Protein Science (1999), *8*:1982–1989. Cambridge University Press. Printed in the USA.

Local water bridges and protein conformational stability

MICHAEL PETUKHOV,[1] DAVID CREGUT,[1] CLÁUDIO M. SOARES,[2] AND LUIS SERRANO[1]

[1]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg D-69117, Germany
[2]Instituto de Tecnologia Química e Biológica, UNL, Apartado 127, 2780 Oeiras, Portugal

(RECEIVED March 3, 1999; ACCEPTED July 5, 1999)

Abstract

Recent studies have pointed out the important role of local water structures in protein conformational stability. Here, we present an accurate and computationally effective way to estimate the free energy contribution of the simplest water structure motif—the water bridge. Based on the combination of empirical parameters for accessible protein surface area and the explicit consideration of all possible water bridges with the protein, we introduce an improved protein solvation model. We find that accounting for water bridge formation in our model is essential to understand the conformational behavior of polypeptides in water. The model formulation, in fact, does not depend on the polypeptide nature of the solute and is therefore applicable to other flexible biomolecules (i.e., DNAs, RNAs, polysaccharides, etc.).

Keywords: local water structure; protein conformational stability; solvation potential

The interaction between a protein and stable local water structures has been the subject of many recent experimental and theoretical studies (Otting et al., 1991; Hummer et al., 1996; Israelachvili & Wennerstrom, 1996; Kovacs et al., 1997; Cheng & Rossky, 1998). Water–protein interactions were modeled using many theoretical approximations. From one hand explicit water molecules are considered (mainly in molecular dynamics simulations) (Kovacs et al., 1997; Bonvin et al., 1998; Cheng & Rossky, 1998). On the other hand, a variety of continuum approximation models based on both the accessible protein surface area (ASA) (Eisenberg & McLachlan, 1986; Ooi et al., 1987; Williams et al., 1992; von Freyberg et al., 1993) and several successful electrostatic models (Warshel & Russell, 1984; Sharp & Honig, 1990) are used. The last model (Warshel & Russell, 1984) considers solvent as a regular grid of polarizable dipoles. These dipoles are in the self-consistent electrostatic force field consisting of the protein dipoles and those of the surrounding solvent molecules. Therefore, this model can be considered as a bridge between the explicit water box and the continuum models.

The explicit water models have been proved to adequately account for protein solvation. However, they are extremely computationally demanding not only because of the drastic increase of interacting atoms in the models, but, even more importantly, they require long equilibration times for the water box itself to estimate the contribution of solvation to the free energy of each protein conformation. The electrostatic based models (Warshel & Russell, 1984; Sharp & Honig, 1990) are more computationally effective; however, they are still not fast enough to study unfolded states of proteins where many millions of very dissimilar conformations are to be considered.

On the other hand, ASA-based theoretical models are fast enough but, as well as all implicit models, suffer from the lack of atomic details and cannot account for stable local water structures near the protein surface. It is also important, in our view, to bear in mind there are cases where ASA-based solvation models were proved to be invalid. For instance, hydration free energy of some hydrocarbons significantly deviates from linear dependence on ASA (Ooi et al., 1987). In addition, several groups derived atomic solvation parameters (ASP) using almost identical sets of experimental data of hydration of small organic compounds. However, the selection of basic atom types and its van der Waal's radii differ from one model to another. As a result, the derived sets of ASPs significantly deviate from each other. While the calculations of protein hydration using some sets of ASP predict the unfolded state to be more stable than the folded, others yield precisely the opposite (Juffer et al., 1995). Nevertheless, it is our belief that ASA plays a major role in protein solvation and that ASA-based models can be corrected for all necessary effects to provide a simple and robust way to account for protein solvation.

An important aspect of protein solvation that has seldom been explicitly considered is the ability of water molecules to mediate hydrogen bond bridges with two polypeptide atoms (backbone–backbone or backbone–side chain), as frequently found in protein crystals (Thanki et al., 1990; Thanki et al., 1991; Morris et al., 1992). Water molecules can simultaneously donate and accept two hydrogen bonds, and therefore there are four possible types of water bridges that must be considered—namely, donor–donor (DD), acceptor–acceptor (AA), donor–acceptor (DA), and acceptor–donor (AD), as illustrated in Figure 1. Due to entropic reasons, it is intuitive to think that the formation of two hydrogen bonds

Reprint requests to: Michael Petukhov, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg D-69117, Germany; e-mail: Petukhov@EMBL-Heidelberg.de.

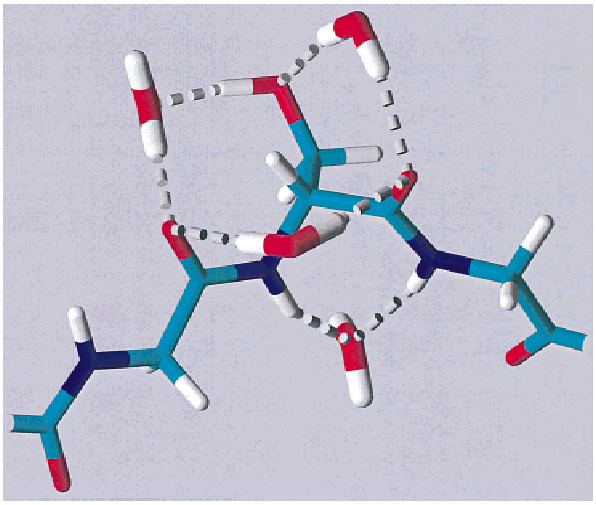

Fig. 1. An example of a cluster of several water bridges involving the side chain of a Ser residue in the central position of a Gly-Gly-X-Gly-Gly peptide.

between the protein and a water molecule is more favorable energetically than the formation of two hydrogen bonds with two different water molecules.

Results and discussion

Theory

The combinations of empirical force fields (particularly ECEPP) with an ASA-based solvation potential have been used to account for solvation in protein free energy calculations (Williams et al., 1992; von Freyberg et al., 1993). To incorporate water bridges in a simple and computationally effective ASA-based solvation model, we need to consider that (1) water bridges can form a complicated network of water–protein hydrogen bonds (see Fig. 1); (2) a competition of different water bridges for the same space is possible in some peptide conformations; and (3) in some cases water bridges can prevent the access of nearby protein atoms to interact with bulk water.

If one assumes for simplicity that there is only one water bridge in a given peptide conformation, then the conformational energy of a molecule (i.e., peptide), E_{conf}, including its solvation equals

$$E_{conf} = E_{pp} + \sum_{i \notin wb} \sigma_i ASA_i + P_{wb} G_{wb} + (1 - P_{wb}) G_{nwb} \qquad (1)$$

where E_{pp} is the energy of intrapeptide interactions, $\sum_{i \notin wb} \sigma_i ASA_i$ is the surface-based solvation of all protein atoms not affected by the presence of a water bridge, P_{wb} is the probability of water bridge formation, G_{wb} and G_{nwb} are solvation free energies resulting from the presence or absence, respectively, of a water bridge. Assuming that G_{nwb} is proportional to the solvent accessibility of the corresponding atoms (i.e., $G_{nwb} = \sum_{i \in wb} \sigma_i ASA_i$), and using the classical relation between the change of free energy of a two state chemical reaction and its equilibrium constant

$$\Delta G_{wb} = G_{wb} - G_{nwb} = -RT \times \ln\{P_{wb}/(1 - P_{wb})\}, \qquad (2)$$

we can simplify Equation 1:

$$E_{conf} = E_{pp} + \sum_i \sigma_i ASA_i - RT \times P_{wb} \ln[P_{wb}/(1 - P_{wb})] \qquad (3)$$

where R is the gas constant and T is the temperature in Kelvin. It is straightforward to generalize Equation 3 for any number of water bridges that do not overlap with others in a given peptide conformation. The water overlaps will be considered below.

Provided we know ΔG_{wb}, we could calculate the probability of the water bridge P_{wb} using the equivalent form of Equation 2:

$$P_{wb} = 1/(1 + e^{\Delta G_{wb}/RT}). \qquad (4)$$

The value of ΔG_{wb} includes, apart from the interaction energy of the water bridge with the peptide E_{wb}, a variety of different enthalpic and entropic effects such as loss of hydrogen bonding to bulk water, changes in entropy upon water bridge formation, and probably less energetically important distortions of local water structures around protein surface. Therefore, it can be represented as $\Delta G_{wb} = E_{wb} + C$, where C is an entity that includes all kind of enthalpic and entropic contributions except for water–protein nonbonded interactions. Formation of a water bridge implies the loss of two water–water hydrogen bonds [3.2 kcal/mol per hydrogen bond (Feyereisen et al., 1996)] and the entropy loss of fixing a water molecule [0.92 kcal/mol at room temperature (Franks, 1982)]. This results in a value for C of ~7.4 kcal/mol. Provided that we include all context depending effects associated with desolvation of the water molecule in E_{wb}, we can consider C as an approximate constant. Thus,

$$P_{wb} = 1/(1 + e^{(E_{wb}+C)/RT}). \qquad (5)$$

The value of E_{wb} can be effectively calculated with a standard force field for a particular water bridge in a given peptide conformation (Momany et al., 1975; Nemethy et al., 1983). Essentially, using standard geometric criteria for hydrogen bond formation between water and protein groups, we can identify all possible water bridges for any particular conformation of a molecule (see Materials and methods). Knowing C, we can computationally effectively estimate P_{wb} and ΔG_{wb}. Alternatively, provided that we know P_{wb} and E_{wb}, we can determine C and test the above hypothesis. P_{wb} can, in principle, be estimated from molecular dynamic (MD) simulations for a particular fixed peptide conformation. Comparison of the P_{wb} values determined from MD and those obtained using Equation 5 with various C values allows us to estimate C and to test its context independence (see below).

In general, for a particular solute, it is possible that there could be more than one water bridge and they can compete for the same space position. Assuming we can separately calculate the energy of water–proteins interactions E_{wb} for all water bridges of an overlapped area and its weighted average, $E^{aver}_{overlap}$, the probability of having a water bridge in the overlapped area is given by Equation 6:

$$P_{overlap} = 1/(1 + e^{(E^{aver}_{overlap}+C)/RT}). \qquad (6)$$

This probability can be used to calculate the average free energy contribution of water bridges in an overlapped area in the same way, as in Equation 3. Finally, the conformational energy of a peptide including solvation can be expressed by Equation 7:

$$E_{conf} = E_{pp} + \sum_i \sigma_i ASA_i - RT \times \sum_{iwb \notin overlap} P_{iwb} \ln[P_{iwb}/(1 - P_{iwb})] - RT \times \sum_{ioverlap} P_{ioverlap} \ln[P_{ioverlap}/(1 - P_{ioverlap})]. \quad (7)$$

We assume here that the probability of nonoverlapping water bridges to interact is quite small, and therefore they are considered to be independent.

It is worth noting that since our model accounts for all possible water bridges with two water–protein hydrogen bonds, the rare but possible cases of three and four water–protein hydrogen bonds are included automatically. This is simply because they present a subset of the water bridges with two hydrogen bonds. The proposed scheme for free energy calculations of water bridges also automatically accounts for more than two coordinated water bridges. This is because the final energy E_{wb} results from energy minimization using ECEPP force field with the list of all possible hydrogen bonds, and those water bridge hydrogen bonds that are not satisfied with protein or bulk water are penalized (see Materials and methods). Only those waters that have none or one hydrogen bonds to the protein are accounted for in the ASA-based part of the model.

Agreement with experimental data

There are several systems in which the above formulation can be tested, but one of the simplest ones for which a complete nuclear magnetic resonance (NMR) description is available (Bundi & Wüthrich, 1979; Fiebig et al., 1996; Plaxco et al., 1997) is the so-called "random coil" Gly-Gly-X-Gly-Gly pentapeptides (where X is any of the 20 amino acids). MD simulations (see Materials and methods) were performed for a representative set of 15 water bridges in some of the Gly-Gly-X-Gly-Gly pentapeptides with Gly, Ser, Thr, Asn, and Gln at the central position. The 15 conformers used in these MD calculations were selected based on their Ewb values [that is, minimized energy of water–protein nonbonded interactions in a particular water bridge conformation (see Materials and methods)]. Three conformers for each peptide sequence at low, moderate, and high values of Ewb have been selected for the same type of water bridges. It is noteworthy in all cases that the relative ranking of water bridge stability in conformation triples as determined by MD was found to be in accord with its Ewb ranking. Thus, MD runs served as a kind of blind test allowing estimation of the predictive power of our model.

The value of P_{wb} (that is, the fraction of the time when two hydrogen bonds of a particular water site are satisfied by a water molecule) has been used as a measure of water bridge stability. Figure 2A shows the time dependence for P_{wb} as calculated from MD trajectories of Gly, Ser, Thr, Asn, and Gln substituted pentapeptides. One can see from Figure 2A that 350 ps MD was enough to reach equilibrium for water exchanges in the water bridges under consideration. The value of P_{wb} is related, of course, to the average residence time of water molecules at a given site. The higher the P_{wb}, the longer the average residence time. However, water exchanges indeed occur frequently even at highly stable water bridges. We observed a few cases where a water molecule lived as long as 150 ps of MD trajectory in a particular water bridge. However, in the majority of cases, the average water residence time of a water bridge was in the range of 10–20 ps. This

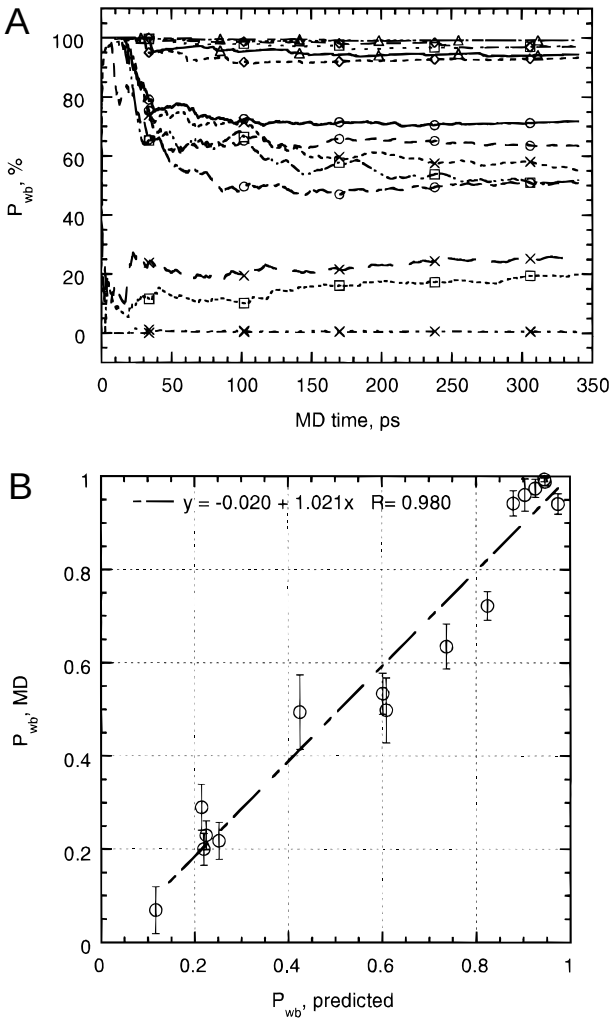

Fig. 2. A: Time course from the MD simulations performed for a representative set of 15 water bridges in some of the conformations of the Gly-Gly-X-Gly-Gly pentapeptides with Gly (○), Ser (×), Thr (△), Asn (◇), and Gln (□) at the central position. The values P_{wb} were calculated from MD trajectories as described in Materials and methods. **B:** Correlation between the probability for a water bridge to occur as determined from MD calculations and that calculated from Equation 5. The standard deviations of the MD based probabilities for the last 250 ps of the MD simulations are shown as error bars.

is in agreement with available NMR data that water residence time at protein surface is in the subnanoseconds interval (Otting et al., 1991). Thus, dynamical properties of water in a stable water bridge can differ from that of moderately stable water bridges by as much as a factor of 15 and probably even more for bulk water molecules.

As expected, the MD simulations showed rather diverse conformational stabilities for the water bridges depending on the type and peptide conformation. However, we find a surprisingly good correlation (see Fig. 2B) between P_{wb} determined from MD and that calculated from Equation 5 with a constant C value of 7.7 kcal/mol [similar to the value expected from basic physical principles (see Results and discussion, Theory)]. These results corroborate our hypothesis that C should be approximately independent on peptide conformation, sequence context, or water bridge type. The

value of C determined here could be dependent on the force field used for the Ewb calculations. However, since all available force fields have similar energy parameters for hydrogen bonds and van der Waal's interactions (the main contributions to Ewb), the changes should be minor.

It is noteworthy also that only the minimized values of Ewb were found to correlate well with P_{wb} determined from MD. This is quite understandable given the sensitivity of hydrogen bond potential to small changes in hydrogen bond distances and angles. Moreover, it was recently shown that hydrogen bonds are to a significant extent of covalent nature (Martin & Derewenda, 1999); therefore, it is expected that protein solvation should have a very significant nonelectrostatic contribution from water protein hydrogen bonding. Thus the exact positions of the water bridges are essential to determine its stability and free energy contribution.

Having determined E_{conf} (Equation 7) for all possible $\phi,\psi,\chi1$ to $\chi4$ conformations with grid steps of 10 or 20° for the central residue in Gly-Gly-X-Gly-Gly peptides, we can determine their relative populations and from those the average NMR parameters. In Table 1 we show the results of the correlation analysis between the experimental NMR $^3J_{\mathrm{HN}\alpha}$ coupling constant values (which are related to the equilibrium distribution of the ϕ dihedral angle)

Table 1. *Correlation between the experimental and the predicted $^3J_{HN\alpha}$ coupling constants for naturally occurring amino acids in unordered peptides*

AA	$^3J_{\mathrm{HN}\alpha}$ (Bundi & Wuthrich, 1979)	$^3J_{\mathrm{HN}\alpha}$ (Dyson[b])	$J_{\mathrm{HN}\alpha}$ (Plaxco et al., 1997)	$J_{\mathrm{HN}\alpha}$ (Fiebig et al., 1996)	$J_{\mathrm{HN}\alpha}$ (Average experiment)	$^3J_{\mathrm{HN}\alpha}$, PDB (Serrano, 1995)	$^3J^{aver}_{\mathrm{HN}\alpha}$ ECEPP/2	$^3J^{aver}_{\mathrm{HN}\alpha}$ +ASA	$^3J^{aver}_{\mathrm{HN}\alpha}$ +WB	CS($\mathrm{H}_{C\alpha}$)[c] +WB
A	6.5		6.1	6.1	6.23	6.49	6.4	6.3	6.4	−0.0306
C	7.7	7.2	7.3		7.40	7.17	6.2	6.5	7.0	0.0090
D	7.0	7.3	7.8	7.3	7.35	7.08	6.5	6.6	7.0	−0.0028
E	7.0	7.1	6.7		6.93	7.09	6.1	6.1	7.2	0.0584
F	9.4[d]	6.9	7.3		7.10	7.56	6.8	6.8	7.0	0.0399
G	5.6	5.7			5.65	5.52	5.4	5.5	5.5	0.0212
H	8.0	7.4	7.8	7.5	7.68	7.47	6.4	6.4	7.0	0.0242
I	7.0	6.6	7.1	7.0	6.93	7.57	7.1	7.1	7.2	0.0929
K	6.5	6.5	7.0	7.0	6.75	7.21	6.4	6.5	7.0	0.0489
L	6.5	6.6	6.8	6.9	6.70	7.21	6.3	6.5	6.7	0.0076
M	6.8	6.7	7.1		6.87	7.26	6.4	6.4	6.6	−0.0106
N	7.5	7.4	7.7	7.3	7.48	7.48	6.0	6.4	7.0	−0.0053
Q	6.0	7.4	7.1		6.83	7.32	6.2	6.5	6.9	0.0354
R	6.9	6.5	6.9	6.7	6.75	7.35	6.2	6.1	6.6	0.0119
S	6.5	6.4	7.0	6.5	6.60	6.93	6.2	6.1	6.6	−0.0182
T	6.9	7.6	7.9	7.3	7.43	7.72	6.4	6.2	6.9	0.0307
V	7.0	6.8	7.2	7.4	7.10	7.78	7.1	7.1	7.2	0.0801
W		7.3	7.0	6.5	6.93	7.08	6.4	6.7	6.9	0.0249
Y	6.8	6.8	7.8	7.0	7.10	7.32	6.9	6.9	7.0	0.0292
R1	0.328	0.210	0.163	0.237	0.427	0.690				
R2	0.387	0.341	0.235	0.322	0.512	0.687				
R3	0.646	0.678	0.464	0.778	0.807	0.863				

[a]The $^3J_{\mathrm{HN}\alpha}$ coupling constants were calculated using Karplus's equation (Vuister & Bax, 1993):

$$^3J_{\mathrm{HN}\alpha} = 6.51 \times \cos^2(|\phi - 60|) - 1.76 \times \cos(|\phi - 60|) + 1.6$$

for each point of the conformational grid. The weighted average value of the $^3J_{\mathrm{HN}\alpha}$ coupling constant was calculated using classical Boltzmann–Gibbs distribution:

$$^3J^{aver}_{\mathrm{HN}\alpha} = \sum_{igrid}\left[\left(\exp(-E_{conf}/RT)\Big/\sum\exp(-E_{conf}/RT)\right)\times {}^3J_{\mathrm{HN}\alpha}\right]$$

where E_{conf} is conformational energy calculated with Equation 7, R is the gas constant, and T is the temperature in kelvin. R1, R2, R3 are the correlation coefficients between the experimental scales and predicted $^3J_{\mathrm{HN}\alpha}$ coupling constants using (1) the ECEPP/2 force field (Momany et al., 1975; Nemethy et al., 1983) without solvation potential; (2) the ECEPP/2 force field plus ASA-based potential (Ooi et al., 1987); and (3) the ECEPP/2 ASA-based solvation potential plus the explicit water bridging energy term as shown in Equation 7. The experimental errors of the $^3J_{\mathrm{HN}\alpha}$ values is ~0.5 Hz.

[b]H.J. Dyson, pers. comm.

[c]The conformational shifts as calculated with SHIFTS computer program (Osapay & Case, 1991). The average of the conformational shifts is 0.024 ppm, and its standard deviation is 0.03 ppm. A perfect prediction should result in zero values for all the amino acids. The error in the prediction of the experimental values in proteins is ≈0.1 ppm (Osapay & Case, 1991).

[d]This value has been excluded from the correlation analysis because it is too high to belong to an unordered polypeptide.

(Bundi & Wüthrich, 1979; Fiebig et al., 1996; Plaxco et al., 1997), with the results of our calculations using several force field approximations: (1) the ECEPP/2 force field (Momany et al., 1975; Nemethy et al., 1983) without solvation potential; (2) the ECEPP/2 force field plus the ASA-based potential (Ooi et al., 1987); and (3) the ECEPP/2, the ASA-based solvation potential plus the explicit water bridging energy term as shown in Equation 7. It is clear from this analysis that the model shown in Equation 7 correctly reproduces the experimental values within the experimental errors. The right column of Table 1 shows the C_α proton conformational shifts values relative to the random coil chemical shifts. Our predictions are within the errors in the calculation of the experimental values in proteins ($\approx$0.1 ppm) (Osapay & Case, 1991) and close to zero, which are expected values for a perfect prediction.

The model using ECEPP/2 force field and ASA-based solvation without water bridges reproduces correctly the NMR parameters for Gly, Ala, Cys, nonpolar (Leu, Ile, and Met), aromatic (Phe, Tyr, and Trp), and positively charged Lys and Arg amino acids. The reason for it is that water bridges with backbone atoms only occur at high energy areas or steep slopes of energy walls in the (ϕ,ψ) map (at least 3–5 kcal/mol above global energy minimum) (Fig. 3A). Therefore, although water bridges stabilize these strained peptide conformations and they can be detected in protein crystals (Thanki et al., 1990, 1991; Morris et al., 1992), their energy contribution is not large enough to produce a significant shift in the peptide conformational equilibrium. In the case of small polar residues (Ser, Thr, Asn, Asp, His, Gln, and Glu), Table 1 shows strong changes for coupling constants if the free energy contribution of water bridges is included. Therefore the consideration of water bridges is essential to reproduce the experimental data for the short polar residues capable to form water bridges between its side chains and the backbone. Unlike the water bridges in the backbone the water bridges (DD, AA, DA, and AD) between side chain and backbone atoms occur in low energy conformations of the protein backbone (regions of right- and left-handed α-helices, parallel and antiparallel β-structures, as well as variety of extended conformations) (Fig. 3B). As a result, water bridge contribution should significantly affect the conformational equilibrium of peptides with small polar side chains. In the case of Lys and Arg residues, water bridges play a less important role because of the large entropy penalty for fixing four side-chain dihedral angles in a suitable conformation for a water bridge.

Figure 3B indicates that water bridges of different type and energy are present in all populated backbone conformations of the peptide with Ser at central position. The similar situation was found for other polar and charged residues. However, the distribution and depth of energy minima in the (ϕ,ψ) maps were found to be dependent on the type of amino acid under consideration (data not shown). Generally the water bridges are found to be the most stable in the areas around $\phi = -120°$ and $+60°$ that correspond exactly to maximums of the Karplus's equation for the $^3J_{HN\alpha}$ coupling constant (see footnote to Table 1). That is the reason the presence of the water bridges always tends to increase the equilibrium values of the coupling constants, that depends on ϕ only, as seen from Table 1, especially for the short polar residues. The preferable ψ areas for water bridge formation are dependent on amino acid types. Figure 3B shows, as a typical example, the distribution of low energy areas in (ϕ,ψ) map of Ser. Very similar pattern was found also for Thr. For other polar and charged residues, the list of low energy areas of (ϕ,ψ) map is given in the caption of Figure 3.

Conclusions

In conclusion, we must also note that accounting for water bridges is essential for understanding peptide and protein conformational equilibrium in the unfolded state, and probably it is important for the energetics of the folded conformation in proteins. The model present here provides accurate and computationally effective way ($\sim$0.1 s vs. $\sim$10 h in terms of demanding computational time for this model and MD with explicit water box, respectively) to estimate the energy of protein solvation for any given protein conformation. We must also stress that the solvation model presented here is applicable not only for proteins, which was the subject of this work, but also to other biomolecules (i.e., DNAs, RNAs, polysaccharides, etc.) that have flexible geometry and contain many polar/charged groups capable of water bridge formation.

Materials and methods

Positions of water bridges

Approximate positions of water molecules were calculated for any given peptide conformation using a complete list of all possible pairs of protein atoms that can participate in hydrogen bonds with water as follows. For each pair of protein atoms with overlapping water shells (the spheres of radii that correspond to ideal hydrogen bond distances plus 0.2 Å to account for possible deviations from ideal hydrogen bonds), the overlapped area has been filled up with uniformly distributed positions for possible water bridge oxygens. The distance between neighboring positions was kept to be less than 0.2 Å. All positions in the overlapped area that had van der Waals clashes with any protein atom (except those that are considered to be bridged by the water molecule) were eliminated. In addition, those water positions that are closer to the corresponding protein donor/acceptor by more than 0.2 Å than its ideal hydrogen bond distance were eliminated as well. The remaining water positions were checked to satisfy the following geometrical conditions: (1) the hydrogen bond angle Donor-H…acceptor must be in the range of 140–180° as was found in organic crystals (Taylor & Kennard, 1984); (2) for those bridges where water accepts a hydrogen atom from the protein, the water acceptor angle ($H\text{-}O_{water}\ldots H$) must be in the range of 100–200°, which was found in ab initio calculations of water dimers (Finney, 1982); (3) for those bridges where water donates its hydrogen to carbonyl and hydroxyl oxygens of a protein, the Protein-$O\ldots H_{water}$ angle must be in the range of 100–140° as was found in protein crystal structures at high resolution (Thanki et al., 1990, 1991; Morris et al., 1992). The positions of the remaining two hydrogen bonds of a water bridge molecule that should be satisfied by the bulk water molecules were checked for absence of steric clashes with protein atoms. In the presence of a clash, the respective water bridge was penalized by 3.2 kcal/mol [this value corresponds to one water-water hydrogen bond (Feyereisen et al., 1996)]. From a few hundreds, on average, of possible water positions in each overlapped area of water shells of protein donor and/or acceptor atoms, the position with the lowest water–protein interaction energy was selected and the energy was optimized using a conjugate gradient method.

Free energy calculations

The energy profiles of the Gly-based pentapeptides substituted with all (except Pro) naturally occurring amino acids at its central

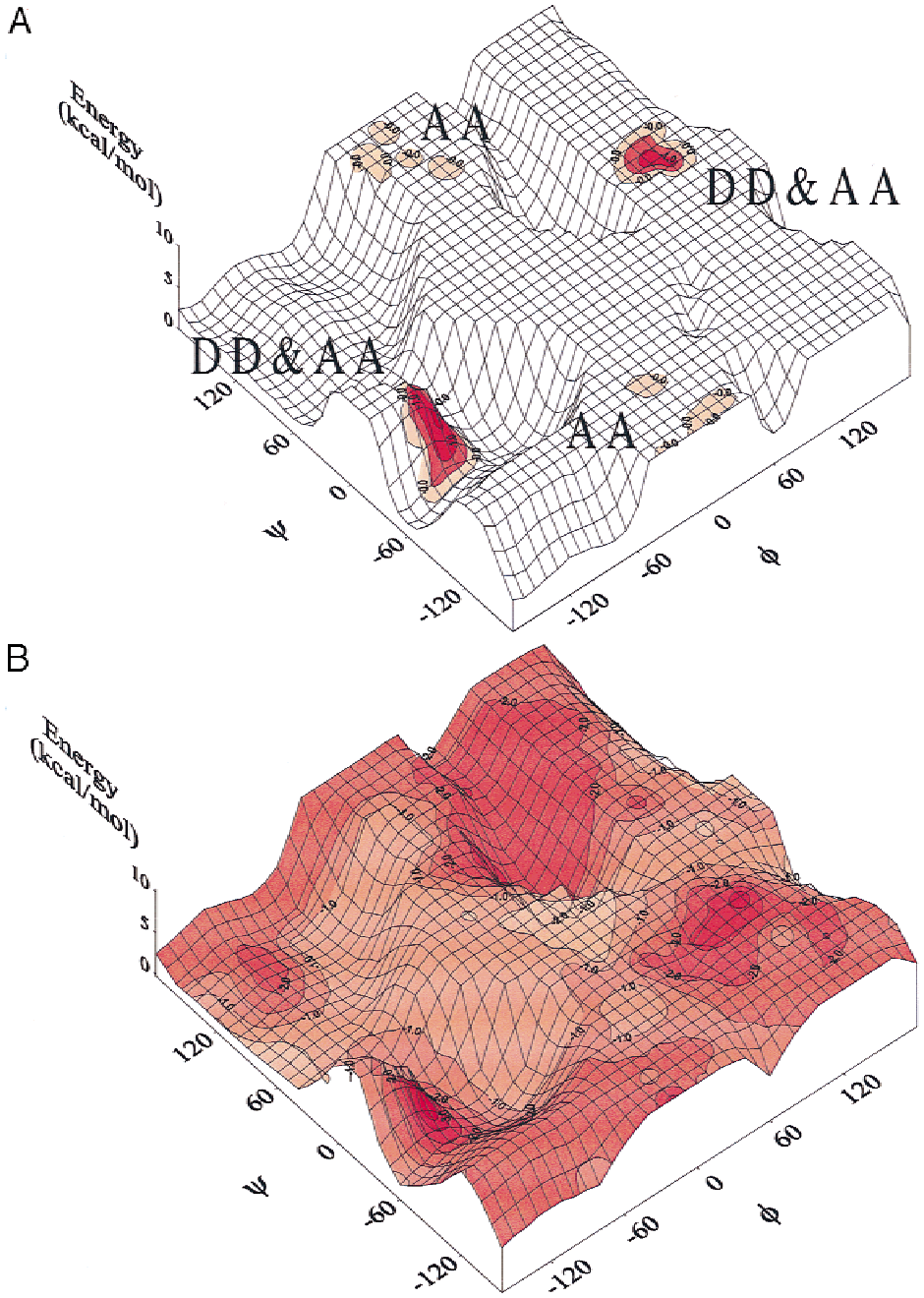

Fig. 3. Energy surface landscape for **(A)** Ala and **(B)** Ser in the Gly-Gly-X-Gly-Gly pentapeptides. Shown in white is the energy of intrapeptide interactions E_{pp} as calculated with the ECEPP/2 (Momany et al., 1975; Nemethy et al., 1983) force field plus the ASA-based solvation (Ooi et al., 1987) potential for every point in the conformational grid for Ala and Ser at central position. Shown in color is the energy contribution of different water bridges. The lowest energy conformation of the side chain of Ser is selected for each particular (ϕ,ψ) pair. Very similar pattern of low energy minima was found also for Thr. For other polar and charged amino acids, the most preferable (ϕ,ψ) areas were found to be following: Asp, Glu: wide areas of (160–260°; 0–360°) and (60–100°; 0–360°); Asn: (0–100°; 60–190°), (180–300°; 300–360°), and (180–300°; 70–190°); Gln: (180–300°; 0–20°), (180–300°; 100–120°), (180–300°; 160–180°), (180–300°; 300–350°), and (0–60°; 60–140°); Arg: (180–300°; 0–60°); (180–300°; 100–160°); (180–300°; 180–200°); (180–300°; 280–320°), and (30–100°; 100–160°); Lys: (160–200°; 60–160°), (160–200°; 260–340°), (280–300°; 0–360°), and (20–400°; 30–80°); His: (160–200°; 50–100°), (270–300°; 270–290°), and (30–60°; 0–120°).

positions were calculated on a grid of backbone and side-chain dihedral angles (ϕ, ψ, χ1 to χ4) with grid steps of 10° (Gly and Ala substituted peptides) and 20° for other amino acids. Each point in the grid was minimized by 200 steps of energy minimization by the conjugate gradient method. The dihedral angles ϕ,ψ and ω of the backbone of flanking residues were initially set to 180° and allowed to vary by a conjugate gradient energy minimization algorithm. The energy calculations were made with the BKS mo-

lecular modeling program (Abagyan & Mazur, 1989) using the ECEPP/2 force field (Momany et al., 1975; Nemethy et al., 1983). All atoms in the peptides were treated explicitly. Bond lengths and bond angles were fixed at their standard values during the energy calculations and minimization. van der Waals', electrostatic, hydrogen bond, and torsion potentials were included in the energy calculations. A dielectric constant of 81 was used for protein–protein interactions to mimic the water screening of electrostatic interactions within the protein (Finkelstein, 1977; Warshel & Papazyan, 1998). Electrostatic interactions are an integral part of the hydrogen bonding potential of the ECEPP/2 force field, and the parameter set of the force field has been verified using a dielectric constant of 2. Therefore, we use this dielectric constant to reproduce correct values for hydrogen bond energies in bridge water–protein interactions and for the interactions between the nonhydrogen bonded atoms of water molecule and protein atoms that have covalent bonds to protein bridging atoms. The high dielectric constant of 81 was used for remaining part of water–protein nonbonded interactions. Although ECEPP/2 does not include any additional functions involving bond and dihedral angles, the above-mentioned approximation of the water–protein interactions showed reasonably good angular dependence of the hydrogen bond energy. Similar results were reported recently (No et al., 1995). Complete lists of nonbonded interactions of peptide series were used in all calculations to avoid the update of interaction pair lists. The protein surface based solvation energy term was modeled by the continuum approximation model for protein solvent interactions (Ooi et al., 1987). Accessible surface area was calculated with the NSC program (Eisenhaber et al., 1995). The source code of the algorithm was kindly provided by Dr. Eisenhaber (EMBL). The van der Waals' radii and atomic solvation parameters were taken from (Ooi et al., 1987).

MD simulations

The MD calculations were performed with AMBER 4.1 package (Cornell et al., 1995; Pearlman et al., 1995). Each peptide structure was immersed in a box of explicit TIP3P waters, with walls at least 10 Å away from any peptide atom. The water box was then truncated to an octahedron, and periodic boundary conditions were employed to eliminate boundary effects. All peptide conformations were kept rigid during the simulations. Both N- and C-termini were uncharged. Nonbonded interactions were evaluated at every step, applying a 12 Å residue-based cut-off. The SHAKE algorithm (van Gunsteren & Berendsen, 1977) was used to constraint all bonds during the MD simulations, and the time step was set to 0.002 ps. All calculations were performed on a Silicon Graphics Octane/R10000 workstation. MD simulations were calculated at 293 K for at least 340 ps, and Cartesian coordinates were saved on disk every 0.04 ps during the course of the trajectories, leading to sets of 8,500 frames. The following strategy was used to prepare each system to the MD runs: all water molecules were minimized, subjected to 10 ps of MD at constant volume to allow for the reorientation and relaxation of the water dipoles, and minimized again. After this procedure to randomize the water box, the system was heated gradually from 10 to 293 K during 20 ps, and then the temperature was maintained at 293 K for the rest of the constant pressure MD simulations. Analysis of H-bonds was performed with the CARNAL module from the AMBER 4.1 package. The probability of water bridge P_{wb} was calculated as the fraction of time that two H-bonds between a water bridge and the corresponding protein atoms are formed. It was evaluated during the last 250 ps of the MD simulations (~5,000 frames), thus allowing the system to equilibrate during the initial 100 ps. The two H-bonds were considered to be formed when distance between heavy atoms of a donor and an acceptor was $\leq$3.5 Å, and the H-Donor-Acceptor angle was $\leq$30°. To estimate the errors in the water bridge probabilities, the partial P_{wb} were calculated for five consequent 50 ps intervals in the last 250 ps of the MD trajectory. The average values of P_{wb} and its standard deviations are shown in Figure 2B.

Acknowledgments

M.P. acknowledges financial support from EC grant BIO4-CT97-2086. C.M.S. acknowledges financial support from PRAXIS XXI (BPD/4151/94), JNICT (PBIC/C/BIO/2037/95), and EC (BIO04-CT96-0413).

References

Abagyan RA, Mazur AK. 1989. New methodology for computer-aided modeling of biomolecular structure and dynamics. 2. Local deformations and cycles. *J Biomol Struct Dyn 6*:833–845.

Bonvin AMJJ, Sunnerhagen M, Otting G, van Gunsteren WF. 1998. Water molecules in DNA recognition II: A molecular dynamics view of the structure and hydration of the trp operator. *J Mol Biol 282*:859–873.

Bundi A, Wüthrich K. 1979. 1H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers 18*:285–297.

Cheng YK, Rossky PJ. 1998. Surface topography dependence of biomolecular hydrophobic hydration. *Nature 392*:696–699.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz JKM, Fergusson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc 117*:5179–5197.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 319*:199–203.

Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. 1995. The dobule cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J Comput Chem 16*:273–284.

Feyereisen MW, Feller D, Dixon DA. 1996. Hydrogen bond energy of the water dimer. *J Phys Chem 100*:2993–2997.

Fiebig KM, Schwalbe H, Buck M, Smith LJ, Dobson CM. 1996. Toward a description of the conformations of denaturated states of proteins. Comparison of a random coil model with NMR measurements. *J Phys Chem 100*:2661–2666.

Finkelstein AV. 1977. Electrostatic interactions of charged groups in water environment and their influence on the polypeptide chain secondary structure formation. *Molek Biol (USSR) 10*:811–819.

Finney JL. 1982. Towards a molecular picture of liquid water. In: Franks F, ed. *Biophysics of water*. New York: John Wiley & Sons Ltd. pp 73–95.

Franks F, ed. 1982. *Water a comprehensive treatise*, Vol. 7. New York: Plenum Press.

Hummer G, Garcia AE, Soumpasis DM. 1996. A statistical mechanical description of biomolecular hydration. *Faraday Discuss 103*:175–189.

Israelachvili J, Wennerstrom H. 1996. Role of hydration and water structure in biological and colloidal interactions. *Nature 379*:219–225.

Juffer AH, Eisenhaber F, Hubbard SJ, Walther D, Argos P. 1995. Comparison of atomic solvation parametric sets: Applicability and limitations in protein folding and binding. *Protein Sci 4*:2499–2509.

Kovacs H, Mark AE, van Gunsteren WF. 1997. Solvent structure at a hydrophobic protein surface. *Proteins 27*:395–404.

Martin TW, Derewenda ZS. 1999. The name is bond—H bond. *Nature Struct Biol 6*:403–406.

Momany FA, McGuire RF, Burgess AW, Scheraga HA. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions and intrinsic torsional potential for the naturally occurring amino acids. *J Phys Chem 79*:2361–2381.

Morris AS, Thanki N, Goodfellow JM. 1992. Hydration of amino acid side chains: Dependence on secondary structure. *Protein Eng 5*:717–728.

Nemethy G, Pottle MS, Scheraga HA. 1983. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. *J Phys Chem 87*:1883–1887.

No KT, Kwong OY, Kim SY, Jhon MS, Scheraga HA. 1995. A simple functional representation of angular-dependent hydrogen-bonded system. 1. Amid, carboxylic acid and amid-carboxylic acid pairs. *J Phys Chem 99*:3478–3486.

Ooi T, Oobatake M, Nemethy G, Scheraga HA. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc Natl Acad Sci USA 84*:3086–3090.

Osapay K, Case DA. 1991. A new analysis of proton chemical shifts in protein. *J Am Chem Soc 113*:9436–9444.

Otting G, Liepinsh E, Wüthrich K. 1991. Protein hydration in aqueous solution. *Science 254*:974–980.

Pearlman DA, Case DA, Caldwell JC, Ross WS, Cheatham TE, Fergusson DM, Seibel GL, Chandra Singh U, Weiner P, Kollman PA. 1995. *AMBER 4.1*. San Francisco: University of California.

Plaxco KW, Morton CJ, Grimshaw SB, Jones JA, Pitkeathly M, Campbell ID, Dobson CM. 1997. The effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. *J Biomol NMR 10*:221–230.

Serrano L. 1995. Comparison between the phi distribution of the amino acids in the protein database and NMR data indicates that amino acids have various phi propensities in the random coil conformation. *J Mol Biol 254*: 322–333.

Sharp KA, Honig B. 1990. Electrostatic interactions in macromolecules: Theory and applications. *Ann Rev Biophys Biophys Chem 19*:301–332.

Taylor R, Kennard O. 1984. Hydrogen-bond geometry in organic crystals. *Acc Chem Res 17*:320–326.

Thanki N, Thornton JM, Goodfellow JM. 1990. Influence of secondary structure on the hydration of serine, threonine and tyrosine residues in proteins. *Protein Eng 3*:495–508.

Thanki N, Umrania Y, Thornton JM, Goodfellow JM. 1991. Analysis of protein main-chain solvation as a function of secondary structure. *J Mol Biol 221*:669–691.

van Gunsteren WF, Berendsen HJC. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys 34*:1311–1327.

von Freyberg B, Richmond TJ, Braun W. 1993. Surface area included in energy refinement of proteins. A comparative study on atomic solvation parameters. *J Mol Biol 233*:275–292.

Vuister GW, Bax AJ. 1993. Quantitative J correlation. A new approach for measuring homonuclear three-bond J(HNHa) coupling constants in 15N enriched proteins. *J Am Chem Soc 115*:7772–7777.

Warshel A, Papazyan A. 1998. Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. *Curr Opin Struct Biol 8*:211–217.

Warshel A, Russell ST. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys 17*:283–422.

Williams RL, Vila J, Perrot G, Scheraga HA. 1992. Empirical solvation models in the context of conformational energy searches: Application to bovine pancreatic trypsin inhibitor. *Proteins 14*:110–119.