 109–129.
(15) Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Mol. Biol.* 1984, *44*, 97–179.
(16) Blake, C. C. F.; Pulford, W. C. A.; Artymiuk, P. J. X-ray Studies of Water in Crystals of Lysozyme. *J. Mol. Biol.* 1983, *167*, 693–723.
(17) Mitra, J.; Ramakrishnan, C. Analysis of O–H···O Hydrogen Bonds. *Int. J. Pept. Protein Res.* 1977, *9*, 27–48.
(18) Chidambaram, R.; Sequira, A.; Sikka, S. K. Neutron diffraction studies of the structure of potassium oxalate monohydrate: lone-pair coordination of the hydrogen-bonded water molecule in crystals. *J. Chem. Phys.* 1964, *41*, 3616–3622.

one of the two strongest hydrogen bonds made by the probe is donated and the other is accepted), then $Ep = \cos^2 p$ is used. If the hydrogen bond geometry is triangular, then $Ep = \cos^2 p_b$ for an accepted hydrogen bond where p_b is the angle between the hydrogen bond and the bisector of the probe's lone-pair orbitals (see Figure 1c).

The Treatment of Multiple Hydrogen Bonds. The hydrogen bond energy of probes which are able to form three or four hydrogen bond is calculated in a similar manner to that for probes which are able to form two hydrogen bonds.[4] An asymmetric configuration is assumed with one hydrogen bond linear, (along the O–H or N–H bond, the lone-pair direction or the bisector of the lone pairs), and of greater energy than all the others. The combination of hydrogen bonds to the probe that is most energetically favorable is determined.

For sp^3-hybridized hydroxyl and water probes, both triangular and tetrahedral hydrogen bond coordination geometries are considered, and that closest to the actual hydrogen bond arrangement is assumed to be the reference for calculating angle Ep.

If two hydrogen bonds are accepted, the optimum hydrogen bond coordination geometry is always assumed to be tetrahedral, and therefore the energies of all but the strongest hydrogen bond, whether accepted or donated, are computed with $Ep = \cos^2 p$ where angle p is defined in Figure 1a.

If only one hydrogen bond is accepted, the determination of the hydrogen bond energy is dependent on the relative strengths of the hydrogen bonds formed. If three hydrogen bonds are formed and the two donated hydrogen bonds are stronger than the accepted hydrogen bond, then $Ep = \cos^2 p$ is used to calculate the energy of the weaker donated hydrogen bond. Then, the coordination geometry is determined from the deviation of the accepted hydrogen bond from the bisector of the lone pair orbitals given by angle p_b. If $p_b > 55°$, the hydrogen bond lies outside the lone-pair orbitals and the water probe is assumed to be tetrahedrally coordinated and $Ep = \cos^2 p$ for the accepted hydrogen bond. If $p_b < 55°$, the accepted hydrogen bond lies within the lone-pair orbitals and the water probe may be triangularly or tetrahedrally coordinated. Therefore, $Ep = \cos^2 p_b$ and $Ep = \cos^2 p_{op}$ are calculated for the accepted hydrogen bond and the factor resulting in the most favorable hydrogen bond energy is taken (see Figure 1b and 1c).

If three hydrogen bonds are formed (one accepted, two donated) and one of the donated hydrogen bonds is weakest, angle p_{ad}, the angle subtended at the probe by the accepted hydrogen bond and the strongest donated hydrogen bond is calculated first. This angle is used in order to determine which optimum geometry (triangular or tetrahedral), the actual hydrogen bond arrangement resembles most closely. If the closest optimum geometry is tetrahedral ($p_{ad} < 118°$), then the energy of the weaker hydrogen bonds is computed using $Ep = \cos^2 p$. If the closest optimum geometry is triangular ($p_{ad} > 128°$), the calculation of Ep is dependent on whether the accepted hydrogen bond is strongest or not. If it is, it is assumed to lie on the bisector of the probe's lone-pair orbitals and the energies of the donated hydrogen bonds are reduced by the $Ep = \cos^2 p$ term. In this case, the probe's hydrogen atoms are assumed to lie on a circular locus around the bisector of the lone-pair orbitals. On the other hand, if the second strongest hydrogen bond is accepted, the energy of this hydrogen bond is computed using an $Ep = \cos^2 p_b$ term.

If only two hydrogen bonds are made, one accepted and one donated, then Ep is determined from the value of p_{ad}, the angle they subtend at the probe. If $p_{ad} > 128°$, the spatial distribution of the hydrogen bonds around the probe is closer to a triangular arrangement. Therefore, if the weaker hydrogen bond is accepted, $Ep = \cos^2 p_b$ is calculated for it; if it is donated, $Ep = \cos^2 p$ is calculated for it, assuming that the stronger accepted hydrogen bond is aligned along the bisector of the probe's lone-pair orbitals. If $p_{ad} < 110°$, the spatial distribution of the hydrogen bonds is closest to a tetrahedral hydrogen bond coordination geometry, and $Ep = \cos^2 p$ is calculated for the weaker hydrogen bond. If $110° < p_{ad} < 128°$, a single preferred hydrogen bond geometry cannot be assigned, and therefore, it is assumed that $Ep = 1.0$ for the weaker hydrogen bond. If only two hydrogen bonds are made and they are both donated, the hydrogen bond coordination geometry is not defined and the energy of the weaker hydrogen bond is calculated with $Ep = \cos^2 p$.

In situations where the closest optimum geometry cannot be unambiguously defined, Ep and, hence, the hydrogen bond energy tend to be overestimated so that possible binding sites will not be missed by program GRID.

Examples of the Application of the Hydrogen Bond Functions

Hemoglobin Subunit Interactions. An important difference between deoxyhemoglobin (deoxyHb) and oxyhemoglobin (oxyHb) is the presence of eight salt bridges between the subunits in deoxyHb that are broken when the subunits rearrange to form oxyHb.[19] These salt bridges tend to stabilize deoxyHb relative to oxyHb. Two of them occur between the $-NH_3^+$ group of Lys 127 of one α subunit and the carboxyl group of the C-terminal residue 141 of the other α subunit. Here, we describe the use of program GRID to interpret the arrangement of the side chain of Lys 127 at the interface between the α subunits.

The structure of human deoxyHb has been solved at 1.74-Å resolution with an R factor of 16%.[19] Program GRID was run for an $-NH_3^+$ probe (able to donate three hydrogen bonds) in the vicinity of residue 127 of one of the α subunits (labeled C) in the structure of human deoxyHb (using Brookhaven Protein Databank[20] file 2HHB) after all the side chain atoms of Lys 127C had been removed. The net atomic charge on deoxyHb was calculated at pH 7 by program GRID as +1. The probe had a charge of +0.66 e corresponding to that for a lysine side chain $-NH_3^+$ group in the GRID force field, and therefore, its binding sites were influenced by electrostatic as well as hydrogen bond and van der Waals interactions. Energy maps, which are shown in Figure 2, were calculated both with and without computing factor Ep to account for the distribution of the hydrogen bonds at the probe.

The local energy minimum in both maps is within the large contoured region about 1.9 Å from the position of the side chain nitrogen (NZ) of Lys 127C. It is clear that

(19) Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. The Crystal Structure of Human Deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **1984**, *175*, 159–174.

(20) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Bryce, M. D.; Rogers, J. R.; Kennard, O.; Shikanouchi, T.; Tasumi, M. The Protein Data Base: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.

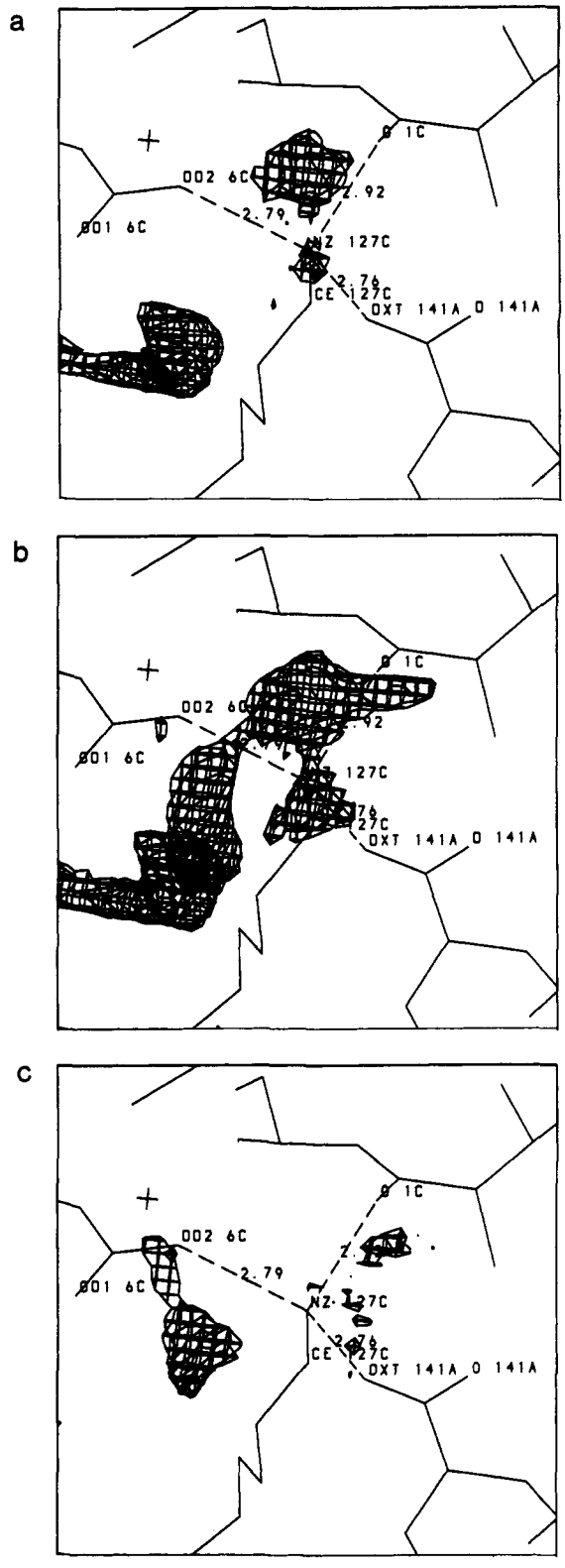

Figure 2. Part of the interface between the two α subunits of deoxyhemoglobin surrounding Lys 127C. (a and b) Grid energy contours (at −9.5 kcal/mol) indicating favorable binding sites for an amine $-NH_3^+$ probe are shown (a) with *Ep* calculated in order to account for the spatial distribution of the hydrogen bonds at the probe and (b) with *Ep* assigned an arbitrary value of 1.0. These energy maps were calculated after the side chain atoms of Lys 127C had been removed from the protein target. (c) Energy map showing differences between the maps in (a) and (b). The contours show the regions where the probe–target interaction is at least 2.5 kcal/mol less favorable when *Ep* is computed correctly as in (a) than when it is not as in (b). See text.

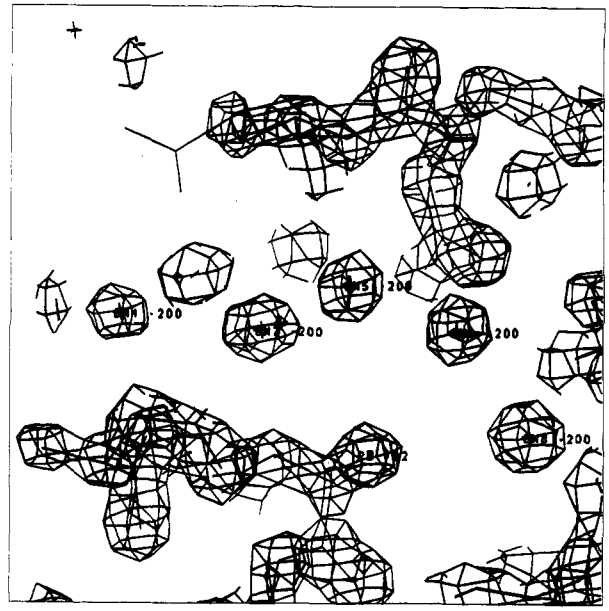

Figure 3. Part of the electron density map of human lysozyme with contours at 0.8 e Å^{-3} showing the four internal waters (200/5–8) and water 200/4 near the protein surface. The internal waters are grouped around Ala 92 and water 200/8 may stabilize a reverse turn in the protein.

this position is physically unrealistic because it could only be occupied by a probe oriented in the opposite direction to NZ Lys 127, i.e. with the covalently attached carbon (CE) positioned further into solution than the nitrogen (NZ).

In both maps, however, there is a second energy minimum closer to the observed position and near the CE–NZ bond. At this energy minimum, the probe, like the observed NZ Lys 127C, would make three hydrogen bonds, to the carboxylate oxygen OD2 of Asp 6C and the carbonyl oxygen O of Val 1C of the same α subunit (C) and to the terminal carboxylate oxygen OXT of Arg 141A of the other α subunit (A). The probe at the energy minimum would have a more tetrahedral hydrogen bond coordination geometry than at the experimentally observed position of NZ Lys 127C where the vectors from the nitrogen to the hydrogen bond accepting atoms are approximately perpendicular to the CE–NZ bond (angle CE–NZ–O $\approx$ 90°). We do not know if the nonideal geometry of the observed position of NZ Lys 127C is due to the restrictions imposed by the presence of the remainder of the lysine side chain linking it to the protein backbone, or to the thermal motions of the side chain which prevented accurate fitting of the nitrogen position.

The map in Figure 2c, which shows the energy differences between the maps in Figure 2a and 2b, indicates two major regions in which the use of the factor *Ep* clarifies the map. At one of these, the probe interacts with the carboxylate group of Asp 6C, and at the other, it interacts with the C-terminal carboxylate group of Arg 141A. In these regions, when *Ep* is not calculated, the probe can donate two hydrogen bonds to the carboxylate group, one to each oxygen. When *Ep* is calculated, this is no longer possible, and only one hydrogen bond is donated to the carboxylate group. This is more realistic because a second hydrogen bond to the carboxylate group would deviate considerably from linearity and would, therefore, be very weak.

This example shows that the GRID method may be used to examine the interactions between two macromolecules.

Table I. Prediction of the Internal Water Molecules in Human Lysozyme by Program GRID

predicted water molecule (PW no.)	no. of predicted waters included in the target	closest observed water molecule	distance from observed water molecule (Å)	binding energy at predicted position (kcal/mol)	hydrogen bond partners at predicted position		hydrogen bond partners at observed position	
1	0	200/7	0.12	−13.8	N	Ala 92	N	Ala 92
					O	Asn 88	200/4	
					O	Ala 83	O	Ala 83
							200/5	
2	1	200/5	0.46	−15.2	O	Gln 86	O	Gln 86
					PW1		200/7	
							200/6	
3	2	200/4	0.38	−13.2	N	Asp 91	N	Asp 91
					OD1	Asn 88	OD1	Asn 88
					PW1		200/7	
							212/1	
4	2	200/8	0.19	−11.3	N	Phe 57	N	Phe 57
					O	Tyr 54	O	Tyr 54
							200/6	
5	4	200/6	0.44	−12.8	O	Leu 84	O	Leu 84
					N	Ile 56	N	Ile 56
					PW2		200/5	
					PW4		200/8	

In addition, it illustrates how program GRID might be used to predict the effect of mutations in a protein and to model the mutated side chains. The procedure of removing an amino acid side chain from a target protein structure and then using program GRID to identify binding sites in the vicinity of the side chain for appropriate probes can be readily applied to other amino acids. In this instance, although the calculation of *Ep* for the $-NH_3^+$ probe had little effect on the positions of the energy minima close to the observed lysine $-NH_3^+$ group, it resulted in a more precise and more easily interpreted energy map.

Water Molecules in Crystals of Human Lysozyme (HL). The water structure of crystalline human lysozyme (HL) provides a model system for testing the GRID method. It has been observed by X-ray crystallography at 1.5 Å resolution with an *R* factor of 18.7%.[21] The original crystallographic data and electron density maps were made available by Artymiuk and Blake enabling the predicted water positions to be compared, not only with the coordinates assigned to the observed waters, but also with the electron density map from which their positions were determined. In addition, the water structure of crystals of HL has been analyzed in detail by Blake, Pulford, and Artymiuk[16] and found to contain features typical of globular proteins.

Internal Water Molecules. Program GRID was run with a water probe over the whole lysozyme molecule, which had a net charge of +8, in order to identify water binding sites. The largest region of favorable binding energy in the GRID energy map was in an internal cavity of HL which contains a chain of water molecules (200/4–8). These water molecules are clearly defined in Artymiuk and Blake's electron density map[16] (see Figure 3), but program GRID was used to sequentially position them in the cavity without using the information about their experimental coordinates. This was done by running GRID with a water probe over the whole internal cavity region of the target HL molecule. An energy minimum was found; its position was refined by repeating the GRID calculation in the vicinity of the energy minimum with a finer grid spacing; and a water molecule was then assigned to it (indicated by the square in Figure 4a). Program GRID was then run with a target consisting of HL and this one water. A second energy minimum was found and a second water molecule assigned to it (Figure 4b). This procedure was repeated (Figure 4c and 4d), adding predicted water molecules to the target until, after five water molecules had been added, no suitably deep energy minimum could be found, indicating that the cavity was fully solvated.

The properties of the predicted waters are listed in Table I. They were predicted at a mean distance of 0.3 ± 0.1 Å from the experimentally observed water sites. The distances between the predicted and experimentally observed water sites were all less than the mean displacement of 0.7–0.9 Å of the waters given by their experimental temperature factors.

The necessity of taking account of the geometry of the hydrogen bonds made by the water probe may be demonstrated by a separate prediction of the location of water 200/8. In this calculation, the target consisted of HL and all of the crystallographically observed water molecules except water 200/8. Program GRID was run for a water probe in the region around the observed position of water 200/8. An earlier version of program GRID in which the spatial distribution of the hydrogen bonds around the probe was not taken into account predicted hydrogen bonds to be made by water 200/8 to both the backbone carbonyl oxygen of Leu 84 and water 200/6 (see Figure 4). However, the carbonyl oxygen of residue 84 and water 200/6 subtend an angle of 37° at water 200/8 and thus the simultaneous formation of hydrogen bonds to these atoms would really be unfavorable. The new version of program GRID was able to take account of the angle subtended at water 200/8 by these atoms and so it predicted a hydrogen bond to be made to water 200/6 only. This resulted in an improved prediction for water 200/8 with the distance from the observed position being reduced to 0.36 Å. This improvement was due to the application of restrictions on the hydrogen bond coordination geometry of the probe which prevented the prediction of unfavorable arrangements of hydrogen bonds to the probe.

This prediction of the internal water molecules in HL illustrates how program GRID may be used to determine accurately well-ordered water sites in crevices of macro-

(21) Artymiuk, P. J.; Blake, C. C. F. Refinement of Human Lysozyme at 1.5Å Resolution. Analysis of Non-bonded and Hydrogen-bond Interactions. *J. Mol. Biol.* 1981, *152*, 737–762.

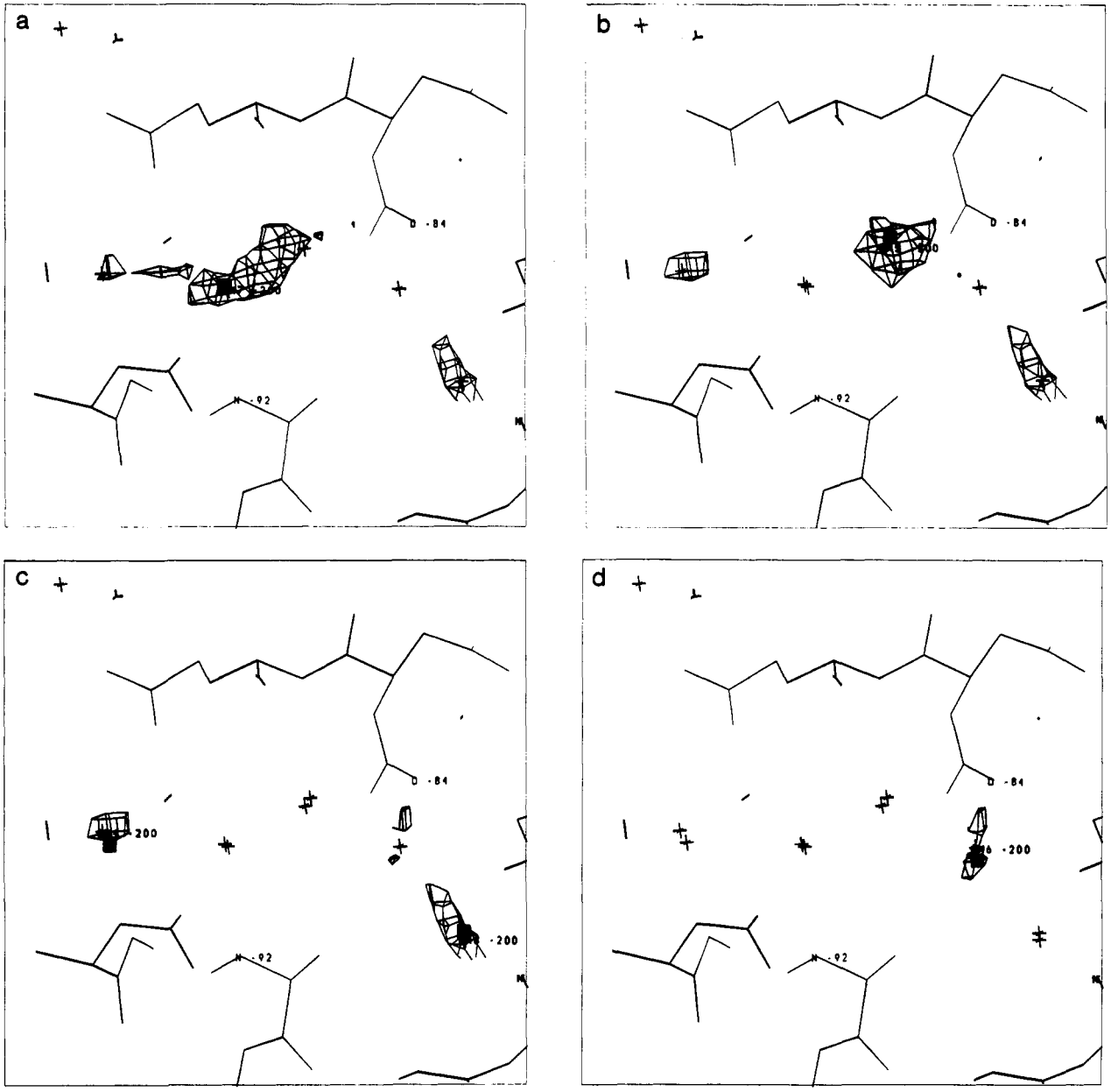

Figure 4. GRID energy maps (contoured at −9 kcal/mol) for a water probe showing the sequential prediction of the water molecules in the internal cavity in human lysozyme (HL) shown in Figure 3. The observed positions of the water molecules are marked by crosses and the predicted positions by squares. (a) The target consisted of HL only. The square marks the position of the energy minimum just 0.1 Å from the observed position of water 200/7. (b) The target consisted of HL and one water molecule at the energy minimum shown in (a). An energy minimum was found near water 200/5 and a water molecule was assigned to it. (c) The target consisted of HL and two predicted waters. Two energy minima could be identified in this map near the observed positions of waters 200/4 and 200/8 and water molecules were assigned to these. (d) The target consisted of HL and four predicted waters. An energy minimum to which a fifth water molecule was assigned was found near the observed position of 200/6. No further waters could be fitted into this cavity. See Table I and text.

molecules. Such water sites are often found at positions where ligands bind to macromolecules, and thus the ability of program GRID to predict water positions is of considerable importance.

Surface Water Molecules. Program GRID is able to determine accurately the position of well-ordered surface water molecules. For example, Figure 5 shows the electron density and GRID energy contours for water 204/8 in HL. Including other crystallographically observed water molecules and symmetry related protein molecules in the target, the local energy minimum was computed to be only 0.1 Å from this water molecule's experimentally observed position. If, on the other hand, the target consisted of one protein molecule alone without its crystallographic images or water molecules, a water binding site was calculated 0.4 Å from the observed position. This water molecule has a tetrahedral hydrogen bond coordination geometry and donates two hydrogen bonds to two carboxyl oxygens in the protein and accepts two hydrogen bonds from two adjacent water molecules (see Figure 5).

A water molecule (205/3) with a planar triangular hydrogen bond coordination geometry is shown in Figure 6. When program GRID was run with a water probe and a target of one HL molecule only, an energy minimum was found 0.4 Å from its observed position where the water probe could donate two strong hydrogen bonds to the backbone carbonyl oxygens of residues 113 and 117 and accept a third hydrogen bond from the guanadino nitrogen

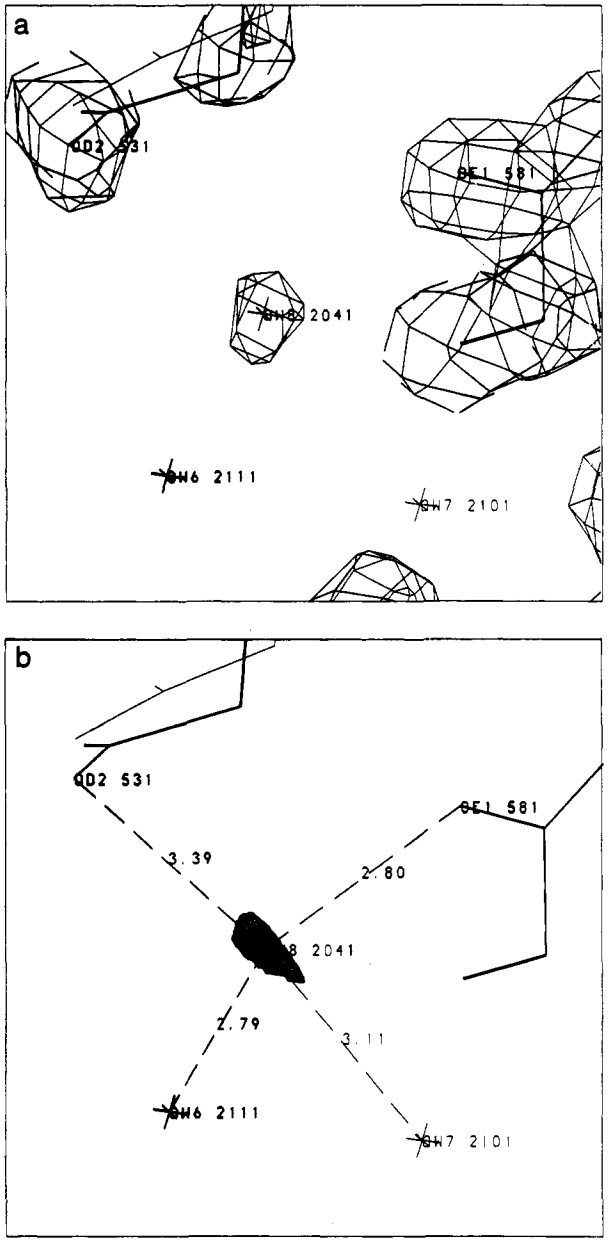

Figure 5. The region of crystalline human lysozyme surrounding the experimentally observed position of the well-ordered surface water molecule 204/8. (a) Part of the electron density map with contours at 0.8 e Å^{-3} showing approximately spherical contours defining the position of water 204/8. (b) Energy contours at −15.7 kcal/mol obtained by running program GRID with a water probe and an HL crystal target consisting of HL, adjacent symmetry related protein molecules, and all water molecules except water 204/8 itself. The energy minimum is 0.1 Å away from the experimentally observed position of water 204/8. When program GRID was run for a target consisting of one HL molecule only, an energy minimum was found 0.4 Å from the observed position of water 204/8. This water site has an approximately tetrahedral hydrogen bond coordination geometry. See text.

(NE) of Arg 119. The latter hydrogen bond lies approximately in the direction of the bisector of the water molecule's lone pairs of electrons close to the plane of the water molecule (containing the oxygen and the two hydrogens). Although the backbone carbonyl oxygen of residue 114 provides a fourth possible hydrogen bond partner for water 205/3 at a distance of 2.95 Å, the GRID

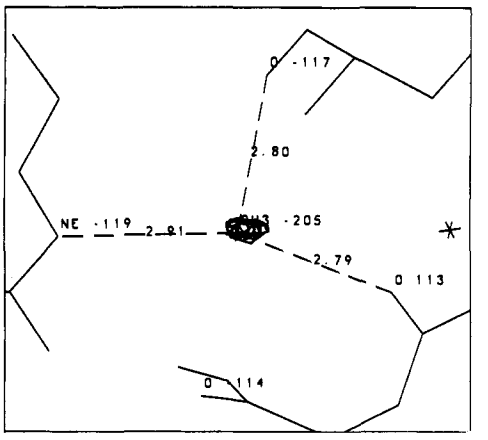

Figure 6. The region of crystalline human lysozyme surrounding the experimentally observed surface water molecule 205/3 with energy contours at −12 kcal/mol. When program GRID was run with a water probe and a target of one HL molecule only, an energy minimum was found 0.4 Å from the observed water site. Program GRID calculated that the three hydrogen bonds shown would be made. This water site has an approximately planar triangular hydrogen bond coordination geometry in which the hydrogen bond accepted by the probe lies approximately in the direction of the bisector of the water molecule's lone pair of electrons and close to the plane of the water molecule defined by the oxygen and hydrogen atoms. In this figure, the water molecule is shown approximately parallel to the plane of the paper. See text.

calculations show that it does not make an effective hydrogen bond. This is because water 205/3 cannot donate more than two hydrogen bonds and the most favorable hydrogen bond energy is obtained by the combination of hydrogen bonds with residues 113, 117, and 119, rather than any combination including a hydrogen bond to residue 114.

These examples demonstrate program GRID's ability to determine water binding sites of both tetrahedral and triangular hydrogen bond coordination geometry. Program GRID has been used previously[22] to determine the positions of all of the observed water molecules in crystalline HL. In that study, 64% (compared to 61% with the earlier version of GRID before improvement of the hydrogen bond term) of all of the experimentally assigned waters had energy minima at a distance from their observed positions that was smaller than the root mean squared displacement ($\langle U^2 \rangle^{1/2}$) from their mean position indicated by their temperature factors ($B = (8/3)\pi^2\langle U^2 \rangle$). Program GRID located energy minima within 1.8 Å of the observed water sites of all of the experimentally observed waters with temperature factors less than 56 Å^2 (corresponding to $\langle U^2 \rangle^{1/2} < 1.5$ Å). The more ordered water molecules, including internal waters, waters bound in clefts and crevices of the protein, and waters conserved in different lysozyme crystals, were particularly well determined by program GRID. Water molecules that were less tightly bound, such as those near very mobile protein side chains or in the second hydration shell, were less well determined, and it is not unlikely that inaccuracies in the experimental assignment of these waters may be comparable to errors in the GRID calculations. Improvements in the prediction of water sites with the new GRID energy function were most conspicuous for specific water sites which had previously been predicted with physically unrealistic hydrogen bond arrangements.

(22) Wade, R. C. Ligand–Macromolecule Interactions. D. Phil. Thesis, University of Oxford, 1988.

Recently, other methods of predicting water binding sites on proteins have been proposed. Vedani and Huhta[23] have developed a method based on hydrogen bond directionality. Pitt and Goodfellow[24] have followed a knowledge-based approach to determine solvent positions around polar groups in proteins. These two methods have not been applied to human lysozyme, and therefore, only an approximate comparison with the GRID method can be made. Nevertheless, the accuracy of the results appears to be similar. E.g. GRID predicts 48% of all of the crystallographic waters in human lysozyme within 1.0 Å of their observed positions and 64% within 1.5 Å. Vedani and Huhta predict an average (for five proteins) of 42% within 1.0 Å and 70% within 1.5 Å. Pitt and Goodfellow predict an average (for seven proteins) of 41% within 1.0 Å and 59% within 1.4 Å.

The examples given here for human lysozyme show that the GRID method may be useful for assigning electron density in the solvent regions of X-ray maps. In particular, it may be of value in the assignment of ordered waters following structure refinement by simulated annealing. Note that GRID runs for this purpose may be performed without extensive computation. For example, less than 1 h of CPU time on a Silicon Graphics IRIS 4D/70 was required to calculate an energy map for a water probe for all of human lysozyme with a grid spacing of 0.5 Å.

Conclusions

The optimization of ligand–macromolecule interactions is fundamental to the design of therapeutic agents. The GRID method may be used as a tool for this purpose that can locate selective, energetically favorable ligand binding sites on molecules of known three-dimensional structure. It uses an empirical energy function with an explicit hydrogen bond term which is dependent on the chemical nature of the donor and acceptor atoms, their separation, and their relative orientation.

In the preceding paper,[4] the modeling of the dependence of hydrogen bond interactions on the orientation of probes capable of making up to two hydrogen bonds was discussed. Here, the methods have been extended to probes capable of making up to four hydrogen bonds. The interaction between the probe and the target is computed by analytically determining the orientation of the probe which optimizes the total interaction energy taking into account the spatial distribution of the hydrogen bonds at the probe and the relative positions of the probe's lone pairs of electrons and hydrogen atoms. The analytic procedure for orienting these probes provides greater accuracy and requires less computation than a numerical approach. Some probes may have more than one preferred hydrogen bond coordination geometry and these alternative geometries are considered explicitly when determining the orientation of the probe. An example of such a probe is water, which is experimentally observed to have tetrahedral and planar triangular hydrogen bond coordination geometries.

The application of the methods described here for NH_3^+ and water probes shows that the correspondence between the computed binding sites and the experimentally observed ones is improved when the spatial distribution of the hydrogen bonds at the probe is modeled. The GRID energy maps become more precise and easier to interpret.

Particular emphasis has been placed on the prediction of water binding sites. This is partly because there is a wealth of experimental data on water sites in biological macromolecular systems with which to compare the GRID calculations. However, it is also because a ligand binding to a target biomolecule may displace water from its surface, or may bind via bridging water molecules, and it is, therefore, important to be able to characterize the hydration properties of the target. Studies on human lysozyme show that program GRID can be used to predict the location of well-ordered waters with reliable accuracy and may be useful in assigning water sites during crystallographic refinement.[22] The method has also been used to study the solvation of the active site of cytochrome P450-cam.[25] It should be noted that it is possible to use program GRID to test automatically whether a given probe at a given position would interact with the target more favorably via a direct interaction or via an indirect interaction mediated by one bridging water molecule.[26]

Acknowledgment. We thank Peter Artymiuk and Colin Blake for providing crystallographic data for the structure of human lysozyme. Support from the Science and Engineering Research Council and the Medical Research Council is gratefully acknowledged. R.C.W. would like to thank Professor J. A. McCammon for the use of computational facilities for a part of this work.

Registry No. Lysozyme, 9001-63-2; ammonium, 14798-03-9; hydroxyl, 3352-57-6; water, 7732-18-5.

(23) Vedani, A.; Huhta, D. W. An Algorithm for the Systematic Solvation of Proteins Based on the Directionality of Hydrogen Bonds. *J. Am. Chem. Soc.* **1991**, *113*, 5860–5862.

(24) Pitt, W. R.; Goodfellow, J. M. Modelling of solvent positions around polar groups in proteins. *Protein Eng.* **1991**, *4*, 531–537.

(25) Wade, R. C. Solvation of the Active Site of Cytochrome P450-cam. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 199–204.

(26) Goodford, P. J. *GRID User Guide*; Molecular Discovery, Oxford, 1991.