Journal of Computer-Aided Molecular Design, 5 (1991) 5–20
ESCOM

J-CAMD 100

The inclusion of electrostatic hydration energies in molecular mechanics calculations

Michael K. Gilson and Barry Honig*

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, NY 10032, U.S.A.

Received 5 June 1990
Accepted 30 June 1990

Key words: Force fields; Solvation energies

SUMMARY

The problem of including solvent effects in molecular mechanics calculations is discussed. It is argued that the neglect of charge–solvent (solvation) interactions can introduce significant errors. The finite difference Poisson–Boltzmann (FDPB) method for calculating electrostatic interactions is summarized and is used as a basis for introducing a new pairwise energy term which accounts for charge–solvent interactions. This term acts between all pairs of atoms usually considered in molecular mechanics calculations and can be easily incorporated into existing force fields. As an example, a parameterization is developed for the CHARMm force field and the results compared to the predictions of the FDPB method. An approach to the realistic incorporation of solvent screening into force fields is also outlined.

INTRODUCTION

Molecular mechanics calculations involve the use of a potential energy function which yields a molecule's energy as a function of conformation. The extent to which a calculation simulates physical reality thus depends critically on the accuracy of the energy function that is used. Standard energy functions typically include terms for the stretching, bending, and twisting of chemical bonds, as well as terms which describe van der Waals and electrostatic interactions between nonbonded atoms. For molecules in solution, the calculations are complicated by the fact that solvent interactions can have profound effects on conformational energies.

In the past few years, it has become common practice to account for solvent effects by including solvent molecules directly in the molecular mechanics force field. In this way, microscopic properties of the solvent are accounted for explicitly. However, the calculations tend to be computation-

*To whom correspondence should be addressed.

ally demanding, particularly for electrostatic interactions, whose long range implies that a large number of solvent molecules must be included in the simulation system. Moreover, most available potential functions neglect electronic polarizability which can make large energetic contributions, particularly to interactions involving charged groups. A proper treatment of electronic polarizability will inevitably increase the computational demands of solvent simulations. For these reasons, it would be valuable to have a potential function which accounts for solvent effects without requiring the explicit description of solvent molecules.

Such a function should be particularly useful in the analysis of conformational energies and in energy minimizations, since in this type of calculation the explicit inclusion of solvent molecules offers little advantage. This is because thermodynamic properties, such as dielectric behavior, cannot be reproduced when the thermal motion of individual solvent molecules is neglected. Rather, when calculating the conformational energy of a solute molecule it is more appropriate to describe the solute–solvent interactions in terms of an averaged property, such as a dielectric constant.

In the past few years, we have applied the Poisson equation to molecules in solution in an attempt to develop a physically rigorous description of electrostatic interactions in such systems [1–3]. In the next section, we briefly summarize the theoretical and computational basis of our methodology and discuss the implications of our findings to date for the development of improved force fields. We then outline a new approach to the inclusion of electrostatic hydration energies in molecular force fields.

Hydration energies of charged groups can make enormous contributions to conformational energies. For example, the energetic cost of even partially removing a charged group from water can be tens of kcal/mol [4, 5]. A related problem arises in considering hydrogen bonding. Intramolecular hydrogen bonding energies are well described in most force fields yet there is no term that accounts for the fact that the donor and acceptor groups can make hydrogen bonds with solvent molecules of approximately comparable strength. Consequently, in comparing different conformations, it is possible to conclude erroneously that a particular structure has the lowest energy simply because it contains the largest number of intramolecular hydrogen bonds.

The first step in developing the new electrostatic hydration term is to make the general observation that the interaction of a solute atom with the solvent can be described as an interaction of opposite sign with the other atoms of the solute molecule. For example, an atom which interacts favorably with solvent can be thought of as being repelled by all other solute atoms, since they displace solvent. An atom which interacts unfavorably with the solvent will be attracted to other atoms of the solute. The physical basis for the electrostatic hydration term developed in the present paper is that the interaction of a charged atom with the high dielectric solvent can be viewed as a charge-induced dipole interaction with the bulk solvent. Consequently, the interaction of a charged atom with another atom that displaces solvent has a $1/r^4$ distance dependence. We derive a simple pairwise expression for the hydration energy which has this distance dependence and suggest specific parameters for the electrostatic hydration terms of charged amino acid side chains and of peptide groups. The new hydration energy term may be incorporated directly into molecular mechanics force fields and thus can be used in energy minimizations, as well as in the rapid evaluation of energies in conformational search procedures. The present paper also sketches an approach to improving the treatment of solvent screening of charge–charge interactions.

The actual calculations reported in this paper refer specifically to proteins. However, the methodology is quite general and parameters could easily be derived for small molecules as well as for other macromolecules.

METHODS

Finite difference Poisson-Boltzmann calculations

The starting point for our discussion is the Poisson equation

$$\nabla \cdot \varepsilon \nabla \varphi = -4\pi\rho \tag{1}$$

where φ is the electrical potential and ρ describes the charge distribution. φ, ε, and ρ vary throughout space. Coulomb's law is a special case solution to the Poisson equation for a homogenous isotropic medium and point charges. When electrolytes are present in solution, a $\sinh\varphi$ term appears in Eq. 1 which then becomes the Poisson–Boltzmann (PB) equation. In this work, we will ignore all ionic strength effects. In the past few years, we have shown how the PB equation can be applied to a variety of electrostatic problems, both by developing a numerical method of solving the equation for complex systems and by relating atomic properties such as radii, charges and polarizability to the spatial variables ε and ρ [1–3].

The numerical method involves partitioning space into a 3-D lattice which makes it possible to solve the PB equation with a finite difference algorithm (the FDPB method). We have developed a package of programs, DelPhi, which takes molecular information (atomic coordinates, charges, etc.) as input, and outputs values of the electrical potential both inside the molecule and in the surrounding solvent. In contrast to previous versions, the current algorithm is quite fast yielding results in a few minutes of μVAX CPU time [6]. The numerical algorithms, as well as an analysis of the validity of the equation when applied at the molecular level, are discussed in a number of recent publications [1–3, 7].

In our work to date, every lattice point lying within a solute molecule is assigned a molecular dielectric constant. In applications to conformational analysis, the molecular dielectric constant is usually 2, which accounts for electronic polarizability [4, 8]. In fact, assigning a dielectric constant of 2 is the physical equivalent of assigning a point polarizability to each atom. Both are approximations and it is not clear at this point which is the more accurate (see Refs. 7 and 8 for a more detailed discussion). In most applications, including those in the present work, the solvent is water so that all lattice points outside the molecule are assigned a dielectric constant of 80. Although boundary waters, especially near ions, are often believed to have a much lower dielectric constant than bulk, we have recently carried out Monte-Carlo simulations suggesting that the dielectric constant remains high, even in the first hydration shell [9]. As is standard in molecular mechanics calculations, all charges are initially placed at atomic nuclei, although in the FDPB method, they are then spread over neighboring lattice points as part of the numerical algorithm [2].

The FDPB method has been quite successful in reproducing experimental observations on the electrostatic properties of proteins [10–13]. A number of important generalizations have emerged from the studies that have been carried out so far. (1) The shape of the molecule as well as its charge distribution can have profound effects on electrical potentials. (2) Electric field lines will prefer to go 'through solvent' rather than 'through molecule'. Thus, even if a low dielectric region separates two charges in a molecule, the effective dielectric constant for their interaction can be high because the field lines will pass primarily through solvent. (3) Effective dielectric constants are sensitive functions of the position of the two charges in the molecule, of the shape of the molecule and of the distance separating the two charges [14]. The fact that no single dielectric constant is appropriate for all pairs of interacting atoms creates significant difficulties in molecular mechanics calculations.

The FDPB method has been extended to the calculation of hydration energies of molecules ranging in size from simple ions to large proteins [5]. For small ions where experimental data are available, the FDPB results are in good agreement with experiment and appear to be more accurate than free energy simulations [9]. In applications to proteins, it was found that the hydration energies of ionizable groups make a large contribution to protein stability. It is worth emphasizing in this regard that hydration as used here implies any interaction with water. An isolated ion in water is 'fully' hydrated. A charge 4 Å below the surface of a protein is partially hydrated because it can still interact with water, although it is not in direct contact. An ion is fully dehydrated only in the hypothetical case where it is infinitely removed from the aqueous phase.

Functional form of the electrostatic hydration term

The proposed electrostatic hydration energy term estimates the electrostatic interaction of a charged atom with the polarizable dielectric medium surrounding it as a function of the conformation of the molecule to which the atom belongs. A theoretical expression for the term may be derived from the following argument. The electrostatic energy density around an atom i which bears charge q_i in a solvent of dielectric constant ε_s is given by continuum electrostatic theory as

$$d\Delta G^o/dV = (332/8\pi\varepsilon_s)q_i^2/r^4 \tag{2}$$

where r is the distance of volume element dV from the charge, and the units are proton-charges, Ångstroms, (Å), and kcal/mol. This equation suggests that if an element of solvent is replaced by an atom j having internal dielectric constant ε_m (molecular) and volume V_j, the electrostatic energy of the system will rise by an amount:

$$\Delta\Delta G^o = (332/8\pi)q_i^2/r_{ij}^4 V_j(1/\varepsilon_m - 1/\varepsilon_s) \tag{3}$$

In essence, the contribution to the total electrostatic energy coming from volume V_j becomes more positive when it is filled with a low dielectric atom, an effect which results in a repulsive interaction between atoms i and j.

Equation 3 is only an approximation to the predictions of continuum electrostatic theory, since it assumes that the electrostatic field of charged atom i is not distorted by the low dielectric region represented by atom j. In fact, however, the field lines will tend to curve around atom j and to stay in the high dielectric solvent. When more atoms are present, more complicated field distortions will occur. It should also be pointed out that Eq. 3 does not account for the influence of a dissolved electrolyte on hydration energies. However, this influence is insignificant compared with the overwhelming effect of the high dielectric solvent [5].

The dielectric constant assigned to each atom need account only for electronic polarizability since molecular mechanics calculations treat dipolar reorientation explicitly. On this basis, ε_m is assigned a dielectric constant of 2 [8] while for water, $\varepsilon_s \simeq 80$. With these dielectric constants, the change in free energy on bringing atom j up to atom i is positive. This makes sense because j partially desolvates i by displacing a volume V_j Å^3 of high dielectric solvent. Similar reasoning has been used as the basis of a model for salting out of organic molecules in solution [15].

The repulsion described in Eq. 3 will act between each charged atom and each other solvent-excluding atom of a macromolecule. Thus, the work of removing atom i from a hypothetical location where it is fully solvated and placing it in a molecule where it is close to other atoms, j, is given by:

$$\Delta\Delta G_i^o = (332/8\pi)(1/\varepsilon_m - 1/\varepsilon_s)q_i^2 \cdot \{\sum_{j \neq i}^{N} V_j(1/r_{ij}^4) - \sum_{l \neq i}^{M} V_l(1/r_{il}^4)\} \quad (4)$$

where N is the number of solvent-displacing atoms in the molecule being considered and the sum is over atoms j in that molecule. The sum over the M atoms in the fully hydrated state, indexed with 'l', is zero in the present context since the fully hydrated atom should be infinitely separated from all other atoms (M = 1). However, as discussed below, for groups which consist of more than one atom, there are cases where it is convenient to choose a hydrated reference state in which atom i belongs to a group of atoms that are treated as a unit. Thus, atom i might be an oxygen in a carboxylate group. In such a case, the sum over the M atoms in the hydrated state will be non-zero since atom i will always be close to one or more neighboring atoms (M > 1).

Equation 3 yields free energy. Its derivative with respect to distance, r_{ij}, gives a force acting between charged atom i and the dehydrating atom j. Thus,

$$F_{ij} = (-332/2\pi)V_j q_i^2/r_{ij}^5(1/\varepsilon_m - 1/\varepsilon_s) \quad (5)$$

Note that this force must act not only on the charged atom i, but in an equal and opposite manner on atom j. If atom j happens to be charged as well, the net hydration repulsion acting between i and j will be the sum of F_{ij} and F_{ji}. This repulsive force presumably may be treated as additive with the electrostatic charge–charge interaction which will act between the two charged atoms.

In evaluating the hydration term for atoms in a single molecule, it is preferable to neglect the interactions of atom i with its bonding partners (1–2 interactions), and with their bonding partners (1–3 interactions). This prevents confusion of the hydration force with bond-stretching and bending force constants. In this work, we do include 1–4 interactions because significant hydration changes could result from changes in dihedral angles. For simplicity, we have assumed a united atom representation so that the volume occupied by hydrogen atoms is not considered explicitly. Thus, atom j in the above equations will never represent a hydrogen. However, we do treat polar hydrogens as charged atoms subject to dehydration by other atoms, so atom i may be a hydrogen.

Parameterization for single atoms

It might appear at first that a parameterization of the electrostatic hydration term is unnecessary, because all the values in Eq. 3 are well defined. However, several problems arise. For one thing, although the volume of atom j could be defined in terms of its van der Waals surface, in fact, the volume of interest is the volume of solvent effectively displaced by the atom, which depends not only on the atom's van der Waals volume, but also on the details of atomic packing in the macromolecule. In addition, the electrostatic hydration term is an approximation, as noted above, and obtaining a good fit to either continuum theory, or experimental values, requires a parameterization.

If we assume that all atoms j are of equal effective volume, V_j can be removed from the sum and incorporated into a constant so that Eq. 4 can be written in the form:

$$\Delta\Delta G_i^o = A_i q_i^2 \{S_i^{mol} - S_i^{hyd}\} \quad (6)$$

where $S_i = \sum_{j \neq i}^{N} 1/r_{ij}^4$ and the superscript mol refers to the first sum in Eq. 4 and indicates that atom i forms part of a molecule. The superscript 'hyd' refers to the second sum in Eq. 4 and denotes the hydrated state. A_i will be given by Eq. 4 as $(332/8\pi)V_j(1/\varepsilon_m - 1/\varepsilon_s)$. Choosing $V_j = 18$ Å^3 [based on

the average packing density of lysozyme (see below)], $\varepsilon_m = 2$, and $\varepsilon_s = 80$ yields $A_i = 120$ kcal/mol/proton^2/Å^4. However, A_i may also be regarded as a parameter which can be determined in a number of other ways.

In this section, we obtain values for A_i by setting $\Delta\Delta G_i^\circ$ equal to the change in solvation energy, $\Delta\Delta G_i^\circ$(solv), for the case where atom i is removed from water and placed in a medium of dielectric constant equal to ε_m. This medium may be thought of as a hypothetical, infinitely large macromolecule. $\Delta\Delta G_i^\circ$(solv) can be deduced from experimental quantities or from FDPB calculations. If S_i^{deh} and S_i^{hyd} represent the values of S_i for the maximally dehydrated and hydrated states of atom i of a specific type, we have that

$$\Delta\Delta G_i^\circ(\text{solv}) = A_i\, q_i^2 \{S_i^{deh} - S_i^{hyd}\} \tag{7}$$

Equation 7 defines a two-point parameterization, which fixes A_i based on the properties of the fully hydrated and dehydrated states (a later section describes the evaluation of A_i based on partially dehydrated states). Note that the difference between Eqs. 6 and 7 is that Eq. 6 concerns a real molecule (hence the superscript 'mol') while Eq. 7 concerns an infinite medium where the atom is fully dehydrated (superscript 'deh'). The strategy in this section is to determine the A_i from Eq. 7 and then use these values in Eq. 6.

In order to determine the value of A_i, we need values for $\Delta\Delta G_i^\circ$(solv), S_i^{deh}, and S_i^{hyd}. (It should be noted that A_i will not depend on q_i, because $\Delta\Delta G_i^\circ$ is proportional to q_i^2, as is the right hand side of Eq. 7.) The precise definitions of the fully hydrated state and the fully dehydrated state are not important, so long as the same definitions are used in the evaluation of $\Delta\Delta G_i^\circ$(solv) and of S_i^{hyd} and S_i^{deh}. We define the fully hydrated state of a given atom type by requiring that it be separated from any atoms which could contribute to S_i. Therefore, it must be separated from the rest of the macromolecule except for its 1–2 and 1–3 bonding partners. With this definition, $S_i^{hyd} = 0$ since no atoms appear in the sum over j.

The fully dehydrated state has the atom buried in a hypothetical infinitely large macromolecule. S_i^{deh} then corresponds to an infinite sum over all atoms in this molecule. In practice, S_i^{deh} can be determined as follows. An atom i of the desired type is selected in a protein whose structure has been solved crystallographically. Contributions to S_i are first calculated for only those atoms that lie within b Å of the selected atom, where b is the depth of the selected atom i from the protein's surface. As described in Ref. 14, we determine atomic depths using the program MS [16], which defines the protein surface as the union of contact and reentrant surfaces. The resulting sum, termed S_i^{near}, will be sensitive to details of packing and bonding around atom i.

In order to estimate the contribution of all the atoms further than b Å from atom i, we note that the atoms in a spherical shell r Å from atom i and of thickness dr will contribute $4\pi N/r^2$ to S_i, where N is the number density, in atoms/Å^3, of non-hydrogen atoms in a folded protein. Thus, based on a simple integration over r, the atoms further than b Å from atom i in our hypothetical infinite protein will contribute $4\pi N/b$, making the value of $S_i^{deh} = S_i^{near} + 4\pi N/b$. It should be clear that the best estimates of S_i^{deh} will be obtained for more deeply buried atoms, for which the more approximate analytical part of S_i^{deh} will make a smaller contribution. We use $N = 0.056$ atoms/Å^3, based on the total volume [8] and the number of non-hydrogen atoms in hen egg white lysozyme.

We have used this method to calculate S_i^{deh} for every occurrence of the following atoms in the protein rhodanese [17]: aspartate Oδ1, Oδ2, Cγ, and Cβ; glutamate Oε1, Oε2, Cδ, and Cγ; lysine Hζ1, Hζ2, Hζ3, Nζ and Cε; arginine Hη11, Hη12, Hη21, Hη22, Nη1, Nη2, Cζ, Nε, Hε, and Cδ;

and main-chain peptide nitrogen. The technique used to generate the coordinates of the polar hydrogen atoms is described below. We find that the values of S_i^{deh} for those atoms at least 2 Å below the protein surface are remarkably uniform, even for the polar hydrogens, with a mean and standard deviation of 0.29 ± 0.04 Å^{-4}, and no clear cut differences between atom types. Thus, for simplicity, we use $S_i^{deh} = 0.3$ for every atom type. With $S_i^{hyd} = 0$, as noted above, Eq. 7 then becomes simply

$$A_i = \Delta\Delta G_i^o(\text{solv})/0.3 \tag{8}$$

As mentioned above, $\Delta\Delta G_i^o$(solv) is the energy required to transfer an atom from water to a medium of dielectric constant 2. It may be calculated by using the FDPB method [5] to compute the energy of transferring atom i together with its 1–2 and 1–3 bonding neighbors from a medium of dielectric constant 80 to a medium of dielectric constant 2.

Parameterization for groups

The treatment in the previous section dealt with individual atoms. *A further complication arises when considering groups of atoms because the electrostatic hydration energy of a group does not equal the summed hydration energies of the individual atoms.* One way of understanding this point is to note that the hydration state of a group of atoms affects not only the interactions of individual atoms with solvent, but also the strength of interactions between atoms making up the group. For example, consider the transfer of a dipolar group from water to vacuum, where the group consists of two spherical atoms of 2 Å radius, joined by a 3Å bond, and having ± 1 proton charge. The transfer free energy of each atom may be estimated by continuum theory as $(166/2)(1 - 1/80) \simeq 80$ kcal/mol. The change in atom–atom interaction energy may be estimated by Coulomb's law as $-(332/3)(1 - 1/80) \simeq 110$, so the net transfer energy equals $2(80) - 110 \simeq +50$ kcal/mol, rather than $\sim +160$ which would be obtained by summing the individual atomic transfer energies.

Simply stated, the solvation energy of a dipole is not equal to the sum of the solvation energies of the component charges. Similarly, the solvation energy of an ionized group is not equal to the sum of the solvation energies of the charged atoms making up that group. Because proteins are commonly described as consisting of groups of atoms (peptides, side chains, etc.), it is useful to develop a parameterization scheme appropriate to these groups.

This is possible if it is assumed that for a relatively rigid group of atoms, the hydration energy will be *proportional* (rather than *equal*) to the summed hydration energies of the individual atoms in the group. Then, in terms of the electrostatic hydration term, we have that

$$\Delta\Delta G_k^o = A_k \, \Sigma_i q_i^2 \{S_i^{mol} - s_i^{hyd}\} \tag{9}$$

where $\Delta\Delta G_k^o$ is the charge–solvent interaction energy given by the electrostatic hydration term for a group k of a given type, A_k is the desired parameter for a group k of the selected type; and q_i is the charge of atom i forming part of the group. The sum runs over all atoms making up the group. Note that Eq. 9 is very similar to Eq. 6, the primary difference being that Eq. 9 involves a sum over all atoms in the group.

The parameterization for groups is established by requiring that

$$\Delta\Delta G_k^o(\text{solv}) = A_k \, \Sigma_i \, q_i^2 \{S_i^{deh} - S_i^{hyd}\} \tag{10}$$

where $\Delta\Delta G_k^o$(solv), the hydration energy of a chemical group of type k, may be obtained either by

the FDPB method or from experimental data. For individual atoms (see above), we set $S_i^{deh} = 0.3$ $Å^{-4}$ for all atom types while S_i^{hyd} was, by definition, zero. For groups S_i^{deh} remains 0.3, but S_i^{hyd} must now be determined for the fully hydrated group, rather than for the fully hydrated atom i, so that S_i^{hyd} need no longer be zero. The fully hydrated group may be defined as the charge-bearing atoms making up the group, along with all their 1–2 and 1–3 bonding neighbors, separated from the rest of the macromolecule and immersed in water. For example, for the CHARMm charge set [18], which places charge on aspartate side chain atoms Oδ1, Oδ2, Cγ and Cβ, the fully hydrated group includes atoms N, Cα, C, Cβ, Cγ, Oδ1 and Oδ2. With this definition, it should be clear that S_i^{hyd} need not equal zero when one is considering a group, because some atoms (j) in the fully hydrated group will be 1–4 or even more distant bonding neighbors to atom i of the group, and will therefore contribute to S_i. It should also be noted that the values of $\Sigma_i(q_i^2 S_i^{hyd})$ and $\Sigma_i(q_i^2 S_i^{deh})$, and thus of A_k, will depend on the charge set used. Values of these sums, based on the CHARMm charge set and on the atomic coordinates of rhodanese [17], are compiled in Table 1 for aspartate, glutamate, lysine and arginine side chains.

Although the left hand side of Eq. 10 may be calculated by the FDPB method, incorporating an experimentally derived energy reduces the uncertainty in the parameterization, because it guarantees that the electrostatic hydration term will yield the best possible estimate of the energy of dehydrating a charged group. Taking the dielectric constant of a macromolecular interior as approximately 2 (see above), and with the dielectric constant of vacuum equal to 1, continuum dielectric theory predicts [9] that the work of dehydrating a charged group by burying it inside a macromolecule will be very close to half the work of dehydrating it by transferring it to vacuum. Thus $\Delta\Delta G_k^o(solv)$ may be set equal to half the vacuum to water transfer energies of appropriate model compounds representing the ionized side chains. Experimentally determined vacuum to water transfer energies for the model compounds acetate, methylammonium and guanidinium are approximately −80 [5], −70 [5], and −50 kcal/mol (B. Honig, unpublished results), respectively. These values permit parameterization of the electrostatic hydration term for the side chains of aspartate and glutamate, lysine and asparagine, respectively. Similarly, experimentally based hydration energies make it possible to obtain parameters for the carbonyl and amino groups in the peptide backbone (see below).

TABLE 1

PARAMETERS FOR IONIZED SIDE CHAINS BASED ON EXPERIMENTAL DATA

Side chain	Hydration energy (kcal/mol)	$\sum_i^M (q_i^2 S_i^{deh})$	$\sum_i^M (q_i^2 S_i^{hyd})$	A_k
Aspartate	−40	0.26	0.02	170
Glutamate	−40	0.26	0.01	160
Arginine	−25	0.42	0.04	66
Lysine	−35	0.16	0.01	230

Hydration energies for transfer from medium of dielectric 2 to water estimated from experiment (see text). Sums in proton^2/Å^4. Symbols are defined in text.

Comparison with continuum calculations

The parameterization method presented above is based only on the extreme cases of fully hydrated and dehydrated charges. However, real applications of the method will involve cases where the charges are located in some molecule and are, therefore, not fully dehydrated. That is, the relevant sums over $1/r^4$ will be S_i^{mol} (Eq. 6) rather than S_i^{deh} (Eq. 7). As a consistency check of the methodology, we have calculated the electrostatic hydration term for atoms and groups in proteins of known structure and compared the values to the energies obtained from FDPB calculations for the transfer of the fully hydrated atoms or groups to their actual locations in the proteins. The FDPB values are then plotted as a function of $S_i^{mol} - S_i^{hyd}$. This makes it possible to determine whether the $1/r^4$ functional form together with a simple parameterization scheme based on solvation energies succeeds in reproducing the results of continuum calculations which provide the physical justification of the method.

Comparisons are made for atom types aspartate Oδ1, arginine Nη1, and the main-chain peptide N, and for the aspartate side chain as a whole. The comparisons use all the aspartates and arginines in rhodanese, and the peptide Ns of residues 29, 86, 111, 113, 116, and 249 of rhodanese [17] and residues 7, 13, 17, 23, 27, 35, 40, and 46 of the small globular protein crambin [19] in its native conformation. For simplicity, the calculations for single atoms all assume a unit charge on each atom examined, but energies for arbitrary q_i can be obtained by multiplying the energies by q_i^2.

Atomic and molecular parameters

The three-dimensional structures of rhodanese [17] and crambin [19] were obtained from the Brookhaven Protein Data Bank [20]. Coordinates were taken from files 1RHD and 1CRN. 'United atom' van der Waals radii [21] and a solvent probe radius of 1.4 Å determined the molecular surfaces used in calculating atomic depths and in the FDPB calculations. The atomic charge set of parameter set 19 of the molecular modelling program CHARMm [18] was used in the studies of ionized side chains. The coordinates of polar hydrogen atoms, required for the parameterization of lysine and arginine side chains, were computed by means of the CHARMm HBUILD command, followed by 200 steps of conjugate gradient energy minimization with only hydrogens free to move, using default energy parameters.

RESULTS

Single atoms

The values of A_i for single atoms are determined by Eq. 8, based on hydration energies for the atom types Asp Oδ1, Arg Nη1, and main-chain peptide N, calculated by means of the FDPB method as described above. The FDPB hydration energies (medium of $\varepsilon = 2$ to water) are respectively -45, -43, and -37 kcal/mol, leading to values of A_i 150, 140 and 120 kcal/mol/proton^2/Å^4, respectively.

Figures 1A,B,C, present scatter plots of charge–solvent interaction energies calculated using the FDPB method versus corresponding values of S_i^{mol}, as described in the Methods section. The data are also listed numerically in Table 2. It is apparent that the hydration energies calculated using continuum theory tend to vary linearly with S_i^{mol}. Linear least-squares fits of the data yield correlation coefficients of 0.90, 0.92, and 0.82 for Asp Oδ1, Arg Nη1, and main-chain N, respectively,

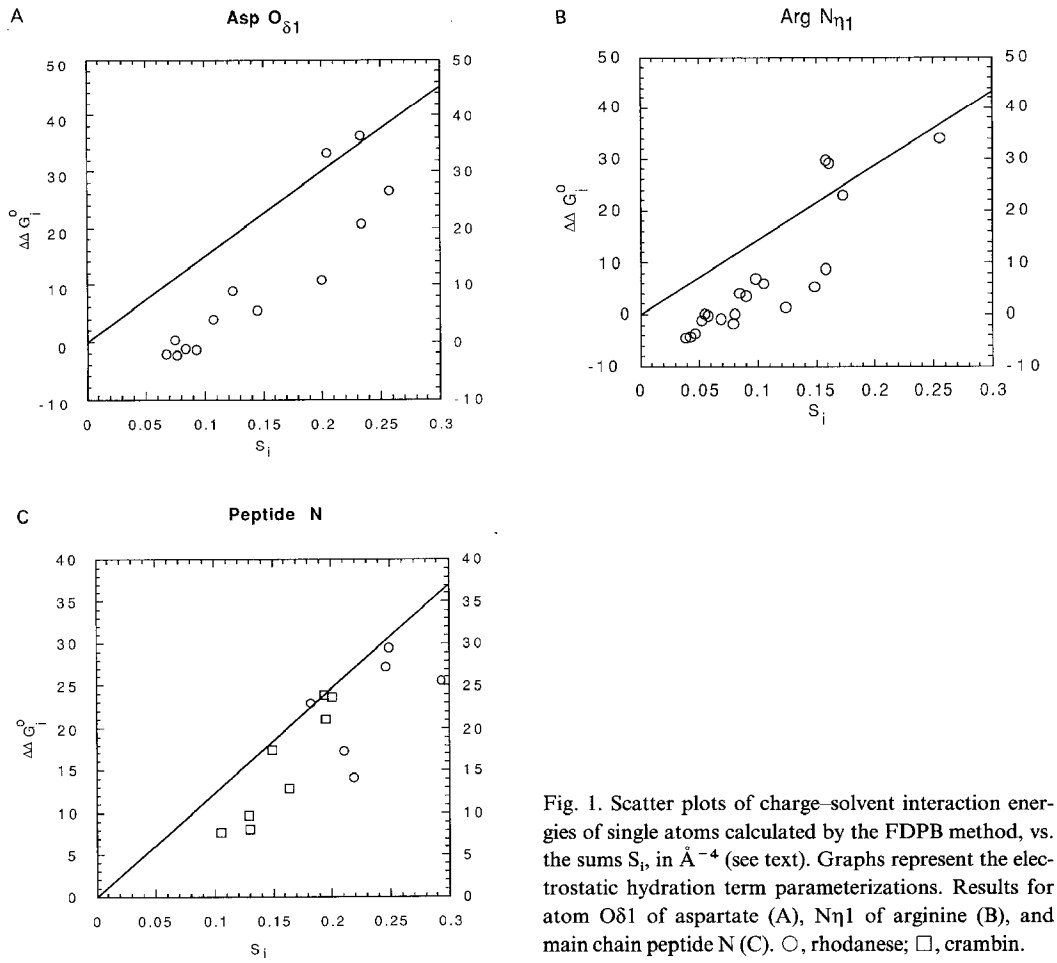

Fig. 1. Scatter plots of charge–solvent interaction energies of single atoms calculated by the FDPB method, vs. the sums S_i, in $Å^{-4}$ (see text). Graphs represent the electrostatic hydration term parameterizations. Results for atom Oδ1 of aspartate (A), Nη1 of arginine (B), and main chain peptide N (C). ○, rhodanese; □, crambin.

and corresponding standard deviations of the data from the linear fits are 6, 6 and 4 kcal/mol. The slopes of the linear least-squares fits of the scatter data of Fig. 1 represent alternative estimates of A_i for the three atom types. The values, 180, 190 and 120 kcal/mol/proton^2/Å^4, respectively, are similar to those obtained by the method described above based on Eq. 8. The linearity of the relationship between S_i^{mol} and hydration energies calculated using the FDPB method supports the validity of using the proposed electrostatic hydration term to approximate continuum results.

Figures 1A,B,C also include the straight lines defined by the parameterization based on solvation energies. The line is obtained by connecting the points corresponding to the fully hydrated state ($S_i^{mol} = S_i^{hyd} = 0$, $\Delta\Delta G_i^o = 0$) and the fully dehydrated state ($S_i^{mol} = S_i^{deh} = 0.3$, $\Delta\Delta G_i^o = \Delta\Delta G_i^o(\text{solv})$). Although the slopes of the lines are close to those defined by the scatter data (see above), the scatter data tend to fall below the lines. That is, the scatter data (corresponding to states of partial hydration) do not fall directly between the fully hydrated and dehydrated states as well as might be expected. This is likely to result from the fact, mentioned above, that the distortion of electrostatic field lines by dielectric boundaries is not accounted for in the electrostatic

TABLE 2

VALUES OF S_i AND ΔG_i^o(FDPB) FOR SELECTED ATOMS, AND OF $\sum_i^M (q_i^2 S_i)$ AND ΔG_k^o(FDPB) FOR ASPARTATE SIDE CHAINS IN RHODANESE

Residue	S_i	ΔG_i^o(FDPB)	Residue	S_i	ΔG_i^o(FDPB)
	Arginine Nη1			*Main chain N*	
				Rhodanese	
7	0.090	3.6	29	0.182	23.0
20	0.052	− 1.2	86	0.219	14.3
29	0.158	8.8	111	0.247	27.2
41	0.148	5.3	113	0.249	29.5
44	0.046	− 3.6	116	0.294	25.6
50	0.068	− 0.7	249	0.210	17.3
64	0.098	6.8			
110	0.158	29.9			
116	0.161	29.0			
121	0.080	0.3		*Crambin*	
131	0.084	4.2	7	0.131	8.1
146	0.079	− 1.6	13	0.201	23.6
158	0.058	− 0.1	17	0.195	21.1
175	0.039	− 4.5	23	0.129	9.6
182	0.256	34.2	27	0.194	23.9
186	0.124	1.5	35	0.149	17.4
205	0.055	0.2	40	0.106	7.7
229	0.042	− 4.1	46	0.165	13.0
248	0.173	23.1			
281	0.105	6.0			
				$\sum_i^M (q_i^2 S_i)$	ΔG_k^o(FDPB)
	Aspartate Oδ1			*Aspartate group*	
32	0.234	20.7	32	0.150	13.5
59	0.124	8.8	59	0.070	2.5
65	0.200	10.7	65	0.101	3.4
82	0.077	− 2.4	82	0.040	− 1.7
92	0.074	0.3	92	0.032	− 1.8
101	0.257	26.6	101	0.142	11.0
102	0.067	− 2.2	102	0.036	− 1.4
180	0.232	36.4	180	0.164	28.1
195	0.093	− 1.3	195	0.059	2.0
200	0.074	0.3	200	0.059	2.3
214	0.084	− 1.1	214	0.063	2.2
238	0.145	5.5	238	0.081	3.2
267	0.107	4.1	267	0.079	3.5
272	0.204	33.1	272	0.133	19.0

Energies in kcal/mol. The occasional negative transfer energies may result from field distortions or from numerical error.

hydration term. This should result in an overestimate of hydration energies by the hydration term which, in contrast to the FDPB calculations, will not allow the electric field to arrange itself in space so as to minimize the free energy. If this explanation is correct, the energies provided by the hydration term will tend to be upper limits to the charge–solvent interactions obtained from the FDPB calculations.

The previous paragraph suggests that it might be more appropriate to obtain parameters from the scatter plots of real proteins rather than from solvation energies, as has been done in this work. However, since the slopes of the scatter data and the straight line defined by solvation energies are similar, the simplicity of the parameterization based on solvation energies suggests that it constitutes the preferable approach.

Group parameters

As described in the Methods section, it is possible to incorporate experimentally determined hydration energies of groups of atoms into the parameterization scheme. This approach guarantees that a molecular mechanics calculation will yield the best possible estimate of the energy cost of dehydrating each type of group. Table 1 provides estimated hydration energies for the side chains of aspartate, glutamate, arginine and lysine, and the resulting parameters A_k for the electrostatic hydration term appropriate to the CHARMm 19 charge set [18].

The ability of the electrostatic hydration term to reproduce continuum results for partially hydrated groups is examined in Fig. 2 for the case of aspartate. The solid line in Fig. 2 is based on the two-point parameterization of Eq. 10. Figure 2 shows that, as in the case of single atoms, the scatter data have an approximately linear relationship, with a correlation coefficient of 0.91 and standard deviation of 4 kcal/mol. This linearity is notable because it indicates that the neglect of

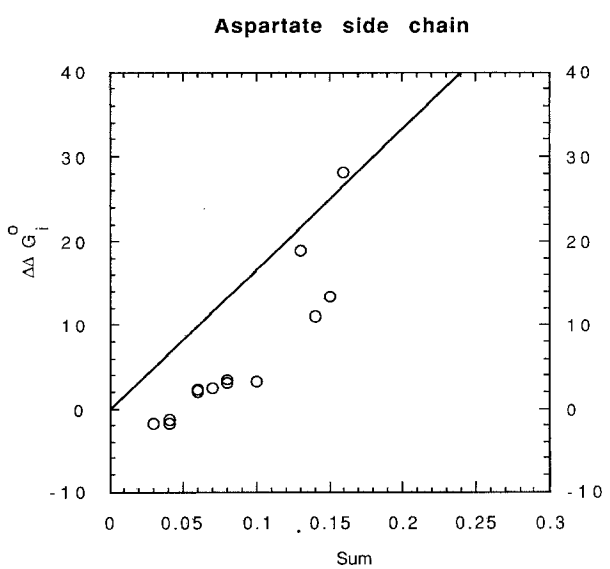

Fig. 2. Scatter plot of charge–solvent interaction energies of aspartate side chain calculated by FDPB method vs. the quantity, $\Sigma_i(q_i^2 S_i^{mol}) - \Sigma_i(q_i^2 S_i^{hyd})$ in proton^2/Å^4 (see text). Line represents electrostatic hydration term parameterization.

changes in intra-group interatomic interactions with changes in dielectric environment is not a bad approximation, at least for this small rigid ionized group. The best fit slope of 150 kcal/mol/proton^2/Å^4 is similar to the value of A_k obtained by the two-point parameterization using an experimentally based hydration energy (see Table 1). Also, as for the single atoms, the scatter points tend to fall somewhat below the parameterization line joining the end points, presumably for the same reasons. One disconcerting feature of the figure is the presence of a few weakly negative transfer energies for groups on the surface of the protein (Sum ~ 0). These are probably due to small numerical errors in the FDPB calculations but could also result from field distortions in crevices that increase solvation energies relative to the free amino acid.

It is possible to use the present approach to derive solvation parameters for dipolar groups as well. As an example, we consider the case of the main chain carbonyl and amide groups. The CHARMm 19 representation of the peptide backbone assigns partial charges of $+0.55$ and -0.55, respectively, to the carbonyl C and O, and charges of -0.35, $+0.25$ and $+0.1$, to the amino nitrogen, hydrogen and Cα, respectively. The vacuum to 'oil' (non-polar solvent) transfer energy of a carbonyl group has been estimated as 2 kcal/mol [22]. Although no number for an NH group is given we assume a value of 3 kcal/mol, which is somewhat less than the assigned value of the slightly more polar OH group [22]. Substitution of these values for $\Delta\Delta G^o(solv)$ into Eq. 10 (and recalling that $S_i^{hyd}=0$ for a group consisting of 2 or 3 bonded atoms) yields $A_{CO}=2/(0.3(0.55^2+0.55^2))=11$ kcal/mol/proton^2/A^4. Similarly, $A_{CNH}=3/(0.3(0.35^2+0.25^2+0.1^2))=51$ kcal/mol/proton^2/A^4. Note that these values for A_k reflect both the charge set and the estimate of the group solvation energy. It is clearly straightforward to derive different values of A_k, if a different charge set is used, or if another estimate of group transfer energies is used.

The sum of the transfer energies for an NH and a CO group as estimated here is 5 kcal/mol. It is worth noting that since many force fields yield hydrogen bond energies of approximately 5 kcal/mol, adding the electrostatic hydration term to such a force field would predict that the energy difference between the totally hydrated carbonyl and amino groups and the fully dehydrated NH...CO hydrogen bond would be about zero. The actual net energetic contribution of an intramolecular hydrogen bond is uncertain, but is likely to be on the order of 0–1 kcal/mol [23]. Given a particular partial charge set, it should be straightforward to 'tune' the value of A_k so that it yields any desired free energy contribution from a hydrogen bond. However, choosing an optimal hydrogen bonding energy and developing appropriate parameters is beyond the scope of this work. Our point here is that the electrostatic hydration term makes it possible to incorporate a realistic physical description of the competing factors involved in hydrogen bond formation, within the pairwise interaction formalism of molecular mechanics force fields.

Forces

Although the results of the previous sections have demonstrated that the electrostatic hydration term yields reasonable hydration energies, it must also yield reasonable forces if it is to be used in energy minimizations. The electrostatic hydration force repelling a charged atom from another atom will be given by

$$F_{ij} = -4\ A_i\ q_i^2/r_{ij}^5 \qquad (11)$$

With $A_i \sim 150$, the force between an atom i of unit charge and another atom j located 3 Å away will be -2.5 kcal/mol/Å. If atom j bears a charge of -1 protons, the total hydration force acting

between the two will be doubled. Thus, a total repulsive force of -5 kcal/mol/Å will act between the two charged atoms.

It is instructive to compare the magnitude of this force to that produced by the Coulombic attraction of two charges of opposite sign. The magnitude of this force will obviously depend on the effective dielectric constant, ε_{eff}, for the interaction. For two charges separated by 3Å and buried in the interior of a protein, $\varepsilon_{eff} \sim 2$, so there will be an attractive force of $332/2/3^2 = 18$ kcal/mol/Å which far outweighs the hydration repulsion of -5 kcal/mol/Å. This is quite reasonable since there will always be a strong attraction between two buried charges. Note that the *net* electrostatic hydration force on each of the charges may be much greater than -5 kcal/mol/Å, since they will be repelled by *all* other atoms, leading to a driving force towards the surface of the macromolecule. However, if the balance of forces is such that the charges will remain buried then they will tend to form a salt bridge.

For charges on the surface of a protein, ε_{eff} can be quite large and the attractive Coulombic interaction will be correspondingly reduced. By setting the Coulombic attractive force equal to the hydration repulsion, it is possible to calculate the equilibrium bond distance on an ion pair on the surface of a protein. Assuming $A_i \sim 150$ for both charges, we have (setting the hydration term in Eq. 11 equal to the Coulombic force) that $r = (600\varepsilon_{eff}/332)^{1/3}$. This predicts ionic bond lengths of 2.6 Å, 3.3 Å and 4.2 Å for $\varepsilon_{eff} = 10$, 20 and 40, respectively. Since these values for ε_{eff} are characteristic of those expected on the surface of proteins, it appears that the electrostatic hydration term when used in conjunction with an appropriate dielectric constant will produce ion pairs which have a reasonable geometry (see also below).

DISCUSSION

We have argued that the absence of a term which accounts for the hydration energy of charged and polar species is a serious deficiency in conventional force fields. A partial solution would be to use FDPB calculations to help distinguish between plausible conformations, but this would not solve the problem of incorporating electrostatic hydration effects directly into force fields. However, in this work we have shown that the electrostatic hydration term developed above serves as a good approximation to the results of the Poisson equation. Thus, we propose including the term

$$\Delta\Delta G_k^o = A_k \Sigma_i q_i^2 S_i^{mol} \qquad (12)$$

in existing force fields. The term is not exact, but, rather, constitutes a reasonable approximation to an important energetic contribution that is usually ignored. Note that Eq. 12 is identical to Eq. 9 except for the omission of S_i^{hyd}, which is used to determine parameters but need not be included in force fields because it is, in most cases, a constant with respect to conformation.

A set of parameters for ionizable groups and backbone dipoles appropriate to the CHARMm force field is given above. Parameters for other residues, as well as complete parameter sets appropriate to other force fields, can easily be obtained using the approach described in this work. The use of the hydration term should improve the performance of molecular mechanics calculations, which at this time generally neglect hydration effects. As parameterized here, the electrostatic hydration term will provide an energy penalty of several tens of kcal/mol for a structure which buries an ionized group, such as an aspartate side chain. Moreover, because ionized side chains will be repelled by the rest of the protein when the electrostatic hydration term is used, the term may diminish the tendency of energy minimizations to form overly compact protein structures.

Incorporating the hydration term will not by itself correct the treatment of hydrogen bonds and salt bridges. As discussed above, the balance between hydration and Coulombic energies will be sensitive to the effective dielectric constant assumed for charge–charge interactions. The effective dielectric constant for charges on the surface of proteins is quite high due to screening by the solvent, while deeply buried groups interact with an effective dielectric constant of about 2 [4, 14]. It is thus essential to develop force fields that incorporate this behavior. Most current molecular mechanics programs allow only a single value of the dielectric constant to be used at one time (or a single functional form such as $\varepsilon = r$) and values ranging from 1–4 are typically chosen. In fact, it is often necessary to choose a small value for ε if a reasonable hydrogen bond is to be obtained. This inevitably leads to unrealistically large electrostatic forces between charged groups on the protein surface. The only way to remedy this problem is to use a dielectric constant which varies as a function of distance from the surface.

This was first attempted by Northrup et al. [24], who scaled the magnitude of charges according to their depth from the surface. An alternative approach, which retains the simplicity of pairwise interaction energies, is to assume that the effective dielectric constant for interactions of charges i and j will be a function of S_i^{mol} and S_j^{mol}. If S_i^{mol} and S_j^{mol} are both small, ε_{eff} should be high since both charges are well hydrated. When S_i^{mol} and S_j^{mol} are equal to their maximum value of 0.3, ε_{eff} should be about 2, corresponding to deeply buried atoms. A least squares fit of a simple functional form of S_i^{mol} and S_j^{mol} (either a product or sum is probably adequate) to values of ε_{eff} obtained from FDPB calculations on pairs of atoms in proteins should yield a simple expression for effective dielectric constants.

It is of interest to compare the electrostatic hydration term with the hydration shell model of Scheraga and coworkers [25]. One important difference is that the latter is designed to account for both electrostatic and hydrophobic effects. However, it should be possible to develop a pairwise hydrophobic term to complement the electrostatic term presented in this work (see below). Another difference concerns the relative contribution of the first hydration layer to the total solvation energy. The hydration shell model assumes that charge–solvent interactions in the first hydration shell account for the total solvation energy. However, continuum theory predicts that waters beyond the first shell make a substantial contribution to the hydration energy. Consequently, the electrostatic hydration term, which is designed to mimic continuum theory, yields a substantially longer range interaction than does the hydration shell model. This, together with its simple functional form, suggests that the electrostatic hydration term should provide a useful alternative to the hydration shell model for charged atoms.

The general approach to treating solvent effects used in this work involves replacing the interactions of an atom with solvent by inverse interactions with atoms which displace solvent. This suggests a somewhat radical approach to conformational energy calculations in which only interactions that distinguish between protein and solvent are included in the potentials. For example, a simple way of treating hydrophobic interactions would be to introduce attractive forces between nonpolar atoms as a substitute for their unfavorable interactions with the solvent. With regard to dispersion forces – the $1/r^6$ term in the van der Waals interactions – the interactions among the atoms of a macromolecule should be similar in magnitude to those between the macromolecule and the solvent. It would thus appear to be valid, to first order at least, to completely ignore such interactions in conformational energy calculations, except perhaps in the final stages of structure optimization when close packing becomes important (see also the discussion by Novotny et al.

[26]). It might be of concern that neglecting dispersion forces, as suggested here, would prevent potential functions from producing protein conformations which have compact structures. However, this problem would probably be overcome by incorporating hydrophobic interactions in the calculations.

Although the suggestions of the previous paragraph are based on a physically reasonable description of solvent effects, they have not yet been implemented. In contrast, the electrostatic hydration term developed in this work can be applied to existing force fields and should provide a good estimate of an interaction that is generally neglected.

ACKNOWLEDGEMENTS

We are grateful to Drs. Richard Fine, Hillary S. Rodman Gilson, Cyrus Levinthal, Peter Shenkin, and Kim Sharp for participating in the discussions which stimulated this work. This work was supported by the NIH (GM-30518) and NSF (DMB88-05434).

REFERENCES

1 Klapper, I., Hagstrom, R., Fine, R., Sharp, K. and Honig, B., Proteins, 1 (1986) 47.
2 Gilson, M., Sharp, K. and Honig, B., J. Comp. Chem., 9 (1988) 327.
3 Jayaram, B., Sharp, K. and Honig, B., Biopolymers, 28 (1989) 975.
4 Gilson, M., Rashin, A., Fine, R. and Honig, B., J. Mol. Biol., 183 (1985) 503.
5 Gilson, M. and Honig, B., Proteins, 4 (1988) 7.
6 Nicholls, A. and Honig, B., J. Comp. Chem., (1991) in press.
7 Gilson, M. and Honig, B., Biopolymers, 25 (1986) 2097.
8 Sharp, K. and Honig, B., Annu. Rev. Biophys, Biophys. Chem., 19 (1990) 301.
9 Jayaram, B., Sharp, K., Fine, R. and Honig, B., J. Phys. Chem., 93 (1989) 4320.
10 Sharp, K., Fine, R. and Honig, B., Science, 236 (1987) 1460.
11 Gilson, M. and Honig, B., Nature, 330 (1987) 84.
12 Sternberg, M., Hayes, R., Russell, A., Thomas, P. and Fersht, A., Nature, 330 (1987) 86.
13 Dao-Pin, S., Liao, D. and Remington, S., Proc. Natl. Acad. Sci. U.S.A., 86 (1989) 5361.
14 Gilson, M. and Honig, B., Proteins, 3 (1988) 32.
15 See e.g. Edsall, J. and Wyman, J., Biophysical Chemistry, Vol. 1, Academic Press, New York, 1958, pp. 263–282.
16 Connolly, M., Science, 221 (1983) 709.
17 Ploegman, J., Drent, G., Kalk, K. and Hol, W., J. Mol. Biol., 123 (1978) 557.
18 Brooks, B., Bruccoleri, B., Olafson, B., States, D., Swaminathan, S., and Karplus, M., J. Comput. Chem., 4 (1983) 187.
19 Hendrickson, W. and Teeter, M., Nature, 290 (1981) 107.
20 Bernstein, F., Koetzle, T., Williams, G., Meyer, Jr., E., Bruice, M., Rodgers, J., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
21 McCammon, J., Wolynes, P. and Karplus, M., Biochemistry, 18 (1979) 927.
22 Engelman, D., Steitz, T. and Goldman, A., Annu. Rev. Biophys. Biophys. Chem., 15 (1986) 321.
23 Fersht, A., Jian-Ping, S., Knill-Jones, J., Lowe, D., Wilkinson, A., Blow, D., Brick, P., Carter, P., Waye, M. and Winter, G., Nature, 314 (1985) 235.
24 Northrup, S., Pear, M., Morgan, J., McCammon, J. and Karplus, M., J. Mol. Biol., 153 (1981) 1087.
25 Kang, Y., Nemethy, G. and Scheraga, H., J. Phys. Chem., 91 (1987) 4118.
26 Novotny, J., Bruccoleri, R. and Karplus, M., J. Mol. Biol., 177 (1984) 787.