PROTEINS: Structure, Function, and Genetics 26:363–376 (1996)

RESEARCH ARTICLES

Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures

Rob W.W. Hooft,[1]* Chris Sander,[1] and Gerrit Vriend[1]
[1]E.M.B.L., P.O. Box 10.220g, D-69012 Heidelberg, Germany

ABSTRACT **A method is presented that positions polar hydrogen atoms in protein structures by optimizing the total hydrogen bond energy. For this goal, an empirical hydrogen bond force field was derived from small molecule crystal structures. Bifurcated hydrogen bonds are taken into account. The procedure also predicts ionization states of His, Asp, and Glu residues. During optimization, side-chain conformations of His, Gln, and Asn residues are allowed to change their last χ angle by 180° to compensate for crystallographic misassignments. Crystal structure symmetry is taken into account where appropriate. The results can have significant implications for molecular dynamics simulations, protein engineering, and docking studies. The largest impact, however, is in protein structure verification: over 85% of protein structures tested can be improved by using our procedure. Proteins 26:363–376**

Key words: hydrogen bond; network; force field; protein structure; validation

INTRODUCTION

The earliest work noting the importance of hydrogen bonding in protein structure was the prediction by Linus Pauling of the α and β secondary structure elements now found ubiquitously in structures of globular proteins.[1,2] Pauling subsequently proposed that hydrogen bonding may be the key force in protein folding and stability. Ever since protein coordinates are available, studies have been performed on the impact of hydrogen bonding on protein structure. A systematic analysis by Baker and Hubbard[3] of hydrogen bonds in 15 well-determined proteins is the key review in which most of the current knowledge of hydrogen bonding in proteins is described.

Currently, protein folding is thought to be primarily driven by removal from solvent of hydrophobic surface of the unfolded chain, but a simple calculation can show that hydrogen bonds must also play an important role.[4] The stability of a folded protein over the denatured state often is around 5–15 kcal mol^{-1}, which is the total interaction energy of one to three hydrogen bonds.[5] So if three independent hydrogen bonds cannot be formed in the folded protein, the complete folding stability might be lost (assuming that in the unfolded state all hydrogen bond donors and acceptors in the protein are hydrogen-bonded to solvent molecules). Indeed, it has been shown that the fraction of unsatisfied hydrogen bond donors in protein structures is very low.[6]

If these facts are considered, it may seem surprising that in many protein structure determinations published to date, hydrogen bonding in the core of the protein is not studied in detail. Difficulties associated with structure determination, using X-ray diffraction, lie at the root of this problem: a difference of one electron cannot be seen directly in a protein structure using X-ray diffraction data at less than atomic resolution. This leads to inherent ambiguities in interpreting the interactions of polar hydrogen atoms that make an important contribution to the overall energy. Consider, for example, the OH groups of Ser, Thr, and Tyr side chains: usually it is not possible to locate the hydrogen atom by using electron density maps. It is similarly impossible to decide which of the two nitrogens in a His residue carries a hydrogen atom. Another consequence of the difficulty resolving a single electron in an X-ray diffraction experiment is that it is often not possible to see which of the two atoms is N and O in amide side chains. Similarly, the nitrogen and carbon atoms in the His side chain cannot be distinguished. However, from the direct environment of Asn, Gln, and His side chains, 75% of the orientations can be assigned unambiguously.[3] In other cases the two equivalent sides are pointing to atoms that can be either donor or acceptor themselves (e.g., water molecules), which makes the problem too complex for manual assignment.

Current refinement procedures for protein structures,[7,8] however, do not allow for "flips" of these side chains to occur automatically; the attention of the crystallographer is needed to correct the conformation for the residues that are built the wrong way around ("flipped").

*Correspondence to: Dr. Rob W.W. Hooft, E.M.B.L., P.O. Box 10.220g, D-69012 Heidelberg, Germany.
Received 19 January 1996; revision accepted 26 February 1996.

A relatively simple way of correcting the side-chain orientations of His, Asn, and Gln residues is to automate the manual procedure described above that looks at each side chain one at a time in its surroundings. This is the approach taken by a procedure HNQCHECK based on the HBPLUS program,[9] which works fine for the 75% of cases where the immediate surroundings leave only one possibility. It cannot take correlated configurations of two neighboring residues or water molecules into account.

Alternatively, it is possible to use a combinatorial approach. This has been coded into the NETWORK program,[10] which does an exhaustive optimization of hydrogen bond schema to make sure that the optimal hydrogen bonding network is found. NETWORK, however, does not allow side-chain flips as optimization steps. Also, like HNQCHECK, it uses a simple on/off force field to describe a hydrogen bond, and it does not take crystal structure symmetry into account. Even with these limitations, a molecular dynamics simulation starting from the NETWORK-optimized hydrogen positions is shown to be much more stable than a similar run using unoptimized hydrogen positions.[10]

The goal of the current project is to go further than HNQCHECK and NETWORK while maintaining the strong points of these methods, and to add as much knowledge about hydrogen bonds and protein structure as possible. The resulting program should find the best positions for polar hydrogens without manual intervention. To achieve this:

1. A hydrogen bond force field is derived specifically for this purpose.
2. The orientation of His, Asn, and Gln residues is taken into account, using an approach that considers the full hydrogen bond network at once.
3. Normal as well as bifurcated hydrogen bonds are scored, using an empirical hydrogen bond force field derived from small-molecule crystal structures.
4. Crystal symmetry is used to locate hydrogen bonds between neighboring molecules in X-ray structures.

COMPUTATIONAL PROCEDURES

To find a (nearly) optimal hydrogen bonding pattern, one needs a force field describing the hydrogen bond interaction energy of any proposed configurations, a definition of the degrees of freedom and the contraints, and a procedure for optimization.

Force Field

Description

The energetics of hydrogen bonds in proteins are subject to the same physical rules as hydrogen bonds in small molecules. To derive a good hydrogen bond "force field" we have used Boltzmann's equation relating the ratio of the frequency of occurrence of two states to the energy difference between them. An analysis of hydrogen bonds obtained in a previous study[11] of the Cambridge Structural Database (CSD)[12] was used as the source of the statistical data. Information about 14625 intermolecular hydrogen bonds was used, obtained from small-molecule X-ray crystal structures consisting only of H, C, N, O, F, and Cl atoms. For our purposes these hydrogen bonds were grouped into 16 classes of donor/acceptor pairs (Table I).

TABLE I. Hydrogen Bond Donors and Acceptors Distinguished for the Collection of Statistics

Accepting group	N-H donors*	O-H donors*
NR_2,	851	619
NR_3	305	350
NHR_2	329	100
NH_2R	164	44
OHR	639	2801
OR	3487	2409
OR_2	602	1064
OH_2	312	549

*The number of representatives of each combination found in the Cambridge Structural Database.

Four parameters δ, λ, ϕ, and θ are used for describing the geometry of a hydrogen bond (Table II). These four parameters can be calculated using the vectors V_0 through V_3 defined for the different acceptor classes in Figure 1:

$$\delta = |V_1| \tag{1}$$

$$\cos\lambda = \frac{V_0 \cdot V_1}{|V_0|\,|V_1|} \tag{2}$$

$$\cos\phi = \frac{V_1 \cdot V_2}{|V_1|\,|V_2|} \tag{3}$$

$$\cos\theta = \frac{(V_1 \times V_2) \cdot (V_2 \times V_3)}{|V_1 \times V_2||V_2 \times V_3|} \tag{4}$$

with $\lambda \geq 0$, $\phi \geq 0$, and $\theta \geq 0$. δ is the hydrogen bond distance. λ can be described as the hydrogen bond angle, ϕ as acceptor angle, and θ as *acceptor torsion angle* or *acceptor dihedral*.

For acceptor classes where V_3 is not defined, the torsion angle θ is set to 0.0; where V_2 is not defined, the angle ϕ is also set to 0.0. As such, ϕ, θ, and δ together form a polar coordinate system describing the position of the H atom in the hydrogen bond with respect to the acceptor.

The interaction energy for a particular donor-acceptor pair can be described as a function ΔE of the four parameters δ, λ, ϕ, θ defined in Equations 1–4:

$$E^0_{HB} = E_{id} + \Delta E(\delta,\lambda,\phi,\theta) \tag{5}$$

Here E_{id}=6 kcal mol^{-1} is the energy for an ideal hydrogen bond, and the function ΔE is chosen such that it is 0.0 for the ideal hydrogen bond configuration, and negative everywhere else. Unfortunately, it is not possible to get reliable statistics for all 18 possible combinations of donor type/acceptor type

TABLE II. Categories of Hydrogen Bond Acceptors Used and the Parameters Used to Describe the Hydrogen Bond Geometry

Acceptor	Used in	Parameters describing bond
NR_2	Histidine	δ, λ, φ, θ
NR_3		δ, λ, φ, θ
NHR_2		δ, λ, φ, θ
NH_2R		δ, λ, φ
OHR,	Serine	δ, λ, φ, θ
OR,	Backbone	δ, λ, φ, θ
OR_2, SR_2	Methionine	δ, λ, φ, θ
OH_2	Water	δ, λ, φ, θ
All others	S–H, anions	δ, λ

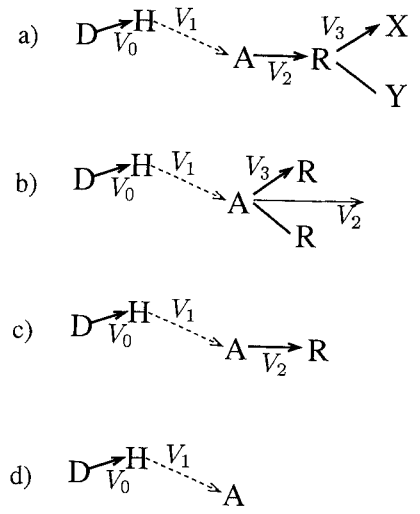

Fig. 1. Description of hydrogen bond geometry. The vectors V_0 through V_3 are used to define the parameters δ, λ, φ, and θ defining the hydrogen bond geometry (see text). D represents the donor atom, A the acceptor atom, and R, X, and Y other atoms in the accepting group. Four types of acceptor are distinguished: (**a**) Carbonyl oxygen acceptors. (**b**) Water acceptors (R = H,H), alcohols (R = H,X) and other acceptors with more than one covalently bound non-H atom; V_2 is the sum of all normalized vectors **A→R**. (**c**) Acceptors with one covalently bound atom except carbonyl oxygen. (**d**) Single atom anionic or S-H acceptors.

pairs in this four-dimensional parameter space. To simplify the problem, the assumption is made that δ and λ distributions neither influence each other nor the φ, θ distribution. This assumption makes it possible to decompose Equation 5 into a reduced form:

$$E^0_{HB} = E_{id} + \Delta E_\delta(\delta) + \Delta E_\lambda(\lambda) + \Delta E_{\phi,\theta}(\phi,\theta). \quad (6)$$

In this case independent statistical analyses can be performed for the δ value, the λ value, and the φ, θ value pairs. The number of observations from the CSD (Table I) is sufficient to get reliable one- and two-parameter distributions.

To verify the lack of correlation between the δ and λ parameters, a density contour diagram of these two parameters for the nitrogen donor/carbonyl acceptor case is given in Figure 2. Figure 3 gives a

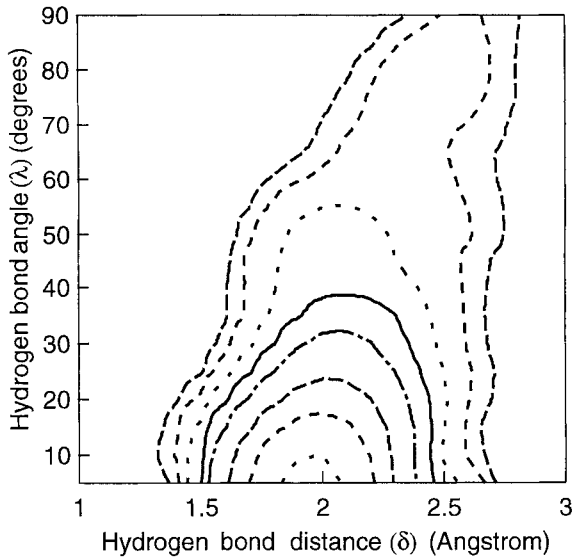

Fig. 2. Contour diagram depicting the number of observations of H-acceptor distance δ versus hydrogen bond angle λ for hydrogen bonds between a nitrogen donor and an oxygen acceptor. The plot shows near-independence of the two parameters: only at very high values of λ is the average preferred δ slightly longer. Contour levels are drawn at logarithmic intervals.

comparable scatter plot of the φ and θ parameters, illustrating that a further simplification is not possible.

To make the use of Equation 6 practical, not all pairs of donors and acceptors in the molecule are considered, but only donor–acceptor pairs that have an energetically favorable interaction (are "hydrogen-bonded"):

$$E_{HB} = \begin{cases} E^0_{HB} & \text{if } E^0_{HB} > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

No attempt was made to derive a functional description for $\Delta E_\delta(\delta)$, $\Delta E_\lambda(\lambda)$ and $\Delta E_{\phi,\theta}(\phi,\theta)$. Instead, the energies were tabulated in steps of 0.1 Å of δ, 5° of λ and φ, and 10° of θ (discretized values for the functions will be written without arguments in parentheses: ΔE_δ, ΔE_λ, and $\Delta E_{\phi,\theta}$). Four separate tables were made for all combinations of N/O donors and N/O acceptors for ΔE_δ and ΔE_λ; 18 separate tables (one for every donor–acceptor combination listed in Table II) were made for $\Delta E_{\phi,\theta}$. Since no statistics were collected for hydrogen bonds to sulfur, the ΔE_δ and ΔE_λ energies were calculated as if the acceptor was a nitrogen atom, but the $\Delta E_\delta(\delta)$ function was shifted by 0.4 Å to take the larger radius of the sulfur atom into account, and the final E^0_{HB} values were divided by 4. This procedure gives a hydrogen bond to sulfur an ideal interaction energy of 1.5 kcal mol^{-1} and has the secondary effect that deviation from the ideal geometry is penalized less. This is a very crude approximation, but unfortunately unavoidable, since not enough data can be obtained.

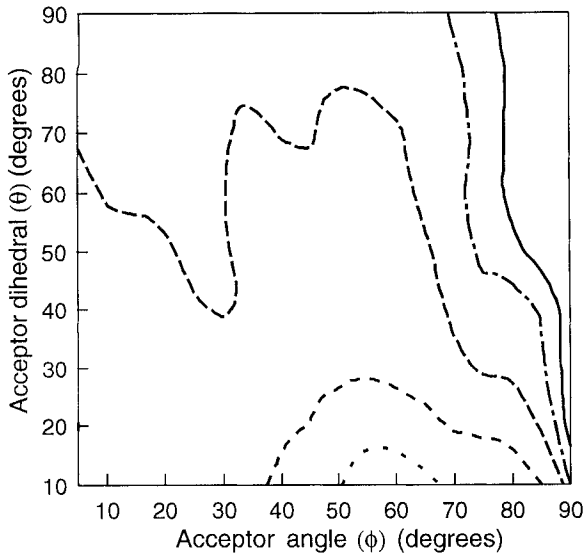

Fig. 3. Contour diagram depicting the number of observations of acceptor angle ϕ versus acceptor dihedral θ for hydrogen bonds between a nitrogen donor and a carbonyl acceptor. The plot shows that the distributions cannot be separated: the distribution of θ changes for different values of ϕ; for low values of ϕ the θ distribution is flat (as expected by symmetry considerations), and at intermediate values of ϕ there is a strong preference for low values of θ. Contour levels are drawn at logarithmic intervals.

Furthermore, since hydrogen bonds to sulfur play a minor role, the exact form of the potential does not influence the results significantly.

Derivation of the force field parameters

The values ΔE_δ were derived from the small molecule statistics, using the inverse of Boltzmann's equation:

$$p = Ae^{-\Delta E/RT} \tag{8}$$

with p representing the frequency of observations, T representing the temperature, A a normalization constant, and $R = 8.31451\ \text{J mol}^{-1}\ \text{K}^{-1}$. However, a blind application of this formula would give biased results. In the absence of any energetical interactions, the number of observations $p(\delta)$ would be proportional to δ^2. To counter this distribution, counting statistics from the CSD were first corrected with a *geometric factor* or *Jacobian* δ^{-2}. So, if the sign from Equation 8 is included in ΔE_δ:

$$p_\delta^c = p_\delta \delta^{-2} \tag{9}$$

$$\Delta E_\delta = RT \ln p_\delta^c + [\text{normalization constant}]. \tag{10}$$

The temperature T is set to 298 K.* The values p_δ are taken from a histogram of δ values found in the CSD. To compensate for the fact that protein structures are expected to be less exact than small molecule structures, a random displacement of up to 0.25 Å was applied to the H atom coordinates from the CSD in compiling the histogram. The value of 0.25 Å was chosen as the average coordinate error in proteins is estimated to be around 0.4 Å,[14,15] but local interatomic distances are expected to be significantly more accurate than this due to refinement procedures using noncrystallographic information (e.g., ref. 7). To prevent the statistics from getting worse due to the partial coordinate randomization, every CSD hydrogen bond was counted 250 times using different random displacements.

The ΔE_λ values were derived in a similar way. The only difference with δ lies in the form of the geometric factor:

$$p_\lambda^c = \frac{p_\lambda}{\sin \lambda} \tag{11}$$

$$\Delta E_\lambda = RT \ln p_\lambda^c + [\text{normalization constant}]. \tag{12}$$

To calculate the $\Delta E_{\phi,\theta}$ tables, a joint histogram was made for ϕ and θ. For the ϕ angle a geometric factor was again taken into account. The combined $\Delta E_{\phi,\theta}$ was derived from

$$p_{\phi,\theta}^c = \frac{p_{\phi,\theta}}{\sin \phi} \tag{13}$$

$$\Delta E_{\phi,\theta} = RT \ln p_{\phi,\theta}^c + [\text{normalization constant}]. \tag{14}$$

The resulting distribution $p_{\phi,\theta}^c$ is fourfold or sixfold symmetric in θ, depending on the acceptor type. To remove this intrinsic symmetry, θ is always reduced to $0 \le \theta \le 90°$ using the equivalences $-\theta \rightarrow \theta$ and $180 - \theta \rightarrow \theta$ in case the acceptor atom carries two substituents and for carbonyl groups. In case the acceptor atom carries three substituents, θ is reduced to $0 \le \theta \le 60°$, using $\theta - 120° \rightarrow \theta$ and $-\theta \rightarrow \theta$.

In Equations 10, 12, and 14, the normalization constants were chosen such that the highest tabulated value of ΔE_δ, ΔE_λ, and $\Delta E_{\phi,\theta}$ is 0.0, so the sum of these terms is 0.0 for an ideal hydrogen bond (as required by Eq. 5).

Completing the force field

To calculate the total hydrogen bonding energy of any particular configuration, the contributions of the individual hydrogen bonds are added. All N and O atoms carrying H atoms are used as donors; S–H and C–H donors are not taken into account in our procedure. All O atoms are considered as acceptors, as are histidine side-chain N atoms that do not carry a H atom, and cysteine and methionine S atoms. Aromatic π electron systems are not considered as acceptors. For each donated H-atom a maximum of two positive E_{HB} values are counted; if any hydro-

*The temperature choice for this application of Boltzmann's law has led to disputes,[13] but we think that using 298 K is the best we can do here. Furthermore, the values will be used in approximately the same way as they were derived, canceling a large fraction of the expected error.

gen atom participates in more than two hydrogen bonds, only the two interactions having the highest E_{HB} values are used. This restriction makes the procedure faster and the memory consumption lower, while it does not significantly alter the results. All atoms with an X-ray occupancy (weight) ≤ 0.5 in the PDB file are ignored to prevent formation of overlapping (and possibly conflicting) hydrogen bond networks.

Energetically unfavorable contacts form exceptions to the simple addition rule:

1. Whenever a hydrogen atom comes within 2.0 Å of a hydrogen atom attached to its acceptor, the interaction is not taken into account. This is done to prevent situations where two O atoms donate their hydrogen atoms to each other.
2. Whenever a hydrogen atom comes within 2.5 Å of a positively charged atom, a penalty of 12 kcal mol^{-1} (just a large value) is subtracted from the total interaction energy. This makes it possible to describe the coordination of metal ions without explicit interaction terms.

In our implementation it is also possible to scale the hydrogen bond energy with a *reliability factor:*

$$\Delta E^{r}_{HB} = \begin{cases} \Delta E^{0}_{HB} * (1 - B/60) & \text{if } B < 60 \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Here B is the highest of the two temperature factors of the donor and the acceptor atom. Using this scaling, hydrogen bonds between highly mobile atoms for which the coordinates are relatively unreliable are given a lower weight in the determination of the hydrogen bond network.

Application of the force field to protein structures

The force field is meant for hydrogen bonds only. Due to the fact that it was derived from an analysis of observed hydrogen bonds, it contains all effects that can influence the preferred geometry of hydrogen bonds. Among these are the actual electrostatic interaction from the hydrogen bond, but also steric effects. For normal force fields (e.g., MM2[16]) one needs to avoid counting forces twice. Since the separation of our hydrogen bond force field into standard components is extremely difficult, addition of the pure hydrogen bond force to a general mechanics program is practically impossible. However, the presence of all effects contributing to hydrogen bonding makes our force field optimally suited for positioning hydrogen atoms, since the energy terms derived here are the only ones to be applied.

Degrees of Freedom and Restrictions

Three classes of degrees of freedom are considered: amino acid side-chain flips, the presence or absence of hydrogen atoms in acidic and basic groups, and the rotational freedom in the placement of hydrogen atoms. All of these degrees of freedom will be referred to as *ambiguities.*

Some restrictions of the freedom are necessary to avoid physically unrealistic situations. A possible way of handling restrictions is to disallow a certain conformation completely. Our program uses the alternative of penalties: situations that are "undesirable" are given an energy penalty. An advantage of this approach is that the undesirability of the situation can be expressed by the size of the penalty.

The penalties, and other numeric values described in the subsections below, can easily be adapted to any specific situation: the numbers presented here are the current defaults in the program. Although sometimes arbitrary, most seem plausible, and the results have been checked for many cases using visual inspection and consultation with the crystallographers solving the structure. The values given are those we used to do the work described in this paper.

Amino acid side-chain flips

Using crystallographic methods it is difficult, even at high (2.0 Å) resolution, to distinguish between two conformations of the side chains of His, Gln, and Asn. In our procedure both the given conformer and the inverted one are considered; if the inverted one is used a *flip penalty* of 1.2 kcal mol^{-1} is subtracted from the interaction energy; that is, we will be approximately 90% certain before proposing a flipped configuration. This precaution is taken to prevent the loss of biological or chemical information used by the crystallographer in nonobvious cases. "Flips" are performed using proper 180° rotations of the last χ angle in the side chain, such that differences in bond length do not disturb the results. In cases where a histidine residue coordinates a metal atom,[17] only the conformation that has an N atom in the coordinating position is considered. If one of the two conformations for a histidine can form a hydrogen bond to a solvent atom that is not explicitly given (i.e., one of the two nitrogens in the side chain has a nonzero solvent accessibility, which indicates that a solvent molecule invisible to the crystallographer is touching the atom), a bonus of 2.4 kcal mol^{-1} is given to this conformation, that is, somewhat less than the average value seen in our force field for a hydrogen bond between protein and an explicitly given water molecule. This way even solvent atoms that have not been found in the X-ray structure can be taken into account. Symmetry-related molecules are taken into account for the calculation of the solvent-accessible surface.

Presence of hydrogen atoms

The N and O atoms in histidine, glutamic acid, and aspartic acid side chains can be either donor or acceptor, depending on the presence of a bound hydrogen atom; this introduces a new class of degrees

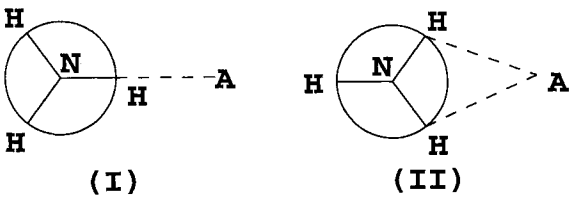

Fig. 4. NH_3 groups tend to have a very small barrier to rotation. Situation (I) shows a configuration with a single hydrogen bond from the NH_3 group to an acceptor A and situation (II) shows a situation (the NH_3 group turned by 60°) with two hydrogen bonds of lower quality. The hydrogen bond energy of these two states is very similar.

of freedom. If the hydrogen atom on one of these atoms is present, the carrying atom is only regarded as a donor; if it is absent, the atom is only treated as an acceptor. Although the carboxylic acid oxygen can, in principle, be donor and acceptor at the same time, this situation is very rarely observed in small molecule structures. Therefore it was decided that this could be neglected in proteins too.

The three different side chains all have two atoms that can be donor or acceptor. They differ in the preferred number of connected hydrogen atoms. For the acidic amino acids, the preferred number of hydrogen atoms is zero. A penalty of 4.5 kcal mol^{-1} is subtracted from the interaction energy if one hydrogen atom is present; this way a hydrogen atom will only be added if there is evidence for a very strong hydrogen bond. For histidine side chains the preferred number of hydrogen atoms is one. A penalty of 6.0 kcal mol^{-1} is subtracted from the interaction energy if no hydrogen atoms are present (this is not expected to occur in practical cases), a penalty of 1.2 kcal mol^{-1} if both hydrogen atoms are present. This last value was chosen such that following a Boltzmann distribution at room temperature about 1 in 10 His side chains is expected to be charged.

The protonation state of NH_3 groups is not changed. They tend to have a very small barrier to rotation (Fig. 4). The low specificity in the local hydrogen bonding scheme makes it difficult to distinguish the ionization states. Furthermore, since the changes in the resulting hydrogen bonding scheme are probably small, it is expected that the inclusion of ionization states would only have a very small effect on the overall hydrogen bonding scheme.

Orientation of hydrogen atoms

For many of the hydrogen bond donors (backbone N, arginine side chain, etc.) accurate coordinates of the hydrogen atom can be calculated by using simple geometric construction: hydrogen atoms are placed at ideal triangular or tetrahedral positions. Other hydrogen atoms (alcoholic OH, lysine NH_3, N-terminal [if not proline] NH_3) are free to move on a circle. Hydrogen atoms connected to water molecules make a H–O–H angle of approximately 110°,

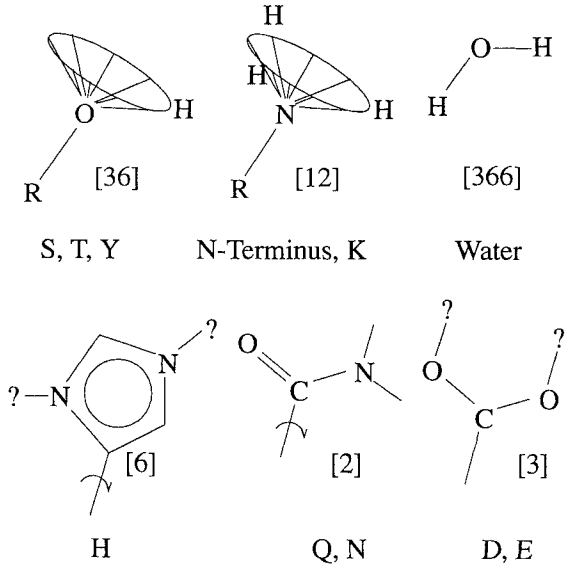

Fig. 5. Diagram showing the variable (ambiguous) hydrogen bond donors in protein structures. Numbers in square brackets represent the number of possibilities that are considered. Amino acids affected are indicated by their one-letter abbreviation.

but the orientation of the molecule is free. In all cases hydrogen atoms are placed at a distance of 1.0 Å from the donor atom.[3]

Both the amino acid side-chain flips and the presence or absence of hydrogen atoms have a discreet number of possibilities. In contrast, the rotational freedom described here is continuous. To simplify the optimization procedure the rotations were also discretized (Fig. 5):

1. NH_3 groups have 12 different angular orientations, each 10.0° apart.
2. OH groups have 36 different angular orientations, each 10.0° apart.
3. Water molecules have 366 different orientations, constructed using pairs of points from a 55-point Fibonacci sphere. This resulted in a homogeneous distribution of orientation without bias for a specific direction (P. Zielenkiewicz, personal communication, 1987).

This discretization method is completely different from the approach used in the NETWORK program.[10] In that program only a strictly limited set of hydrogen positions is considered for each of the rotational degrees of freedom of a donor, each one optimized for a single potential acceptor. As a consequence of this, bifurcated hydrogen bonds (a single hydrogen donated to two different acceptors at the same time) will not be found. Since we want to be able to locate bifurcated hydrogen bonds, this schema, which is computationally considerably less expensive than ours, was not considered.

TABLE III. Calculation of the Number of Possible Hydrogen Bond Networks To Be Considered for PDB Entry 5TIM*

Number	Type	Choices	Total
10	His	8	8^{10}
26	Glu	4	4^{26}
26	Gln	2	2^{16}
16	Asp	4	4^{16}
20	Asn	2	2^{20}
10	Tyr	36	36^{10}
24	Thr	36	36^{24}
36	Ser	36	36^{36}
2	N-term	12	12^{2}
34	Lys	12	12^{34}
279	Water	366	279^{366}
	Total $N_c \approx$	10^{1088}	

*This does not take into account that some ambiguities can be fixed using prior knowledge.

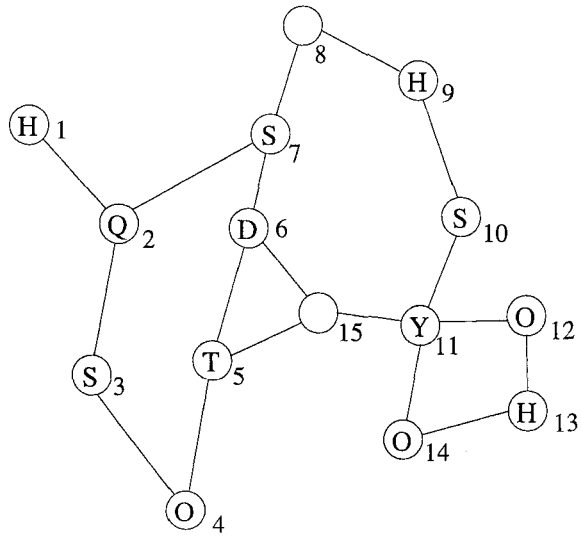

Fig. 6. Network of potential hydrogen bonds in a hypothetical protein. Unambiguous donors/acceptors are given as open circles (8 and 15). Water molecules are indicated by O, and ambiguous side chains of amino acids with their one-letter code. Even for a relatively small example like this, the total number of possible hydrogen bond networks is already enormous. However, the network can be treated as two independent subnetworks (1–7 and 9–14). Furthermore, the optimal position of (1) can be established as a function of only (2), reducing the first subnetwork to (2–7). Similarly, (9) is only dependent on (10), and after that (10) is only dependent on (11). This reduces the second subnetwork to (11–14).

Reduction of the Problem Size

Using our less restricted discretization, the total number of possible configurations to be considered (N_c) in a large protein structure such as triose phosphate isomerase (PDB entry 5TIM[18]) is quite high (Table III). The choice of the right optimization strategy therefore is very important.

Our optimization protocol starts with the construction of a list of all possible hydrogen bonds. Here all possible side-chain conformations of donor and acceptor and/or all possible positions of the hydrogen atom on the donor are taken into account. Acceptors in symmetry related molecules can be considered if desired; this is sufficient to ensure that all intersymmetry hydrogen bonds will be taken into account. From the resulting list all *potential acceptors* for each of the donors are extracted and stored.

After this, for each of the ambiguities[†] a list is made of all *affected donors*: A donor is affected by an ambiguity if the value of the hydrogen bonds originating at that donor can be influenced by changing the status of the ambiguity. For an ambiguous donor the affected donors are the ones for which it is a potential acceptor. For an ambiguous residue the affected donors are all donors that have one or more of the side-chain atoms of the residue as a potential acceptor.

Precalculating lists of affected donors helps because these lists generally have only few entries. This is because the cooperative effect (the effect that a hydrogen bond is stronger if the acceptor is donating another hydrogen itself) is not taken into account in our force field. If the cooperative effect would be taken into account, many more hydrogen bond energies would be influenced by every change in configuration. This could slow down calculations by a few orders of magnitude.

Any ambiguity that cannot influence the total H-bond energy is fixed to a natural (neutral) value at this point. For example, if a histidine side chain cannot form any hydrogen bonds, the conformation of the side chain is kept as given in the input coordinates, and a hydrogen atom is attached to the Nϵ2 atom.

Subsequently, the lists of affected donors and potential acceptors are searched to locate clusters of ambiguities in the hydrogen bond network that can form hydrogen bonds between them, but not to any other ambiguities. The hydrogen bond network is thus subdivided in subnetworks. Each of the subnetworks is now considered independently (Fig. 6).

In a subnetwork some ambiguities (e.g., H1 in Fig. 6) could be linked to only one other ambiguity (Q2). For each of the positions of Q2 the optimal choice of H1 can be calculated independently of all other ambiguities, and these results can be tabulated. When this has been done, H1 does not have to be considered any longer, and is removed from the subnetwork. This *cutting* procedure is repeated until no more ambiguities can be removed.

If a water molecule is only linked to one other water molecule, the total number of evaluations

[†]Terminology: donors for which the hydrogen position is variable and atoms that do not always carry a hydrogen atom are collectively described by the term *ambiguous donors*. Side chains of His, Asn, and Gln are called *ambiguous residues*. Both classes taken together are referred to as *ambiguities*.

needed for cutting is $366^2 = 133{,}956$. Since this number of evaluations requires a considerable amount of time (10 seconds on a fast workstation), no cutting is done. Cutting is only performed when it can be done using less than $10^{4.3} \approx 20{,}000$ evaluations (the *exhaustive search limit,* a parameter of the program).

For the remaining part of the subnetwork an exhaustive evaluation is performed if the total number of evaluations needed is less than the exhaustive search limit. For subnetworks that are bigger than this, a heuristic approach is followed as described in the following section.

Heuristic Optimization Procedure

To find good configurations for subnetworks that are too big to solve by brute force evaluation, the *threshold accepting* method (TA)[19,20] was selected. TA is a stochastic procedure like *simulated annealing* (SA), in which the exponential threshold is replaced by a sharp cutoff. For the traveling salesman problem—a classical problem used in mathematics to test optimization strategies—TA has been shown[19] to be more efficient than SA, and less dependent on a good cooling strategy.

In SA the chance p_a that a new configuration with an energy E_{new} will be accepted is

$$p_a = \begin{cases} e^{(E_{old} - E_{new})/RT} & \text{if } E_{new} > E_{old} \\ 1 & \text{otherwise.} \end{cases} \quad (16)$$

With T the temperature that is decreased following a *cooling strategy,* and $R = 8.31451$ J mol^{-1} K^{-1}. In TA this is replaced by

$$p_a = \begin{cases} 0 & \text{if } E^{new} - E^{old} > E_T \\ 1 & \text{otherwise.} \end{cases} \quad (17)$$

Here E_T is a *threshold energy*. Since in our force field the interaction energy must be maximized instead of minimized, the equation was negated:

$$p_a = \begin{cases} 0 & \text{if } E_{HB}^{old} - E_{HB}^{new} > E_T \\ 1 & \text{otherwise.} \end{cases} \quad (18)$$

Our cooling strategy starts with the threshold at 4.5 kcal mol^{-1}. In every TA step one randomly selected ambiguity is set to a random value different from the previous one, and the hydrogen bonding energy is reevaluated (Only the affected donors for the ambiguity need to be recalculated here). The step is accepted or rejected following Equation 18. After a series consisting of $N_s = \lfloor \log_2 N_c \rfloor$ steps ($\lfloor x \rfloor$ is the highest integer lower than x), E_T is reduced by $\Delta E_T = 0.09$ kcal mol^{-1}. This procedure is repeated until $E_T = 0$ is reached. Values for ΔE_T and N_s are not critical. The values given here are a result of experiments; longer simulations do not result in a measurable improvement in network quality.

To finish the maximization protocol, all ambiguities are individually optimized in single steps to fine-tune the hydrogen bonds without changing the connectivity of the network.

The whole procedure (taking a random starting point, performing a TA maximization, followed by local optimization) is repeated a few times (we normally use 7), and the best of the results is selected. Also a report is generated listing all ambiguities that have different states in the independently optimized results. The final energies of the optimizations normally differ no more than a few kilocalories per mole, mostly due to different orientations of water molecules.

The threshold accepting method was selected after experiments with genetic algorithms (GA) and Monte Carlo simulated annealing (SA) approaches failed to give satisfying results.

RESULTS AND DISCUSSION

Results for a Test Set

A test set consisting of the atomic coordinates of 368 protein structures from the Protein Data Bank[21] (December 1994) was selected based on the following criteria:

1. The word "mutant" does not occur in any SOURCE record.
2. The structure was determined using X-ray crystallography.
3. The structure contains at least one protein molecule.
4. No substrates of more than one atom are present (the capability to consider hydrogen bonding to ligands has since been added to the program).
5. There are at least two water molecules.
6. The symmetry records are valid or correctable, and there are no Van der Waals clashes between symmetry-related molecules.[22]

For all structures in this test set an optimized hydrogen bond network was determined. The total CPU time required on a fast workstation was about 4 days. Numerous examples have been carefully inspected using a graphic display, to confirm that results intuitively look right. Most of the parameters were optimized on the basis of these inspections. Only the final statistics are described in this section.

After optimization, 774 of the 2177 histidine residues in the test set have only Nδ1 as donor, 748 only Nε2, 395 both Nδ1 and Nε2; the remaining 280 do not make side-chain hydrogen bonds with protein or solvent molecules present in the PDB file. This means that between 18% and 21% of the histidine side chains are charged.

Of the 6269 aspartic acid residues in the test set, 10 were found to carry a proton (0.2%). Of the 5727 glutamic acid residues, 12 were found to carry a proton (0.2%).

About 14% of the His and 18% of the Gln and Asn side chains in the test set are "flipped" (terminal dihedral angle of the residue is changed by 180°) in

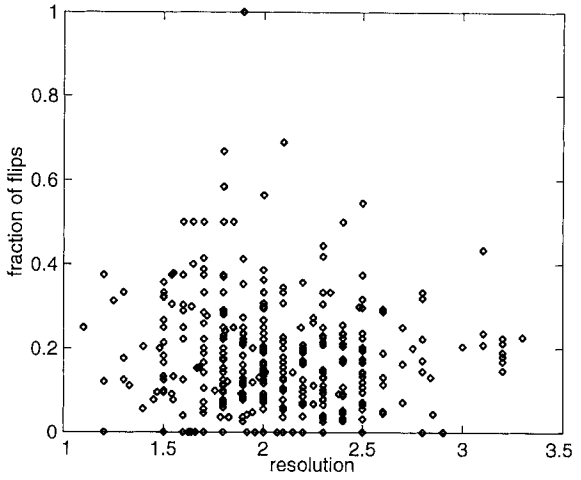

Fig. 7. Fraction of flipped residues as a function of the crystallographic resolution of the structure. No correlation is observed.

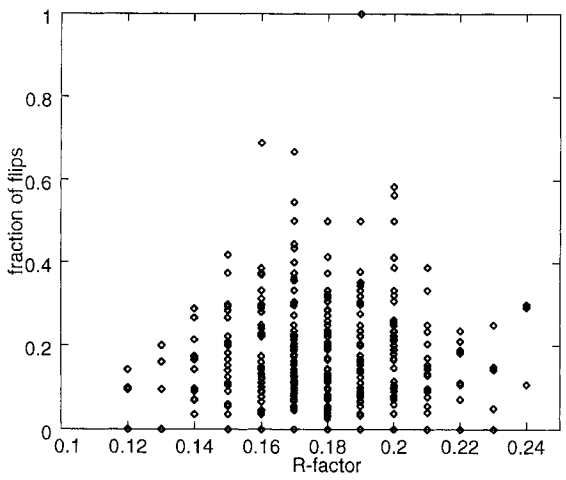

Fig. 8. Fraction of flipped residues as a function of the cyrstallographic R-factor of the structure. No correlation is observed.

the optimized structure (295 of 2177 for His, 701 of 3924 for Gln, and 929 of 5095 for Asn). Only 36 of the 368 protein structures in the test set (10%) do not have any residues that are flipped by our procedure. A number of the flipped side chains are only hydrogen-bonded to water molecules; a sizeable fraction of these might be artifacts due to inaccurate water coordinates. The percentage of flipped residues is not correlated with the resolution (Fig. 7) or the R factor (Fig. 8) of the structures. Contrary to what one might hope, it is also not correlated with the deposition date of the PDB file (Fig. 9). A slight, but insignificant tendency was observed for structures solved at better resolutions to have a better average energy of the hydrogen bonds involving protein atoms (Fig. 10).

For our test set, the HBPLUS program[6] reports 13% of the His, 14% of the Gln, and 14% of the Asn residues either as *highly suspect* or as *slightly suspect*. As expected, these numbers are lower than ours: as described earlier, 25% of all cases cannot be resolved looking only at nearest neighbors. The fraction of flipped His residues for our test set using the HBPLUS program is higher than that reported by the authors[9] (15.1% for Asn and Gln side chains and 9.9% for His side chains). This is most probably caused by a different selection procedure for the test set. 34 residues marked as *highly optimal* in the HBPLUS output are flipped by our program (10 His, 18 Gln, 6 Asn), 80 residues marked as highly suspect were not flipped (47 His, 20 Gln, 13 Asn). Many of the last class are flipped by HBPLUS because of weak hydrogen bonds and/or accessibility. These questionable cases are not flipped in our procedure due to the influence of the flip penalty.

If the flip penalty is increased from 1.2 kcal mol^{-1} to 3.0 kcal mol^{-1}, 9.4% of all ambiguous residues are still flipped, 7.9% of the ambiguous residues do not

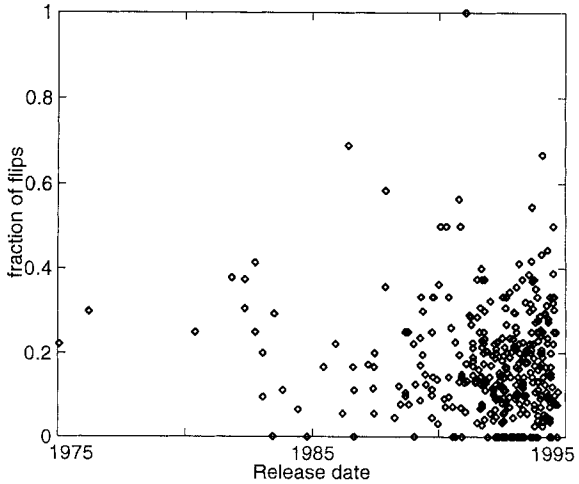

Fig. 9. Fraction of flipped residues as a function of the submission date of the Protein Data Bank file. No correlation is observed.

flip any longer, and 0.3% did not flip before but do so now (this can happen in marginal cases because of the partially stochastic nature of the algorithm). Using this higher flip penalty, none of the residues marked highly optimal by HBPLUS are flipped by our procedure. The higher value of the flip penalty results in more false negatives and less false positives than the standard value; this might be valuable in some cases.

In the standard calculation of the full test set, 9.3% of all hydrogen bonds found (20,415 of 219,479) are donated to atoms in symmetry-related molecules.

A significant fraction of the polar hydrogen atoms in side chains form hydrogen bonds with more than one acceptor at the same time; figures range from 1.5% for Tyr OH to 18.5% for His Nϵ2. Also, 8% of all

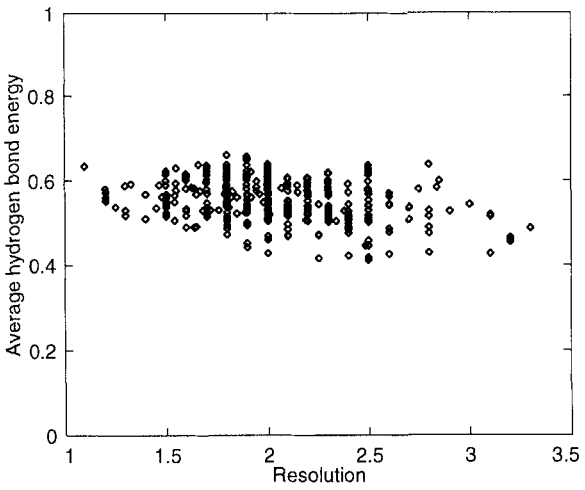

Fig. 10. Average hydrogen bond energy for backbone-backbone hydrogen bonds as a function of the resolution of the structure. An slight decrease of the hydrogen bond quality for lower resolution structures is observed. Higher energy values indicate more favorable conformations.

backbone N donors are bifurcated: in many α helices (with standard i, $i-4$ backbone hydrogen bonding) additional hydrogen bonds between residues i and $i-3$ are found to contribute significantly.[23]

In 12 structures of the test set hydrogen atoms end up within 2.5 Å of a metal ion. In 10 structures this is a caused by a water molecule that is located very close to a metal atom (1.4–1.8 Å, one distance of 0.7 Å was found). The other two cases are a calcium ion coordinated by two backbone nitrogen atoms, and a zinc ion coordinated by the N terminus of the polypeptide chain. The last case arises because our program lacks the flexibility to change the charge on an NH_3 group. The other 11 structures have been correctly signalled as having a problem.

Four proteins from the set were also calculated with the *brute force threshold* changed to slightly above 366[2]. Calculations with this setting take 2–6 times longer. Changes in the final hydrogen bonding energy are minimal, it improves by 0.6–3.6 kcal mol^{-1}. The actual visible changes in the hydrogen bond network are even smaller than this improvement suggests: 3% of the hydrogen bond network is changed (72 hydrogen bonds are replaced by 81 new ones), the majority of these are water–water hydrogen bonds. Only 3% of the changes are between protein residues (two new bonds found). In most cases this very slight improvement in hydrogen bond energy is not worth the extra time taken by the computation; most probably a better result can be obtained by repeating the standard calculation a number of times.

Neutron Diffraction Studies

The current PDB contains a number of structures that were determined using neutron diffraction experiments. For neutron scattering, the "visibility" of the atoms is not determined by the number of electrons they carry, but by the scattering length of the nucleus. The scattering length for hydrogen nuclei is of the same order of magnitude as for non hydrogen atoms, making hydrogens as visible as other atoms. Because of this property the neutron scattering technique can in principle be used to locate hydrogen atoms experimentally.

Predicted hydrogen positions using our procedure were compared with the experimentally determined positions for four neutron diffraction structures from the PDB. Many differences were found that were clearly due to problems in the experimentally determined structure. The comparisons are therefore not useful as a way to assess the accuracy of our calculations.

Examples

Baker and Hubbard[3] give a few interesting examples of clusters of buried water molecules in the actinidin structure (PDB identifier 2ACT[25]) (cf. Fig. 20 in ref. 3). For water molecule 1, surrounded by two donors and three acceptors (Fig. 11), our best hydrogen bond network has one bifurcated hydrogen bond, such that all hydrogen bond potential is satisfied. The pair of water molecules No. 29 and No. 36 fill a surface crevice (Fig. 12). In the final hydrogen bond network two other water molecules, No. 105 (as donor) and No. 59 (as acceptor) complete the hydrogen bonding scheme in an almost perfect tetrahedral way. The buried cluster of eight water molecules for which the hydrogen bonding scheme was completely hand-assigned in Baker and Hubbard[3] is assigned in exactly the same way by our procedure. (Fig. 13).

In the two pairs of closely approaching acidic side chains (Glu 16/Asp 115 and Asp 33/Asp 213)[26] in penicillopepsin (PDB identifier 1APT) none of the four residues is identified as predicted to be protonated using our approach, because the geometry of the hydrogen bond that would be formed is not good enough to overcome the 4.5 kcal mol^{-1} energy penalty for protonation. This may suggest either that this penalty is too high, or that the formation of a hydrogen bond between two acidic groups should be treated specially, possibly because the "local pH" may approach the pK_a of the side chains. In this same structure, OE2 of Glu 45 is protonated and forms a nice hydrogen bond to OD1 of residue Asn 84.

Validation of the Sulfur-Related Force Field

The parameters chosen for hydrogen bonds to sulfur were not directly based on small molecule statistics. Therefore an extra verification is appropriate.

In order to confirm that the results for hydrogen bonds to sulfur are insensitive to changes in the

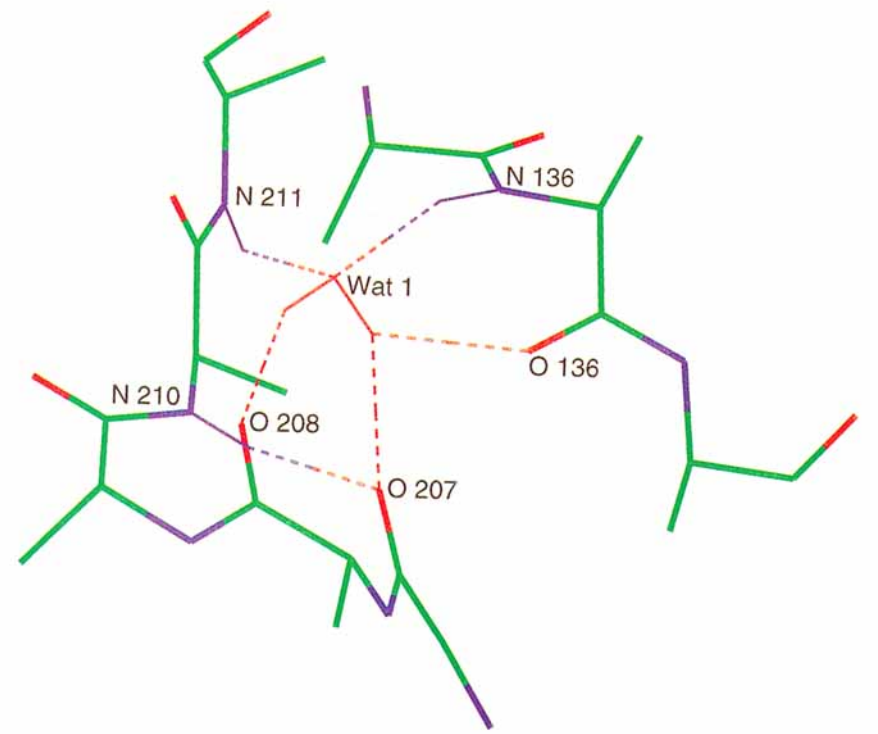

Fig. 11. Buried water molecule in the acinidin structure, with optimised hydrogen bonds indicated as dashed lines. All three surrounding acceptors are satisfied due to one of the hydrogens making a bifurcated hydrogen bond.

force field, the test run was repeated with all energy terms for hydrogen bonds to sulfur doubled. The original calculations resulted in 948 hydrogen bonds to sulfur (305 to methionine, 643 to cysteine). Using the higher force constants 143 new hydrogen bonds to S were found, of which 100 had water donors. The 43 remaining changes were checked manually, and only two of these were considered improvements (in most cases other hydrogen bonds were broken; in a number of cases residues had to be flipped to be able to form a hydrogen bond to S). Due to the partly heuristic nature of the algorithm, also four hydrogen bonds to S that were found with the lower energy force field were not reported in the recalculation with higher energy terms, all of these having water as donor.

Even doubling the importance of hydrogen bonds to sulfur does not significantly change the optimized hydrogen bond network. This indicates that the exact values of the sulfur parameters are not important.

B-Factor Scaling

Many structures in the PDB have a few not very well-defined regions. These regions often have thermal motion parameters (B factors) that are higher than those for the surrounding residues. In our force field normally all hydrogen bonds are weighted equally. This could have as effect that a badly defined residue forces a "flip" onto a neighboring well-defined residue. To prevent such situations, the possibility was created to scale down the scores for hydrogen bonds involving atoms with high thermal motion parameters.

In the few cases which we have inspected, this change did affect parts of the hydrogen bond scheme involving highly mobile loop regions. This modification of the force field might prove valuable in specific cases, but since it changes the empirical force field in a way that is statistically incorrect, we do not think that it is advisable to use the changed force field for all structures.

Water–Water Hydrogen Bonds

The number of possible orientations of water molecules is very large, and the positions of water molecules as determined using X-ray crystallography are less accurate than positions of protein atoms. For these two reasons the hydrogen bonds that are found between pairs of water molecules are less reliable than those that involve protein residues.

Calculations were therefore performed on a few structures with a modified force field in which the hydrogen bonds between water molecules were not scored.[10] For structures that have large water clusters this can speed up the calculation considerably. Resulting hydrogen bond patterns for these calculations were not satisfactory. There were quite a number of changes even in hydrogen bonds between protein and water molecules, as the water molecules in

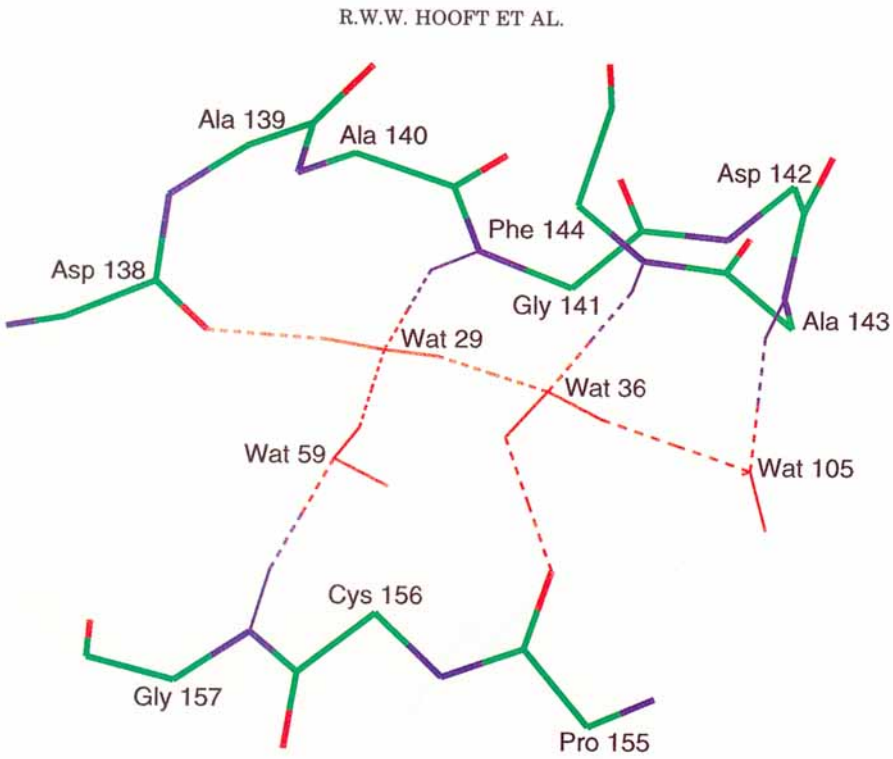

Fig. 12. Pair of buried water molecules in the actinidin structure, with optimiised hydrogen bonds indicated as dashed lines. Two other water molecules help to satify all hydrogen bond possibilities in near-perfect tetrahedral way.

the first hydration shell tended to orient themselves optimally for only one hydrogen bond. Furthermore, in these calculations detailed reliable water–water information is lost for clusters of buried water molecules (Fig. 12 and 13). The idea was therefore not pursued any further.

CONCLUSIONS

A method has been developed to locate polar hydrogen atoms in protein structures. The method makes use of an empirical hydrogen bond force field that was derived specifically for this purpose. Using this force field not only one to one hydrogen bonds but also bifurcated hydrogen bonds can be analyzed. In the optimization procedure so-called "flips" of residues that cannot be deduced directly from X-ray data are taken into account. Hydrogen bonds between symmetry-related molecules are also used.

Calculations on a test set show that the number of flips in structures is not related to classic quality measures such as crystallographic R factor or resolution, nor to the publication date. Compound protein quality measures (e.g., ref. 27) that rely on databases of reliable structures could be miscalibrated because of the presence of over 15% of misassigned His, Asn, and Gln side chains. We therefore plan to produce a database of corrected structures for improving empirical validation methods in the future.

Molecular dynamics simulations have been performed to test our method.[28] In most cases molecular dynamics runs starting from our corrected molecules with optimized hydrogen positions equilibrated quicker to a lower root-mean-square (RMS) value than molecular dynamics runs starting from the uncorrected coordinates with protons placed in standard positions by the molecular dynamics program. The improvement is mostly the result of selecting the correct local minimum for a larger number of hydrogens. Energy minimizations using GROMOS[29] change these hydrogen positions only by a very small amount, indicating a high degree of similarity between our hydrogen bond force field and a molecular mechanics description of a hydrogen bond.

The procedure as described has been incorporated as a module in the WHAT IF program, and is available from the authors under the normal conditions for this program package (a nominal fee for the academic community). As a tool for structure validation it is also part of the freely available WHAT CHECK program that can be downloaded from our ftp server. Further information can be obtained from the World Wide Web on http://www.sander.embl-heidelberg.de/whatif/ or via E-mail to hooft@embl-heidelberg.de or vriend@embl-heidelberg.de.

ACKNOWLEDGMENTS

This work was done in the context of the Protein Structure Database Validation Project funded by the European Commission. This project has as one of its

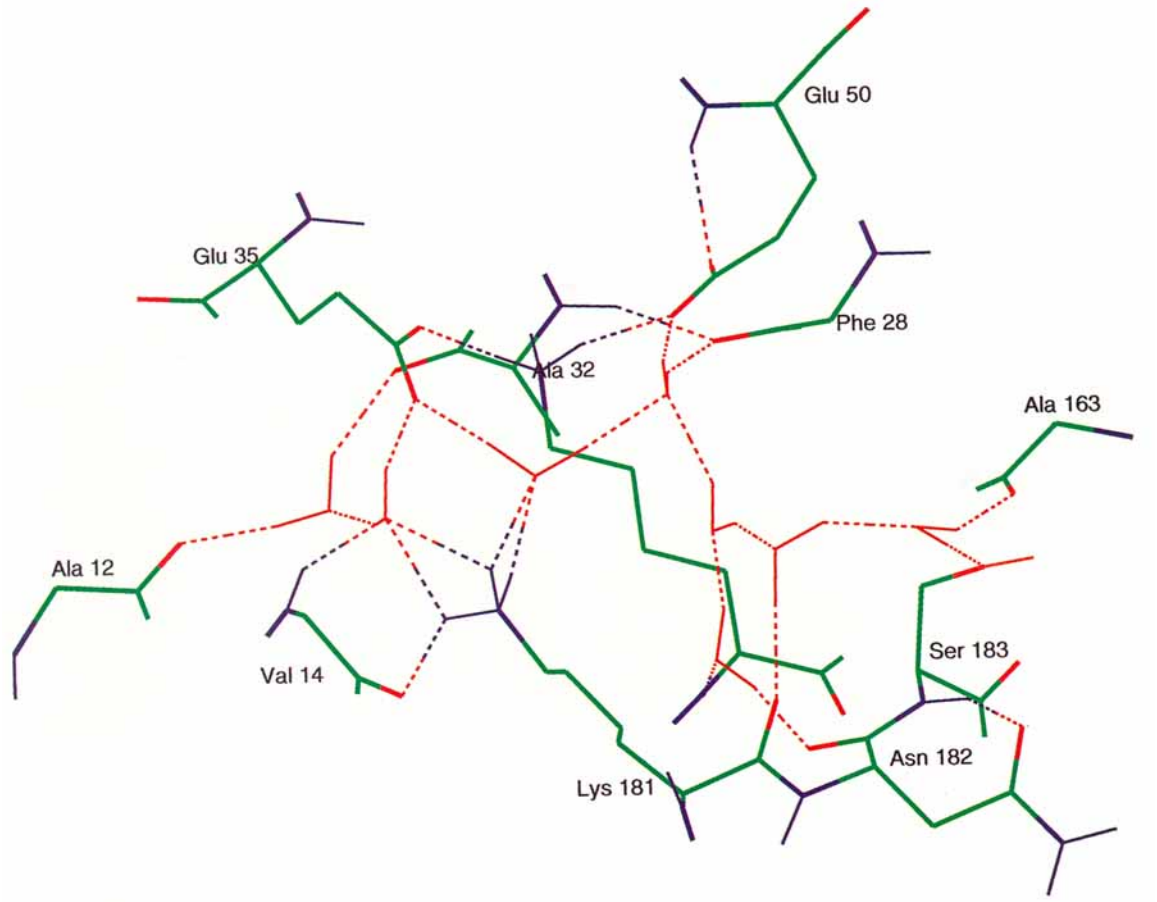

Fig. 13. Cluster of 8 buried water molecules in the actinidin structure, with optimised hydrogen bonds indicated as dashed lines. The hydrogen atoms are placed in the same positions as manually assigned by Baker and Hubbard.[3]

goals to automate the process of data entry and validation in protein structure databases.

We thank Dr. R. Wierenga and Dr. V. Lamzin for productive discussions.

REFERENCES

1. Pauling, L., Corey, R.B., Branson, H.R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Natl. Acad. Sci. U.S.A. 37: 205–211, 1951.
2. Pauling, L., Corey, R.B. Configurations of polypeptide chains with favored conformations around single bonds: Two new pleated sheets. Proc. Natl. Acad. Sci. U.S.A. 37: 729–740, 1951.
3. Baker, E.N., Hubbard, R.E. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44:97–179, 1984.
4. Creighton, T.E. Stability of folded conformations. Curr. Opin. Struct. Biol. 1:5–16, 1991.
5. Branden, C., Tooze, J. "Introduction to Protein Structure." New York: Garland, 1991.
6. Mc Donald, I.K., Thornton, J.M. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 238:777–793, 1994.
7. Brunger, A.T., Kuriyan, J., Karplus, M. Crystallographic r factor refinement by molecular dynamics. Science 235: 458–460, 1987.
8. Tronrud, D.E., ten Eyck, L.F., Matthews, B.W. An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Crystallogr. A 43:489–501, 1987.
9. Mc Donald, I.K., Thornton, J.M. The application of hydrogen bonding analysis in x-ray crystallography to help orientate asparagine, glutamine and histidine side chains. Protein Eng. 8:217–224, 1994.
10. Bass, M.B., Hopkins, D.F., Jaquysh, W.A.N., Ornstein, R.L. A method for determining the positions of polar hydrogens added to a protein structure that maximizes protein hydrogen bonding. Proteins 12:266–277, 1992.
11. Hooft, R.W.W., Kanters, J., Kroon, J. Statistical evaluation of the hydrogen bond: Another view on C-H donors. Submitted.
12. Allen, F.H., Kennard, O., Taylor, R. Systematic analysis of structural data as research technique in organic chemistry. Acc. Chem. Res. 16:146–153, 1983.
13. Finkelstein, A.V. Implications of the random characteristics of protein sequences for their three-dimensional structure. Curr. Opin. Struct. Biol. 4:422–428, 1994.
14. Read, R.J. Improved fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A 42:140–149, 1986.
15. Luzzati, V. Traitement statistique des erreurs dans la détermination de structures cristallines. Acta Crystallogr. 5:802–810, 1952.
16. Allinger, N.L. MM2, Quantum Chemistry Program Exchange no. 423. Chemistry Department, Indiana University. 1982.
17. Chakrabarti, P. Geometry of interaction of metal ions with

histidine residues in protein structures. Protein Eng. 4:57–63, 1990.
18. Wierenga, R.K., Noble, M.E.M., Vriend, G., Nauche, S., Hol, W.G.J. Refined 1.83 Angstroms structure of trypanosomal triosephosphate isomerase, crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. J. Mol. Biol. 220:995, 1991.
19. Dueck, G., Scheuer, T. Threshold accepting: A general purpose optimization algorithm superior to simulated annealing. J. Comput. Phys. 90:161–175, 1990.
20. Dueck, G. New optimization heuristics: The great deluge algorithm and the record-to-record travel. J. Comput. Phys. 104:86–92, 1993.
21. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macro-molecular structures. J. Mol. Biol. 112:535–542, 1977.
22. Hooft, R.W.W., Vriend, G., Sander, C. Reconstruction of symmetry related molecules from protein data bank (PDB) files. J. Appl. Crystallogr. 27:1006–1009, 1994.
23. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen bond and geometrical features. Biopolymers 22:2577–2637, 1983.
24. Baker, E.N., Dodson, E.J. Crystallographic refinement of the structure of actinidin at 1.7 Angstroms resolution by fast Fourier least-squares methods. Acta Crystallogr. A 36:559, 1980.
25. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8 Å resolution. J. Mol. Biol. 163:299–361, 1983.
26. Vriend, G., Sander, C. Quality control of protein models: directional atomic contact analysis. J. Appl. Crystallogr. 26:47–60, 1993.
27. van Aalten, D. to be submitted.
28. van Gunsteren, W., Berendsen, H.J.C. GROMOS, Groningen molecular simulation computer package. University of Groningen, The Netherlands, 1987.