ELSEVIER

Biophysical Chemistry 78 (1999) 113–126

Biophysical Chemistry

Comparison of the potentials of mean force for alanine tetrapeptide between integral equation theory and simulation

Ninad V. Prabhu[a], John S. Perkyns[a], Herb D. Blatt[a], Paul E. Smith[b], B. Montgomery Pettitt[a,*]

[a] *Department of Chemistry, University of Houston, Houston, TX 77204-5641, USA*
[b] *Department of Biochemistry, Kansas State University, Manhattan, KS 66506-3702, USA*

Received 18 January 1999; accepted 26 January 1999

Abstract

The dielectrically consistent reference interaction site model (DRISM) integral equation theory is applied to determine the potential of mean force (PMF) for an alanine tetramer. A stochastic dynamics simulation of the alanine tetramer using this PMF is then compared with an explicit water molecular dynamics simulation. In addition, comparison is also done with simulations using other solvent models like the extended reference interaction site model (XRISM) theory, constant dielectric and linear distance-dependent dielectric models. The results show that the DRISM method offers a fairly accurate and computationally inexpensive alternative to explicit water simulations for studies on small peptides.

Keywords: Integral equation; PMF; Solvent model; Simulation; Alanine; Molecular dynamics; Stochastic dynamics

1. Introduction

The determination of the conformation of peptides in their native state in aqueous solution and when bound to a receptor in different lipid environments is a problem of current interest [1]. There can be a structural relationship between the conformations of the peptides in different environments, and it is believed that the characteristics of binding to a receptor are sometimes determined by the conformations available to the peptide in solution [2,3]. Several experimental methods such as NMR, circular dichroism (CD) and ESR are routinely used to determine the solvent structure of peptides, and computational methods have been used to supplement and interpret this information, as in the case of structures and relative populations of several conformers in fast exchange, that may not be directly discernable from experiments [4].

Molecular dynamics simulations of model pep-

* Corresponding author.

0301-4622/99/$ - see front matter
PII: S0301-4622(99)00014-9

tides in explicit solvents have gained acceptance as a method to determine the conformational properties of peptides in solution [5]. The use of an explicit solvent model makes these calculations computationally expensive, and alternate methods need to be chosen where the balance between accuracy and computational efficiency may be optimized [6]. Also, in sampling studies of some peptides it is advantageous to run a molecular dynamics simulation at elevated temperatures [7,8] which may be impractical if explicit solvent is included [9].

Methods based on integral equations have been developed in which the effects of the solvent on interactions between solute atoms can be represented more accurately than most other kinds of implicit solvent or continuum models [10–15]. Despite the fact that these methods are approximate due to the difficulty involved in calculating some of the integrals in the underlying theory [16], electrostatic interactions and solvent packing effects can be determined with acceptable accuracy. Employing these methods to completely model the solvent interaction with a full multisite polypeptide has proven to be difficult when the number of sites exceeds $O(10^2)$ due to numerical difficulties [17]. New developments in the numerical procedure may, however, alleviate this problem (J.S. Perkyns, unpublished results).

The superposition approximation [13,18] can be used to obtain the solvent contribution to the free energy of solvation of the peptide as a pairwise sum of the free energies of solvation of the constituent atoms of the peptide at infinite dilution. These can be expressed as pairwise potentials of mean force (PMF) that may be appended to a molecular mechanics force field calculation as lookup tables to determine the free energy and its first and second order derivatives. This approximation has been found satisfactory when dealing with atoms in short polypeptide chains that are separated by at least three bond lengths [13,19]. In order to include the dynamical effects of the solvent effects, namely solvent viscosity and solvent collisions with the peptide, a stochastic dynamics simulation based on the integration of the Langevin equation can be used to generate a trajectory using the force fields described above [13,20].

In this study we present the results from a stochastic dynamics simulation of a neutral alanine tetramer using PMFs calculated from the solutions to the dielectrically consistent reference interaction site model (DRISM) integral equations [14,15]. The DRISM theory removes the inherent dielectric inconsistency in the earlier RISM based theories [13,18]. The results are compared with those from an explicit water simulation of the same peptide. A comparison is also done with simulations using the XRISM (extended RISM) theory [13] and other commonly used solvent models like the constant dielectric and linear distance-dependent solvent models.

To set the background for this work, previously published studies on the conformations of alanine based peptides are reviewed in Section 2. The details of obtaining the DRISM integral equation based PMFs are described in Section 3, and the details of the simulations and the basis for comparison are detailed in Section 4. The results are presented and discussed in Section 5 and are summarized in the Section 6.

2. Alanine peptides

Short-length alanine based peptides, especially the alanine dipeptide, have served as central models for the theoretical studies of the backbone conformations in proteins [21,22]. The free energy surface of the alanine dipeptide on the (ϕ,ψ) dihedral space is characterized by the minima describing the C_7^{eq}, C_5, C_7^{ax}, α_L, α_R, β and P_{II} conformations. The C_7 conformation corresponds to a pseudo seven-membered ring and the superscripts *ax* and *eq* indicate the axial or equatorial orientation of the β methyl group, respectively. The C_5 conformation forms a pseudo five-membered ring, α_L refers to the left handed helix, α_R to the right handed helix, P_{II} to poly (L-proline)-II helix and the β minimum corresponds to extended conformations.

The conformers of alanine in the gas phase have been determined by millimeter wave spectroscopy [23] and by gas-phase electron diffrac-

tion [24]. Ab initio studies on the structure of alanine at the restricted Hartree–Fock (RHF) and at various correlated (post-Hartree–Fock) levels of theory have been reviewed by Csaszar [25]. The results of experimental studies are in good accord with theory [23]. The relative energies of these conformers are determined by the type of hydrogen bonds, *cis* vs. *trans* arrangement of the carboxylic functional group, steric strain and the repulsion of lone electron pairs on the nitrogen and oxygen atoms [25].

The gas phase potential surface along the ϕ and ψ dihedral coordinates of the alanine dipeptide was calculated by Head-Gordon et al. [26,27] by ab initio methods at the MP2 level of electron correlation. The minima were found to be in the order C_7^{eq} $(-85°,67°) < C_5(-168°,170°) < C_7^{ax}(74°,-57°) < \alpha_L(64°,33°) < \beta(-128°,30°) < P_{II}$ $(68°,-177°)$ with relative energies of 0.0, 1.26, 2.53, 3.83, 5.95 and 8.16 kcal mol^{-1}, respectively.

The CHARMM [28] molecular mechanics force field exhibits the same number and type of gas-phase minima as the above calculations [27]. Zimmermann et al. [29] have calculated and compared the alanine gas phase energy surface by using several other empirical force fields like AMBER [30] and MM2 [31]. All of these were found to predict C_7^{eq} as the low energy minimum and most of them gave low energy conformations at the C_5, α_R and α_L regions. Recently the gas phase energy surface of the alanine tetramer was examined by Friesner et al. [32] using local MP2 with the cc-pVTZ(-f) correlation-consistent basis set and was compared with a variety of correlated ab initio and molecular mechanics force field calculations.

The crystal structure of the alanine tripeptide was found to be in an antiparallel β-pleated sheet conformation in the hemihydrate crystallized form and a parallel β-pleated conformation in the unhydrated form [33]. The circular dichroism (CD) spectra of the alanine dipeptide show that the major difference between the conformations of the peptide in polar and non-polar solvents is in the distribution of the ψ dihedral, with the proportions of the α_R conformation $(-80°,-50°)$ being dominant in the former kind of solvents and the C_7^{eq} conformation $(-80°,80°)$ being dominant in the latter kind [34]. The IR spectra of the aqueous alanine dipeptide show, on the basis of N–H bond stretching, that the structures were a mixture of hydrogen bonded five-membered ring species (53%) and a non-hydrogen bonded species (47%) [35]. Raman spectra and depolarized rayleigh scattering of aqueous alanine dipeptides show the preferred conformation as $(\pm 65° \pm 5°, \pm 35° \pm 5°)$ [36].

An early computer simulation on the free energy profile of the alanine dipeptide in explicit water on the (ϕ,ψ) coordinates was done by Mezei et al. [37] where the α_R and P_{II} conformations were found to be the most stable minima (Table 1). Anderson et al. [38] found two minima in the α_R region at $(-70°,-40°)$ and $(-120°, -40°)$, but the lowest energy state was identified as the β state at $(-110°,120°)$ and it was approximately 3 kcal mol^{-1} lower in energy than α_R. The free energy barrier between the β and α_R states was found to be between 0.5 and 1 kcal mol^{-1}.

The α helical and C_7 conformations were shown by Tobias et al. [21] to be destabilized relative to the β conformation by peptide–peptide interactions. However, unlike the C_7 case, the helical conformations have an overall stabilization due to favorable peptide–water interactions. The entropic contributions to the free energy were roughly the same for the β and C_7 conformations and were relatively destabilizing to the helical states since the latter bind more water molecules. The authors also state that there is an unfavorable electrostatic component to the contribution of the peptide–water interaction to the free energy when the peptide unit dipoles are aligned, as in the helical and C_7 conformations, as compared to the β conformation where the dipoles are antiparallel. The C_7 conformation, however, has a favorable intramolecular hydrogen bonding interaction. The energy barrier along the ψ dihedral coordinate for an α_R to β transition was a broad one with a maximum of 3 kcal mol^{-1} at approximately $-10°$.

According to the Zimm–Bragg theory [39] for helix to coil transitions, the formation of small helices should be unfavorable since the unfavorable entropic contribution to the free energy is

Table 1
Minima on the appropriate alanine dipeptide free energy surface

Solvent model		Minima on the appropriate free energy surface								
		α_R	α_L	C_7^{ax}	C_7^{eq}	C_5	P_{II}	β	β_2	α'
Explicit solvent[a]	ϕ	−50		90		150	150			
	ψ	−70		−90		−150	−80			
	ΔF	0.0		3.6			0.4			
Explicit solvent[b]	ϕ	−80	60	60				−80		
	ψ	−60	60	−80				120		
	ΔF	0.0	4.1	3.6				0.0		
ECEPP[c]	ϕ	−74	54	78	−84	−154			−150	−158
	ψ	−45	57	−64	79	153			72	−58
	ΔF	1.1	2.3	8.8	0.0	0.4			0.7	1.6
PPMMX[d]	ϕ	−68	55	56	−68	−158			−162	−157
$\epsilon = 80$	ψ	−44	46	−110	126	138			59	−56
	ΔF	0.0	0.6	0.7	0.1	0.4			0.7	0.3
XRISM[e]	ϕ	−67	63	64	−66	−177	−57			
	ψ	−55	49	−69	70	−180	−171			
	ΔF	1.8	1.0	0.0	0.5	0.1	0.0			
HF ORF[f]	ϕ	−75	61	73	−83	171	71			
	ψ	−28	38	−53	63	173	168			
	ΔF	1.6	1.8	2.0	0.0	1.3	5.5			
HF ORF[g]	ϕ		69	75	−73			−118	−112	
	ψ		39	−73	75			133	23	
	ΔF		5.4	5.4	3.5			1.2	0.0	

Notes. The location and the relative free energy (ΔF) of minima on the alanine dipeptide free energy surface as reported from studies using different solvent models and thermodynamic ensembles. The explicit solvent models are from [a]Mezei et al. [37] and [b]Tobias et al. [21]. The continuum solvent model studies are by Koca et al. [46] using [c]ECEPP and [d]PPMMX. [e]The XRISM model is by Lau et al. [47] and the Hartree–Fock with SCRF method are from [f]Shang et al. [27] and [g]from Gould et al. [50]. All the solvent models are described in the text and all free energies are in units of kcal mol^{-1}.

not offset by the favorable enthalpic contribution from the formation of hydrogen bonds. It has been suggested that [40] helical propensity is a cooperative phenomenon that increases with the length of the peptide and that helix propagation would be favored along the N-terminus rather than the C-terminus. It has also been demonstrated [3] that charged groups influence the conformations of the residues next to them. For the zwitterion the folded conformations were found to be preferred, whereas in the neutral peptide the preference was for the extended conformation. For short alanine polymers a 3_{10} helix (60°, −30°) is preferred over the α helix (60°, −50°) [41,42] but the latter is preferred in longer polymers [41,43–45].

A comparative study of the alanine dipeptide (ϕ,ψ) surface was made by Koca et al. [46] using various molecular mechanics programs. In some cases they included simple solvent representations like the constant dielectric models, and some of their results are included in Table 1. The locations and relative energies of the minima in some of their results are in fair agreement with explicit water simulations.

Pettitt and co-workers [13,19,20,47] have used

the XRISM integral equation theory to determine the PMFs and then construct the free energy surface for alanine dipeptide within the superposition approximation [13]. Though the minimum of their free energy surface was at C_7^{ax} (Table 1), the change in the free energy of solvation was found to be far more stabilizing for the extended and helical conformations than for the C_7 conformations [19,47]. On the other hand the entropic contributions to the free energy were found to be similar for the various conformers [47].

Pellegrini et al. [48] have also incorporated the effects of aqueous solvent on the free energy surface of alanine dipeptide by using PMFs in the superposition approximation. However, the correlation functions that give the PMFs were calculated from a Monte Carlo simulation of uncharged solute atoms in explicit water. A dielectric constant of 80 was used to compute the work to discharge and recharge the solute atoms and hence get the coulombic component of the PMF. It was shown from a comparative study of this method with a Monte Carlo simulation with explicit water that the free energy minima agreed within a tolerance of 2 kcal mol^{-1} whereas the barrier heights between the minima were considerably underestimated.

Straatsma et al. [9] determined the PMFs about the dihedral angles of alanine dipeptide in explicit solvent in an isothermal–isobaric ensemble by multiconfigurational thermodynamic integration. These PMFs were then fitted to a functional form and used for molecular dynamics simulations of alanine polymers.

In another continuum solvent approach, the linear Poisson–Boltzmann equation was solved for conformations on a (ϕ,ψ) grid to generate free energy profiles for alanine dipeptide [49]. The comparison of the PMFs with those calculated from an explicit solvent representation was shown to agree well on the positions of the minima and the energy barriers. However, some of the barrier heights were not in as good agreement.

Shang and Head-Gordon [27] used molecular orbital calculations to determine the conformational space of alanine dipeptide using HF/3–21G. The solvent effects were incorporated using the self-consistent reaction field (SCRF) method with the solute being enclosed in a sphere with an appropriately chosen constant dielectric constant. Also, the solute charge distribution was limited to a molecular dipole moment. Some of the approximations in the above study were improved upon by Gould et al. [50]. They used an ellipsoidal cavity for the solute, the dimensions of which were calculated from the van der Waals surface. The charge distribution of the solute was described by a single center multipole expansion up to $l = 7$ and the HF/6–31** basis set was used. The reported minima on the (ϕ,ψ) grid from both these studies are shown in Table 1.

3. Theory

The many properties of an n-atom polyatomic molecule in solution may be determined from the n-point singlet intramolecular distribution function $\rho^1(1)$, where 1 represents the set of vectors r_1^α of each site, α, of the molecule [19]. The distribution may be expressed as a product of the solvent influence function or the intramolecular singlet cavity distribution function, $y(1)$ and $e^{-\beta U(1)}$, the intramolecular Boltzmann factor. Here $U(1)$ represents the intramolecular potential and β^{-1} is the product of the absolute temperature, T, and Boltzmann's constant k_B. Thus $y(1)$ represents the solvent contribution to the intramolecular free energy [18].

For the cases where the atoms of the solute do not overlap and no clear interior exists, previous work [13,18,19] in this laboratory has shown that the n-point cavity distribution function may be approximated by the superposition of the two-site cavity distribution functions

$$y(1) = \prod_{\alpha,\gamma} y_{\alpha,\gamma}(r), \tag{1}$$

where $y_{\alpha,\gamma}(r)$ is the cavity distribution function between non-overlapping atomic sites α and γ. This approximation has been found satisfactory when dealing with atoms in short polypeptide chains that are separated by at least three-bond lengths [13,18,19]. For the one-component case, the radial distribution function, $g(r)$, is related to the cavity distribution function by

$$g(r) = y(r)\mathrm{e}^{-\beta U(r)}, \tag{2}$$

where $U(r)$ is the intramolecular potential in the gas phase. The relation between $g(r)$ and the PMF or work function, $W(r)$, is expressed as

$$g(r) = \mathrm{e}^{-\beta W(r)}. \tag{3}$$

By separating the PMF into its potential and solvent induced contributions,

$$W(r) = U(r) + \Delta W(r), \tag{4}$$

it can be seen from Eq. (2) that $\Delta W(r)$ is the cavity potential since

$$y(r) = \mathrm{e}^{-\beta \Delta W(r)}. \tag{5}$$

Using this framework the solvent contribution to the free energy of a polypeptide can be determined from the set of atomic pair cavity distribution functions, $y_{\alpha,\gamma}(r)$.

The cavity distribution function may be computed by several techniques, for example, by applying the umbrella sampling method to computer simulations. However, here we have used analytical methods based on the numerical solutions of integral equations which have a computational cost advantage over other methods. We have used the dielectrically consistent reference interaction site model equation (DRISM) [14,15] with a hypernetted-chain closure. This theory is closely related to the Ornstein–Zernike equation based-RISM [10,11] and extended-RISM (XRISM) theories [12]. The DRISM theory removes the inherent dielectric inconsistency in the earlier RISM based theories particularly for solutions with a finite salt concentration [51]. In earlier work the XRISM theory was used to calculate the solvent effects on a peptide at infinite dilution in pure water [19]. There is, however, a small difference in the quality of short-range correlations calculated by using XRISM and DRISM (J.S. Perkyns, unpublished results).

In fourier space the DRISM equation is expressed as

$$\left(\rho\tilde{\mathbf{h}}\rho - \tilde{\chi}\right) = (\tilde{\omega} + \tilde{\chi})\tilde{\mathbf{c}}(\tilde{\omega} + \tilde{\chi}) + (\tilde{\omega} + \tilde{\chi})\tilde{\mathbf{c}}\left(\rho\tilde{\mathbf{h}}\rho - \tilde{\chi}\right), \tag{6}$$

where $\tilde{\omega}$ is the intramolecular correlation matrix with elements $\tilde{\omega}(i,j) = j_0(-kd_{ij})$ with d_{ij} being the distance between site i and site j within the same rigid molecule. When $d_{ij} = 0$, $\omega(i,j) = 1$ and when i and j are on different sites, $\omega(i,j) = 0$. The diagonal matrix ρ represents the number density of various sites where u and v represent solute sites and solvent sites, respectively. Also, $\mathbf{h} = \mathbf{g} - \mathbf{1}$ is the matrix of intermolecular total correlation functions and $\mathbf{c}$ is the matrix of direct correlation functions. $\tilde{\chi}$ is a matrix consisting of elements

$$\tilde{\chi}_{i,j}(k) = -j_0(-kd_{ix})j_0(-kd_{iy})j_1(-kd_{iz}) \times j_0(kd_{jx})j_0(kd_{jy})j_1(kd_{jz})\left[\rho_v\tilde{h}_c(k)\right], \tag{7}$$

and d_{ix} is the distance between site i and its molecular center of charge in the x-direction (with the dipolar direction arbitrarily chosen as the z-direction) and $j_0(kd)$ and $j_1(kd)$ are zeroth and first order spherical Bessel functions, respectively. The form of this equation ensures that the functions have the correct asymptotic form in the limit of small k. The function $\tilde{h}_c(k)$ is approximated as

$$\tilde{h}_c(k) = \left[\frac{\epsilon - 1}{y\rho_v} - 3\right]\mathrm{e}^{-ak^2}, \quad a > 0, \tag{8}$$

where $a = 0.15$ for this application. Also, ϵ is the dielectric constant and $y = 4\pi\beta\rho_v\mu^2/9$, where ρ_v and μ are the solvent density and the solvent dipole moment, respectively. For a monoatomic solute at infinite dilution $\mathbf{h}$ may be separated into three equations for $\mathbf{h}^{vv}$, $\mathbf{h}^{uv}$ and $\mathbf{h}^{uu}$ as [13,19]

$$\left(\rho\tilde{\mathbf{h}}^{vv}\rho - \tilde{\chi}\right) = (\tilde{\omega} + \tilde{\chi})\tilde{\mathbf{c}}^{vv}(\tilde{\omega} + \tilde{\chi}) + (\tilde{\omega} + \tilde{\chi})\tilde{\mathbf{c}}^{vv}\left(\rho\tilde{\mathbf{h}}^{vv}\rho - \tilde{\chi}\right), \tag{9}$$

$$\tilde{\mathbf{h}}^{uv}\rho = \tilde{\mathbf{c}}^{uv}(\tilde{\omega} + \tilde{\chi}) + \tilde{\mathbf{c}}^{uv}\left(\rho\tilde{\mathbf{h}}^{vv}\rho - \tilde{\chi}\right), \quad (10)$$

$$\tilde{\mathbf{h}}^{uu} = \tilde{\mathbf{c}}^{uu} + \tilde{\mathbf{c}}^{uv}\rho\tilde{\mathbf{h}}^{vu}, \quad (11)$$

and since the polypeptide is considered to have independent monoatomic sites at infinite dilution, $\tilde{\rho}$, $\tilde{\omega}$ and $\tilde{\chi}$ refer to the solvent only. The second equation that relates **h** and **c** is the hypernetted chain (HNC) which in r space is written as

$$c_{ij} = e^{-\beta U_{ij} + h_{ij} - c_{ij}} - h_{ij} + c_{ij} - 1, \quad (12)$$

where h_{ij} and c_{ij} are the appropriate solute or solvent sites i and j and U_{ij} is the pair potential which has Lennard–Jones and coulombic components,

$$U_{ij}(r) = \frac{q_i q_j}{r} + 4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r}\right)^{12} - \left(\frac{\sigma_{i,j}}{r}\right)^{6}\right], \quad (13)$$

where $\sigma_{i,j}$ and $\epsilon_{i,j}$ are parameters for the Lennard–Jones diameter and the well depth.

The two equations can be solved by the Picard method [52]. The correlation functions for the solvent are first ascertained and then used to generate the solute–solvent correlation functions. These in turn are used to obtain the final solute–solute correlation functions. Finally, the PMF for the solute atoms in solution is obtained as a function of the radial separation of the solute atoms by

$$W_{ij} = -kT\ln(h_{ij} + 1). \quad (14)$$

The PMF incorporates the equilibrium effect of the solvation of a polypeptide, and the dynamical effects of the solvent may be incorporated by using a stochastic force with damping in the equations of motion [13]. The approach used here is an approximation to the generalized Langevin equation and is written as [20]

$$m_i\ddot{r}_i = -\sum_{j=1}^{N}\nabla W_{ij}(r_{ij}) - m_i\gamma_i\dot{r}_i + A_i(t), \quad (15)$$

where W_{ij}, as earlier, represents the PMF, γ_i represents the friction constant which is an approximation to the friction kernel of the generalized Langevin equation [20] and $A_i(t)$ is a random force.

The combination of using the PMF from the DRISM theory and the Langevin equation of motion is an approximation that reduces a many-body Hamiltonian problem to an effective few-body problem and results in a large reduction in computation time [20].

4. Method

The potential of mean force lookup table was created using atomic parameters from the CHARMM 22 all-atom parameter set [53] for the peptide atoms, and the TIP3P water model parameters, [54] with a modification to avoid catastrophic overlap, [55] were used for determining the correlation functions. A total of 19 different atom types was used in the PMF table, each representing a different set of values for $\sigma_{i,j}$, $\epsilon_{i,j}$ and coulombic charge. This set also includes atom types representing the reduced parameters that are used for the purpose of modeling 1,4 non-bonded interactions.

These tables provided all the non-bonded free energy while the other intramolecular interactions for the peptide were calculated from the CHARMM molecular mechanics force field [28] using the CHARMM 22 all-atom topology and parameter sets. The stochastic dynamics simulations (SD) were done using a friction coefficient of 2 ps^{-1}. This choice of friction coefficient was made to ensure a thorough sampling of the conformational space of the polypeptide following the work of Loncharich et al., who have reported that the rate of isomerization of alanine dipeptide was the maximum at that value [56]. The simulations were performed at a temperature of 300 K and the total time of the trajectories was 15 ns, computed with timesteps of 2 fs. Structures were saved every picosecond for further analysis. All other SD simulations which are used in this study were performed with the same conditions of time, temperature and friction coefficient and using the CHARMM 22 all-atom parameter set.

The peptide being studied here is NH_2–Ala–Ala–Ala–Ala–COOH which will be referred to as Ala_4. In the CHARMM force field

the non-bonded interaction between atoms that are three bondlengths apart (1,4 interactions) are calculated with atomic parameters that are different from those used for the rest of the non-bonded calculations. In order to check the optimum choice of the parameters to calculate the PMFs for 1,4 interaction, we performed three different simulations using different methods to incorporate the PMFs.

The method where the calculated PMFs for the 1,4 interactions did not use different parameters from those for the rest of the non-bonded interactions turned out to be the optimum choice. The simulation using this scheme will be referred to as DRISMSD and is used for the comparative study with the other solvent models.

The explicit water molecular dynamics simulation of Ala_4 was also done with the CHARMM 22 all-atom parameter set for the peptide atoms and TIP3P for the water model parameters. The simulation of the peptide was done in a cubic box having a length of 25 Å and using periodic boundary conditions. The box contained 512 water molecules and the electrostatic interactions were calculated using an Ewald sum [57]. The simulation had a total time duration of 10 ns using a timestep of 2 fs. The structures that were used for analysis were saved every tenth of a picosecond. More details on the simulation can be found in another publication [3].

We have also done a SD simulation with PMFs calculated using the XRISM integral equation theory for comparison. The methodology for calculating the PMFs is similar to earlier work [13]. However, the peptide and solvent parameters used are the same as those for the DRISM calculations, and this simulation will be referred to as XRISMSD.

A SD simulation using a constant dielectric of 1 will be used as a reference representing the gas-phase results (EPS1). In addition, SD simulations with some commonly used and computationally inexpensive solvent models are also used for comparison. One such simulation is done using a constant dielectric of 80 and will be referred to as EPS80. Another uses a linear distance-dependent dielectric (RD1) and the final simulation used a dielectric that is linearly-dependent on four times the distance in Angstroms (RD4).

The free energy surfaces in Fig. 1 were generated by the adiabatic mapping method [19]. Here the free energy of the peptide was determined on the relevant (ϕ,ψ) grid at intervals of 30°. The rest of the dihedrals were constrained to 180° at each grid point and this was followed by thorough minimization of the molecule. The free energy was evaluated after the constraints were removed.

The comparison between the various solvent models was done by first determining the population distributions of central dihedrals ϕ_2, ψ_2, ϕ_3 and ψ_3, from structures generated from the various methods. Next, the free energy profiles along the ϕ and ψ coordinates were determined from population Boltzmann distributions,

$$\frac{N(\chi)}{N(\chi_{\max})} = e^{-\beta W(\chi)}, \tag{16}$$

where $N(\chi)$ is the population at a particular value, χ, of the dihedral angle. Also, $N(\chi_{\max})$ is the population at the dihedral where the maximum number is observed for a given distribution, β is the inverse of the product of the Boltzmann constant and temperature (300 K) and $W(\chi)$ is the difference between the free energies at χ and $\chi_{\max}$.

For further analysis, the relative populations of alpha carbon pseudo-dihedrals were obtained from structures from the dynamics trajectories. This dihedral angle is formed by the four alpha carbon atoms (C_α) of the peptide and gives a good picture of its overall shape [58]. Low absolute values of the dihedral correspond to folded states where the ends of the peptide are in close proximity. Conversely, high absolute values indicate that the overall structure is extended. The population distribution of the pseudo-dihedrals was used to get the free energy profile along this coordinate.

5. Results and discussion

It can be seen from Fig. 1 how the free energy surface of Ala_4 is markedly changed by the inclu-

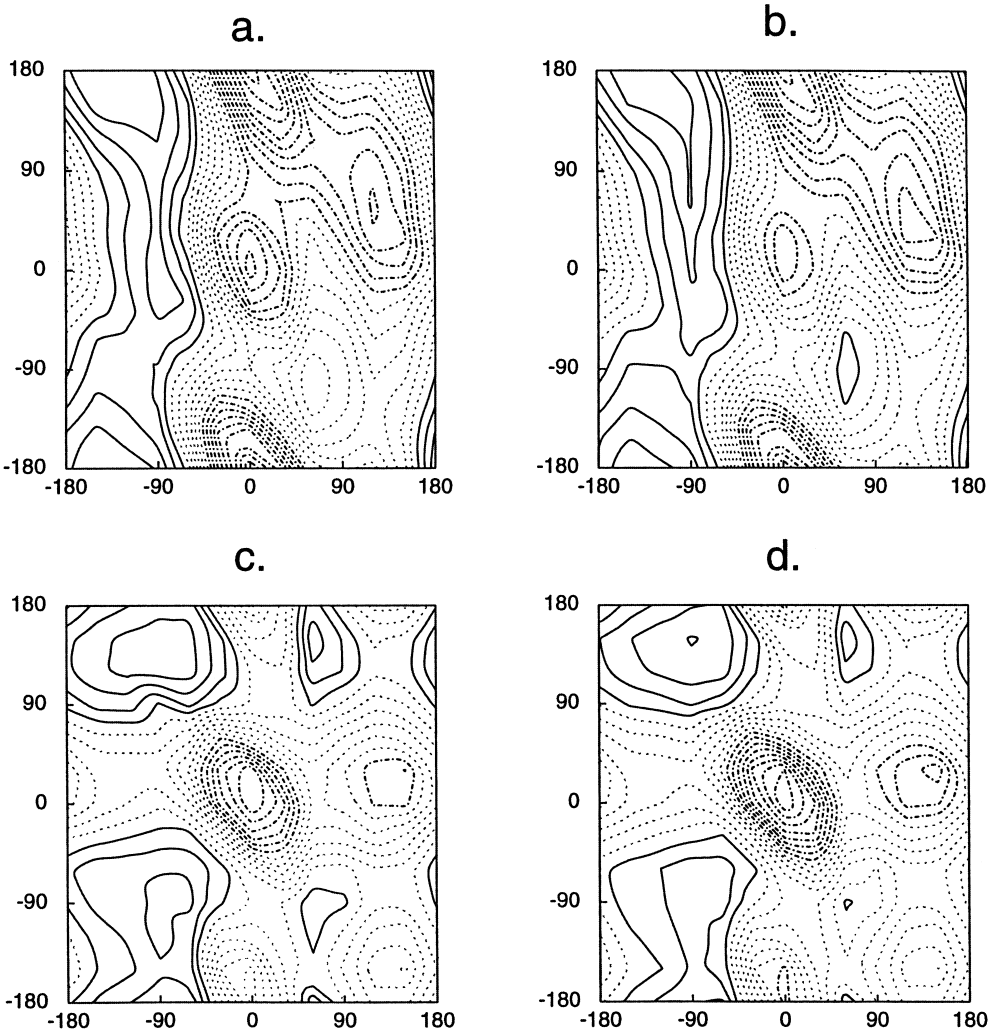

Fig. 1. The free energy surface of Ala_4 by adiabatic mapping. The contour levels are 1.0 kcal mol^{-1} each. The solid contours are for 0–4 kcals mol^{-1}, the dashed for 4–10 kcals mol^{-1} and the dot-dashed for >10 kcals mol^{-1}. The various free energy surfaces are: (a) (ϕ_2,ψ_2) for EPS1 model. (b) (ϕ_3,ψ_3) for EPS1 model. (c) (ϕ_2,ψ_2) for DRISMSD model. (d) (ϕ_3,ψ_3) for DRISMSD model.

sion of solvent effects from the DRISMSD PMFs. The minimum for EPS1 on both the (ϕ_2,ψ_2) (Fig. 1a) and (ϕ_3,ψ_3) surfaces (Fig. 1b) is in a region around $(-120°,180°)$. Most of the region bounded by $-60° < \psi < -180°$ and $-180° < \psi < 180°$ is within 2 kcal mol^{-1} from the minimum in both surfaces. However, a relatively higher energy region is seen around $(-160°,0°)$ in both cases. There are large energy barriers centered around (0°,0°), (0°,180°) and (60°,120°) and the rest of the two surfaces are of intermediate energy.

For the PMF modified case, the (ϕ_2,ψ_2) free energy surface had two regions corresponding to energy minima (Fig. 1c). The first was bounded by ϕ_2 with values roughly between $-60°$ and $-100°$ and ψ_2 between $-75°$ and $-140°$. The second region was bounded by ϕ_2 with values roughly between $-60°$ and $-130°$ and ψ_2 values between 120° and 170°. The region roughly bounded by ϕ_2 values between $-50°$ and $-180°$ and ψ_2 values between $-40°$ and $-180°$ and between 100° and 180° are within 2 kcal mol^{-1} of the energy mini-

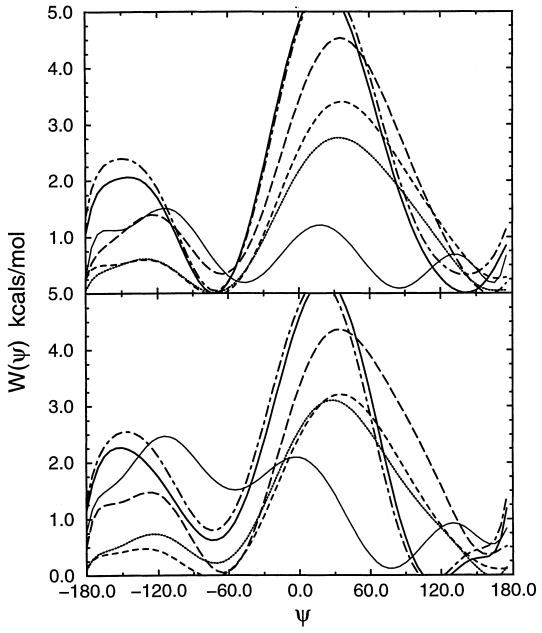

Fig. 2. The free energy profile along (a) ψ_2 (top panel) and (b) ψ_3 (bottom panel). The various curves represent: EPS1, thin solid line; DRISMSD, thick solid line; EWMD, long dashed line; XRISM, alternating long and short dashed line; EPS80, short dashed line; and RD4, dotted line.

mum. Low free energy regions along $\phi_2 = 60°$ are also observed. Relative to the EPS1 case, the free energy barrier around (0°,0°) is higher whereas those around (0°,180°) and (60°,120°) are lower in energy. Also, the region around (−90°,0°), which is fairly stable in the EPS1 case, is relatively destabilized by the solvent. Similar features are seen in Fig. 1d for the solvent modified free energy surface for (ϕ_3,ψ_3).

We now compare with the presumably correct molecular dynamics results. The free energy curves around ψ_2 (Fig. 2a) show that both the DRISMSD and EWMD results have two minima of nearly the same well depth (within 0.5 kcal mol^{-1}). One is around −60° corresponding to the *gauche*$^-$ state and the other around 150° which is the *trans* state.

The EPS80 and RD4 results also have two minima, in the *gauche*$^-$ and *trans* regions, for ψ_2 (Fig. 2a). The EPS1 curve (Fig. 2a) has equally stable minima in the *gauche*$^-$, *gauche*$^+$ and *trans* regions which are separated by energy barriers of approximately 1.5 kcal mol^{-1}. It can be seen that in the environment of explicit solvent there is a shift in the center of the *gauche*$^-$ minimum and a substantial destabilization of the *gauche*$^+$ state is caused. Also, the height of the free energy barrier around 0° increases by over 4 kcal mol^{-1}. The EPS80 model underestimates this barrier by >1 kcal mol^{-1} whereas the RD4 model underestimates it by approximately 1.5 kcal mol^{-1}. On the other hand in the DRISMSD case the barrier is overestimated. The other free energy barrier around −120° is overestimated by DRISMSD by approximately 0.5 kcal mol^{-1}. In the EPS80 and RD4 models this barrier is underestimated by almost 1 kcal mol^{-1}, having a value of only 0.5 kcal mol^{-1}, thus making the barrier region thermally accessible at room temperature. The RD1 results (figure not displayed) have one minimum around −90° and another one which is approximately 1.5 kcal mol^{-1} higher in energy in the *trans* region. The RD1 model thus shows a considerable deviation from the EWMD results and hence has been excluded from further analysis.

Unlike the DRISMSD and RD4 models where there are two equal minima, the *trans* minimum in the EWMD case is more stable than the *gauche*$^-$ minimum by almost 0.5 kcal mol^{-1}. In the EPS80 free energy profile the *gauche*$^-$ minimum is more stable but by an even smaller number. The DRISMSD and XRISMSD curves show a very good agreement though small differences exist. The DRISMSD curve is in slightly better agreement with the EWMD which is discernable from the regions belonging to the *trans* minimum and the barrier around −120°.

The free energy curves around ψ_3 (Fig. 2b) have a different quantitative character from those seen for ψ_2. The EPS1 curve has its most stable minimum in the *gauche*$^+$ state around 75°. The *trans* minimum is approximately 0.5 kcal mol^{-1} higher and the *gauche*$^-$ one is approximately 1.5 kcal mol^{-1} higher in energy. The EWMD curve has two minima, the more stable one at *gauche*$^-$ and the slightly higher one at *trans* being within 0.5 kcal mol^{-1} of each other. In the DRISMSD case the *trans* minimum is the more stable by approximately 0.5 kcal mol^{-1} and, this is also

displaced to around 120° from around 150° in EWMD. The two minima are almost equal in energy in the EPS80 and RD4 models though the *gauche*$^-$ is slightly more stable for the former and *trans* more stable for the latter models. Similar to the ψ_2 free energy profiles, in the ψ_3 case the EPS80 and RD4 models underestimate the free energy barriers whereas they are overestimated by DRISMSD. Also, the DRISMSD and XRISMSD curves are fairly similar though once again the former has a slightly better agreement with EWMD.

All these solvent models predict a dominant free energy minimum around $-90°$ for the ϕ dihedrals. The DRISMSD and XRISMD models have a minimum in the *gauche*$^+$ region unlike EWMD, which is not unexpected for DRISMSD since this is also observed in Fig. 1c,d. The free energy profiles from the SD simulations (figure not shown) reveal that the free energy well around $-90°$ has very similar characteristics for all the models. Also, the *gauche*$^+$ minimum for DRISMSD is more than 2 kcal mol^{-1} more unstable than the *gauche*$^-$ minimum. The XRISMSD *gauche*$^+$ minimum is even higher compared to its *gauche*$^-$ minimum and hence in this case is in slightly better agreement with EWMD. However, the *gauche*$^+$ region is more than 4 kcal mol^{-1} more unstable than the *gauche*$^-$ minimum for all the continuum models and hence the free energy profiles along the ϕ dihedrals show EPS1, EPS80 and RD4 to be in better agreement with the explicit solvent models than DRISMSD and XRISMSD are. However, to emphasise that the anomalous free energy well at 60° for the continuum solvent models represents only an insignificant number of the overall population, a population distribution is diplayed in Fig. 3.

The final comparative study between these models has been done by studying the free energy profiles around the pseudo-dihedral defined by the four C_α atoms of the peptide (Fig. 4). This pseudo-dihedral is a measure of the overall structure of the peptide. The EPS1 curve has minima at 60° and 180° separated by two barriers of approximately 1 kcal mol^{-1}, one centered around 120° and the other broader barrier with a maximum around 30°. The EWMD curve also has two minima but these are at around 0° and around $-120°$. The free energy barrier around 120° is also much higher compared to EPS1, the maximum being nearly 3 kcal mol^{-1}. The other barrier is centered around $-60°$ and is slightly over 0.75 kcal mol^{-1}.

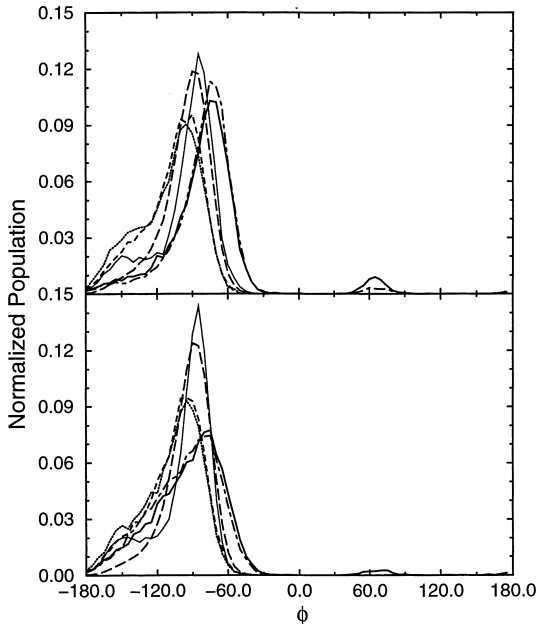

Fig. 3. The normalized probability distribution of (a) ϕ_2 (top panel) and (b) ϕ_3 (bottom panel). The various curves represent: EPS1, thin solid line; DRISMSD, thick solid line; EWMD, long dashed line; XRISM, alternating long and short dashed line; EPS80, short dashed line; and RD4, dotted line.

The DRISMSD curve has the same characteristics as the EWMD curve though the barrier at $-60°$ is slightly overestimated and the barrier at 120° slightly underestimated. The profiles for both RD4 and EPS80 also show maxima at 120° and minima at 0°, but do not have a barrier at $-60°$. The DRISMSD and XRISMSD curves are again qualitatively very similar but the DRISMSD curve has a slightly better match with EWMD.

On the basis of the above results, the EPS80 and RD4 models proved to be fairly good approximations to the aqueous solvent for describing the free energy profiles of Ala_4 along the dihedral coordinates. A small polypeptide, unlike large proteins, has all its atoms either exposed directly

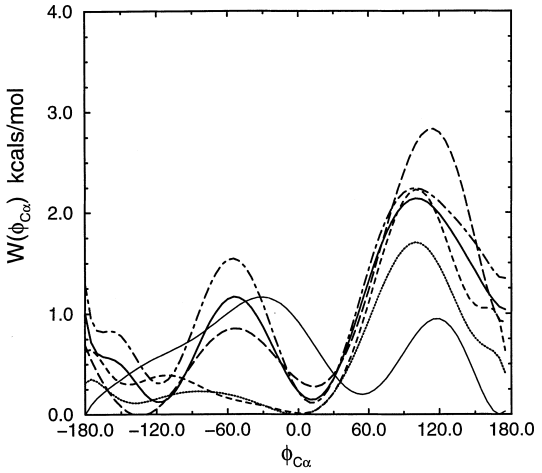

Fig. 4. The free energy profile along the C_α dihedral. The various curves represent: EPS1, thin solid line; DRISMSD, thick solid line; EMWD, long dashed line; XRISM, alternating long and short dashed line; EPS80, short dashed line; and RD4, dotted line.

or in close proximity to the solvent, and hence for some such systems, the EPS80 method has been used successfully [8,59]. On the other hand, the use of a constant dielectric of 80 tends to underestimate the structure making interactions [6] like hydrogen bonding and the coulombic interactions between highly charged groups and hence result in the underpopulation of the structural families where these interactions would be dominant. Structures with strong coulombic interactions would be better represented by the RD4 model than the EPS80 model [60].

Electrostatic interactions are represented far more accurately by the DRISMSD model compared to both EPS80 and RD4 and thus would be expected to give more accurate results for peptides with charged side chains. The DRISMSD and XRISMSD methods show better agreement with the EWMD results than the EPS80 and RD4 methods do. This was seen in the free energy profile around the central ψ dihedral and the C_α pseudo-dihedrals. The anomalous minimum in the ϕ dihedral profiles of DRISMSD and XRISMSD is only sparsely populated at physiological temperatures and thus would not hamper the determination of the conformations of peptides. Ala_4 does not have any charged side chains and this agreement may occur because the RISM based methods account for the packing effects of the solvent unlike EPS80 and RD4.

The DRISMSD results had a slightly better overall match with EWMD than XRISMSD did, though overall the two integral equation models showed fairly similar results and the observed differences are within the expected errors of these methods [61]. However, the DRISM theory shows a much more significant improvement over XRISM in the prediction of thermodynamic properties of salt solutions, and in the near future we shall be using this theory to model peptides in salt solutions at physiological concentrations. Presently, work is in progress in which the DRISM PMFs are used in a high temperature simulation of a tridecamer hormone (α-MSH) to determine its aqueous conformations. Another application using this method which is in progress is a room temperature SD simulation of α-MSH with appropriate friction coefficients to model the dynamical effect of the aqueous solvent.

6. Summary

In this study the DRISM integral equation theory was used to incorporate the effect of aqueous solvent as a superposition of PMFs between the atoms of a peptide. The model was then tested by comparison with an explicit water molecular dynamics simulation. The basis of comparison were the energy profiles around the peptide's central ϕ and ψ dihedrals and the pseudo-dihedral defined by the four C_α atoms. Simulations using PMFs calculated from the XRISM theory and those using other frequently used solvent models were also compared.

The results from the DRISM model exhibited the best match with the qualitative and quantitative aspects of the results from the explicit water simulation. The RD1 model was found to be inaccurate in this study. The DRISM and XRISM results were in fairly good agreement with each other as would be expected for such closely related theories. The EPS80 and RD4 models proved to be fairly good in predicting the minima on the free energy profiles though some of the free energy barriers were too low. The low com-

putational cost of the EPS80 and RD4 models makes them fairly useful for studies on some small peptides. The DRISM method has a computational cost comparable to the EPS80 and RD4 models but offers more accurate results and is thus the best alternative to the far more computationally expensive explicit water dynamics methods for simulations of small peptides.

Acknowledgements

The authors thank the N.I.H. and Robert A. Welch Foundation for partial support. HDB was supported by a fellowship from the Keck Center for Computational Biology and the NSF. Molecular Simulations, Inc. is thanked for visualization software. Dr Gillian Lynch and Michael Feig are thanked for stimulating discussion.

References

[1] S.G. Jacchieri, A.S. Ito, Int. J. Quantum Chem. 53 (1995) 335.
[2] V.J. Hruby, B.M. Pettitt, in: T.J. Perun, C.L. Propst (Eds.), Computer-Aided Drug Design, Marcel Dekker, New York, 1989, pp. 405–460.
[3] H.D. Blatt, P.E. Smith, B.M. Pettitt, J. Phys. Chem. B 101 (1997) 7628.
[4] D.O. Cicero, G. Barbato, R. Bazzo, J. Am. Chem. Soc. 117 (1995) 1027.
[5] B.R. Brooks, Chem. Scr. 29A (1989) 165.
[6] P.E. Smith, P.M. Pettitt, J. Phys. Chem. 98 (1994) 9700.
[7] B.M. Pettitt, T. Matsunaga, F. Al-Obeidi, C. Gehrig, V.J. Hrudy, M. Karplus, Biophys. J. 60 (1991) 1540.
[8] S.D. O'Connor, P.E. Smith, F. Al-Obeidi, B.M. Pettitt, J. Med. Chem. 35 (1992) 2870.
[9] T.P. Straatsma, J.A. McCammon, J. Chem. Phys. 101 (1994) 5032.
[10] H.C. Andersen, D. Chandler, J. Chem. Phys. 57 (1972) 1918.
[11] H.C. Andersen, D. Chandler, J. Chem. Phys. 57 (1972) 1930.
[12] F. Hirata, P.J. Rossky, Chem. Phys. Lett. 83 (1981) 329.
[13] B.M. Pettitt, M. Karplus, Chem. Phys. Lett. 121 (1985) 194.
[14] J. Perkyns, B.M. Pettitt, Chem. Phys. Lett. 190 (1992) 626.
[15] J. Perkyns, B.M. Pettitt, J. Chem. Phys. 97 (1992) 7656.
[16] J.P. Hansen, I.R. McDonald, Theory of Simple Liquids, 2nd ed., Prentice-Hall Inc., Academic Press, NY, 1986.
[17] G.L. Rame, Conformation of Small Biomolecules in Solution, Ph.D. thesis, University of Houston, 1992.
[18] L.R. Pratt, D. Chandler, J. Chem. Phys. 67 (1977) 3683.
[19] B.M. Pettitt, M. Karplus, P.J. Rossky, J. Phys. Chem. 90 (1986) 6335.
[20] P.E. Smith, B.M. Pettitt, M. Karplus, J. Phys. Chem. 97 (1993) 6907.
[21] D.J. Tobias, C.L. Brooks III, J. Phys. Chem. 96 (1992) 3864.
[22] T. Head-Gordon, F.H. Stillinger, M.H. Wright, D.M. Gay, Proc. Nalt. Acad. Sci. USA 89 (1992) 11513.
[23] P.D. Godfrey, R.D. Brown, F.M. Rodgers, J. Mol. Struct. 376 (1996) 65.
[24] K. Ijima, K. Tanaka, S. Onuma, J. Mol. Struct. 246 (1991) 257.
[25] A.G. Csaszar, J. Phys. Chem. 100 (1996) 3541.
[26] T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C.L. Brooks III, J.A. Pople, J. Am. Chem. Soc. 113 (1991) 5989.
[27] H.S. Shang, T. Head-Gordon, J. Am. Chem. Soc. 116 (1994) 1528.
[28] B.R. Brooks, R.E. Brucolerri, B.D. Olafson, D.J. States, S. Swaminahan, M.J. Karplus, J. Comput. Chem. 4 (1983) 187.
[29] C.H. Lee, S.S. Zimmerman, J. Biomol. Struct. Dynam. 13 (1995) 201.
[30] S.J. Weiner, P.A. Kollman, D.A. Case, et al., J. Am. Chem. Soc. 106 (1984) 765.
[31] N.L. Allinger, Pure Appl. Chem. 54 (1982) 2515.
[32] M.D. Beachy, D. Chasman, R.B. Murphy, T.A. Halgren, R.A. Friesner, J. Am. Chem. Soc. 119 (1997) 5908.
[33] A. Hempel, N. Camerman, A. Camerman, Biopolymers 31 (1991) 187.
[34] V. Madison, K.D. Kopple, J. Am. Chem. Soc. 102 (1980) 4855.
[35] C.I. Jose, A.A. Belhekar, M.S. Agashe, Biopolymers 26 (1987) 1315.
[36] M. Avignon, C. Garrigou-Lagrange, P. Bothorel, Biopolymers 12 (1973) 1651.
[37] M. Mezei, P.K. Mehrotra, D.L. Beveridge, J. Am. Chem. Soc. 107 (1985) 2239.
[38] A.G. Anderson, J. Hermans, Proteins 3 (1988) 262.
[39] B.H. Zimm, J.K. Bragg, J. Chem. Phys. 31 (1959) 526.
[40] W.S. Young, C.L. Brooks, J. Mol. Biol. 259 (1996) 560.
[41] W.R. Fiori, S.M. Miick, G.L. Millhauser, Biochemistry 32 (1993) 11957.
[42] P. Hanson, G. Martinez, G. Millhauser, et al., J. Am. Chem. Soc. 118 (1996) 271.
[43] P.S. Kim, D.J. Lockhart, Science 260 (1993) 198.
[44] J. Tirado-Rives, D.S. Maxwell, W.L. Jorgensen, J. Am. Chem. Soc. 115 (1993) 11590.
[45] L. Zhang, J. Hermans, J. Am. Chem. Soc. 116 (1994) 11915.
[46] J. Koca, P.H.J. Carlsen, J. Mol. Struct. 291 (1993) 271.
[47] W.F. Lau, B.M. Pettitt, Biopolymers 26 (1987) 1817.
[48] M. Pellegrini, N. Gronbech-Jensen, S. Doniach, J. Am. Chem. Soc. 104 (1996) 8639.
[49] T.J. Marrone, M.K. Gilson, J.A. McCammon, J. Phys. Chem. 100 (1996) 1439.

[50] I.R. Gould, W.D. Cornell, I.H. Hiller, J. Am. Chem. Soc. 116 (1994) 9250.
[51] J. Perkyns, B.M. Pettitt, Biophys. Chem. 51 (1994) 129.
[52] M.H. Gillan, Mol. Phys. 6 (1979) 1781.
[53] A.D. MacKerrel, Jr., D. Bashford, M. Bellot, et al., J. Phys. Chem. B 102 (1998) 3586.
[54] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Imprey, M.L. Klein, J. Phys. Chem. 79 (1983) 926.
[55] B.M. Pettitt, P.J. Rossky, J. Chem. Phys. 77 (1982) 1451.
[56] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Biopolymers 32 (1992) 523.
[57] P. Ewald, Ann. Phys. 64 (1921) 253.
[58] N.A. Akhmedov, E.M. Popov, Molekulyarana Biol. 23 (1988) 249.
[59] F. Al-Obeidi, M.E. Hadley, B.M. Pettitt, V.J. Hruby, J. Am. Chem. Soc. 111 (1989) 3413.
[60] V. Dagget, P.A. Kollman, I.D. Kuntz, Biopolymers 31 (1991) 285.
[61] H. Yu, B.M. Pettitt, M. Karplus, J. Am. Chem. Soc. 113 (1991) 2425.