FARADAY

Solvation thermodynamics of amino acids

Assessment of the electrostatic contribution and force-field dependence

Surjit B. Dixit, R. Bhasin,† E. Rajasekaran‡ and B. Jayaram*§
Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

The free energies of hydration of the 20 amino acids in their zwitterionic form, their pK_a shifts and the side-chain free energies of transfer have been calculated using the finite difference Poisson–Boltzmann methodology. A comparison of the results obtained with charge and size parameters from some popular force fields used in modelling biomolecules is presented. The force fields considered include recent versions of AMBER, CHARMM, CVFF, GROMOS and OPLS, PARSE and an *ab initio*-derived charge set. A general agreement between the theoretical predictions, emerging from each of the parameter sets, and experiment is discernible. A critique on the current status of theoretical studies on amino acids in solution is also advanced.

Proteins play a key role in nearly all biological processes. The basic structural units of proteins are amino acids. The side chains of these building blocks differ in size, shape, charge, hydrogen-bonding capacity, hydrophobicity and chemical reactivity. Individually and collectively, these side chains contribute to the structure and function of proteins. Theoretical and computer simulation studies on the thermodynamic properties of amino acids and the role of electrostatics in particular, in this context, become very important in developing a molecular view of how different residues interact with each other and with solvent and ion atmosphere. Such studies can pave the way for investigations on protein structure, function and conformational stability, nature of active sites of enzymes, steric and electrostatic complementarities in protein–ligand, protein–DNA interactions *etc*. Knowledge of the contribution of the individual amino acids to the electrostatic field and energetics of proteins is of considerable value in designing enzymes with enhanced or altered function and stability. Free energies of transfer of amino acid chains can help predict the stability of different conformations of proteins. Calculations of pK_a shifts of amino acids help in explaining complex titration curves, as well as in elucidating reaction mechanisms involving protons and recognition of proteins *etc*.

The study of inter- and intra-molecular interactions in an aqueous environment is very complex. In recent years, however, computer simulation methods such as Monte Carlo (MC) and molecular dynamics (MD) have been applied to biomolecular problems to study both electrostatic and non-electrostatic effects, but these are computationally expensive. Other methods based on the dielectric continuum solvent approach have been found to be extremely useful in accurately and expeditiously assessing the role of electrostatics in biological processes.[1–13]

Here, we present a study of the electrostatic properties of the 20 amino acids calculated using the finite-difference Poisson–Boltzmann (FDPB) methodology.[14–22] More specifically, we report the electrostatic contribution to (*a*) free energies of solvation of amino acids at pH 7, in their zwitterionic forms, (*b*) free energies of solvation of amino acid side chains as obtained with some recent parameter sets of AMBER,[23] CHARMM,[24] CVFF,[25,26] GROMOS,[27] OPLS,[28] PARSE,[29] *ab initio*-derived charges[30] and (*c*) pK_a shifts in amino acid zwitterions. The results have been compared with available experimental data.

Background

Theoretical studies on the solvation thermodynamics of amino acids have taken a three-fold path. The first involves an adaptation of the statistical mechanical principles in a molecular simulation context. The Metropolis MC,[31] MD[32] and integral equation strategies[33] to obtain free energies of solvation, pK_a shifts *etc*., come under this category. The second approach utilizes classical electrostatics in which analytical or numerical solutions are sought for the Poisson or the Poisson–Boltzmann equation. This leads to a determination of the electrostatic potentials and related properties of the molecular system embedded in a solvent treated as a dielectric continuum and salt as a diffuse ionic cloud. The third involves empirical approaches[34–38] which relate free energies of solvation to a PV or γS_a type term where V and S_a denote excluded volume and accessible surface area, and P and γ are the free energy parameters per unit volume and unit area respectively. The P and γ parameters are calibrated against experiment and several computational procedures exist for the evaluation of excluded volumes[34,35] and accessible surface areas of the molecular system on hand.[39–41]

Structure, energetics and conformational preferences of alanine dipeptide in solution have been investigated by a number of workers *via* computer simulations.[42–47] The results indicate that both the α_R and P_{II} conformations are preferentially stabilized by hydration. Bash *et al.*[48] applied a free energy perturbation method in conjunction with MD simulations to estimate the free energies of hydration of amino acids, which were found to be in good agreement with the available experimental data. Ben-Naim, Ting and Jernigan[49,50] proposed a statistical mechanical treatment to deal with solvation effects on proteins and concluded that solute–solvent hydrogen bonds constituted the largest component of the free energy of solvation. A thorough account of the modelling of the conformations of peptides and proteins both *in vacuo* and *in aquo* was presented recently by Vasquez *et al.*[51] and by Brooks *et al.*[52,53]

Continuum electrostatic theory provides a rational and computationally tractable approach to the problem of the determination of electrostatic fields in and around biological

† Present address: Ranbaxy Research Laboratories, New Delhi 110020, India.
‡ Present address: Department of Chemistry, University of Nebraska, Lincoln, NE 68588, USA.
§ E-mail address: bjayaram@chemistry.iitd.ernet.in

macromolecules since they incorporate the essential electrostatic features of the solvent and macromolecules, the dielectric boundaries in the system, the ionic strength and the locations and magnitude of the charges. The FDPB method is one approach to an accurate assessment of the thermodynamic properties of solutes in solution. In this method, the molecule, together with its environment, is mapped onto a cubic grid and finite-difference equations are set up on the grid. A specification of the sites of charges and a suitable definition of the solvent-accessible surface enveloping these charges enables a numerical determination of the electrostatic potentials everywhere in the system. The electrostatic potentials or fields help in the study of solvation energies,[17] pK shifts,[54–56] binding energies, conformational analyses, enzyme active site studies[1] *etc.*

Solvation thermodynamics of amino acids in the dielectric continuum approach have been examined by several workers. Tanford and Roxby[57] devised a computational method for the calculation of hydrogen ion titration curves of proteins in the framework of Tanford–Kirkwood theory and concluded that, although the major perturbation of the acidic and basic groups of proteins arises from electrostatic interactions between charged sites, an accurate prediction of the pK_a values of individual groups was not feasible. Matthew and co-workers (ref. 6 and references therein) introduced a static accessibility Tanford–Kirkwood model to explain protein titration curves. Further extensions of the Tanford–Kirkwood model were recently reported by Karshikoff[58] for a better modelling of titration curves.

Gilson and Honig[59] attempted to account for the observed pK_a shifts in subtilisin by means of a continuum electrostatic model. They noted that the Poisson–Boltzmann model gave satisfactory results for both the magnitude and ionic strength dependence of the electrostatic interactions. Antosiewicz *et al.*[60] described an accurate approach for the pK_as of ionizable groups in proteins. The accuracy was assessed by a comparison of the computed pK_as with 60 measured pK_as in seven proteins. They suggest that a high protein relative permittivity improves the overall agreement with experiment because it accounts approximately for phenomena which tend to mitigate the pK_a shifts and which are not specifically included in the model.

Sitkoff *et al.*[29] calculated the free energies of hydration of amino acid side chains and other small organic molecules using the dielectric continuum solvent model. They investigated the utility of several available parameter sets for free energy of solvation calculations and examined the feasibility of optimizing force field or *ab initio*-derived parameters through either charge or radius scaling. This led to a new simple set (PARSE) of charge and size parameters specifically for the FDPB method. Free energies of solvation of neutral amino acid side chains were reported recently by Schmidt and Fine[61] with a continuum solvation model employing the CFF'91 force-field parameters. Subsequently, Simonson and Brunger[62] assessed the accuracy of the free energies of solvation estimated from macroscopic continuum theory *via* a calculation of vapour-to-water transfer energies and pK_a shifts of about 17 amino acid side chains.

Still *et al.*[63] gave a semi-analytical treatment of solvation which could be implemented in molecular mechanics and dynamics programs. They demonstrated that the calculated energies of hydration of small molecules were of comparable accuracy to those obtained from contemporary free energy perturbation results. The small molecules studied include amino acid side chains. Lim *et al.*[64] calculated the energies of solvation and pK_as of model ionizable side chains of amino acids using continuum dielectric methods and an integral equation approach. They found that energies of solvation, calculated with both continuum and integral equation methods, agreed well with experiment, but not the pK_a values. They suggested a charge reduction scheme to obtain experimental energies of solvation and pK_a values. They reported that the pK_a changes were very sensitive to the solution conformation.

Several empirical methods have been developed to study the energetics of protein folding. Eisenberg and McLachlan[38] described one such method for calculating the stability of protein structures in water, starting from the atomic coordinates. The basic assumption is that the free energy of hydration of the molecular system can be considered as a sum of the contributions from individual atoms. The contribution of each protein atom to the energy of solvation is estimated as the product of the accessible surface area of the atom and its atomic solvation parameter. Applications of this method include estimates of the relative stability of different protein conformations, estimates of the free energy of binding of ligands to protein and an atomic level description of hydrophobicity and amphiphilicity. Scheraga and co-workers developed an excluded volume approach[34] to estimate solvent effects on conformational stability and flexibility of peptides. They also described a method based on accessible surface areas[65] for the inclusion of the effects of hydration in empirical conformational energy computations on polypeptides. They evaluated the constants of proportionality, representing the free energies of hydration per unit area of accessible surface, for seven classes of atoms/groups (present in peptides) by least-squares fitting to experimental free energies of solvation of small monofunctional aliphatic and aromatic molecules. Sternberg *et al.*[66] reported an algorithm for the prediction of electrostatic effects in modelling pK_a shifts.

As evident from the series of studies cited above, the solvation thermodynamics of amino acid zwitterions has not received much attention. In this study, we address this lacuna with a state of the art methodology for determining electrostatic properties, namely the FDPB method, and relate the results to classical electrostatic models developed by Onsager[67,68] and Kirkwood and co-workers.[69–73] Also, a number of new parameter sets were reported recently for modelling amino acids and proteins.[23–30] Some of these appeared in the literature while this work was in progress or nearing completion. This gave us an opportunity to assess the performance of the diverse force fields in predicting the electrostatic properties of amino acid side chains in aqueous media.

Calculations

The electrostatic contribution to the free energies of solvation of all the amino acids in their zwitterionic form is calculated using the FDPB method.[15,16,25] The procedure involves a specification of the Cartesian coordinates of each of the atoms in the molecule, their partial atomic charges and sizes, relative permittivity inside and outside the molecule, ionic strength *etc.*, and seeking a numerical solution to the PB equation. A resolution of 0.25 Å per grid was employed in all the FDPB calculations. The principal property evaluated is the electrostatic potential from which other properties ensue as given below.

The total electrostatic energies of the system in vacuum A_v ($\varepsilon = 1$) and in water A_w ($\varepsilon = 80$) are computed and hence the electrostatic contribution to the free energy of solvation ΔA_{sol} is obtained as

$$\Delta A_{sol}(\text{elec}) = (A_w - A_v)$$

A_v represents the energy to discharge the solute in vacuum and A_w is the energy to recharge the solute in solvent water. The reference state is thus the fully charged molecular system in vacuum. The process by which the ΔA_{sol}(elec) is computed is equivalent to estimating the work done in discharging the molecular species in vacuum and charging it up in solvent and hence the difference between the two total electrostatic ener-

gies, A_v and A_w, is identifiable with the Helmholtz energy of solvation of the molecular system.[73,74]

The atomic coordinates were generated and optimized *in vacuo* using the Biosym software.[25] The protocol followed involved assignment of parameters from the CVFF force field and further energy minimization to optimize the structure. This was done by sequentially performing minimization using steepest descent, conjugate gradient and Newton–Raphson algorithms until the maximum derivative was less than 0.01 kcal $Å^{-1}$ or the total number of iterations was 1000. The structures so obtained were employed for all the calculations reported here. The effect of the solvent on the structure was also considered and discussed below where appropriate.

In the amino acid zwitterion free energy of solvation and the pK_a shift calculations, a simple formal charge model was employed, *i.e.* the positive charge on the amino group was distributed over the three hydrogen atoms attached to the *N*-terminal nitrogen and the negative charge was distributed over the two carboxyl oxygens of *C*-terminal carbons. The inner and outer relative permittivities were taken to be 2 and 80, respectively.

The pK_a shifts or the effective pK_a values of the α-CO_2H group and α-NH_3^+ group of each amino acid are estimated in relation to a suitable reference system.

$$pK_{eff} = pK_{int} + \Delta pK_a$$

Here pK_{int} is related to the intrinsic equilibrium constant in the absence of other charged sites and pK_{eff} is the modified value in the presence of coupling factors.[75] ΔpK_a is calculated from the formula

$$\Delta pK_a = \frac{\phi}{2.303} = \frac{(\phi_1 + \phi_2)/2}{2.303}$$

where ϕ is the potential in kT e^{-1} units[59] at the target site of the functional group undergoing dissociation, due to the presence of other charged groups or atoms, responsible for the shift in pK_a values. The mean potential at the target is taken as in the acetate group, which requires a consideration of the potential at both oxygens.

In the side chain free energy of solvation calculations, the main chain atoms were not considered but the Cα carbon was replaced by a hydrogen atom. The side chain of proline cannot be represented in this manner and hence has been omitted. Our attempt here has been to characterize the charge and radius parameters of the current force fields in use for biomolecular modelling, with regard to their ability to model the electrostatic properties of the side chains. The parameter sets considered were adapted from AMBER,[23] CHARMM,[24] CVFF,[25,26] GROMOS,[27] OPLS[28] and PARSE.[29] *Ab initio*-derived charges of Chipot *et al.*[30] were also tested with radii from AMBER. The radii were calculated from the non-bonded interaction parameters and correspond to the van der Waals radii at which the non-bonded interaction energy is zero rather than the radii at the minimum of the potential well.

To enable a comparison with experiment, both electrostatic and hydrophobic contributions are required. Hydrophobic contributions were determined from a least-squares fit equation relating experimental energy of solvation[76] of branched and linear alkanes to their accessible surface areas (S_a). The S_as of the hydrocarbons were calculated with each of the force-field radii set separately. A line of regression for the solvation energy as a function of the accessible surface area of the form

$$\Delta A_{h\phi} = bS_a + a$$

was obtained, where b and a are constants, the slope and intercept respectively, for the radii set. S_a of the amino acids were calculated with each force-field radii set and substituted in the above equation to obtain the respective hydrophobic contributions to the free energies of solvation. Several hydrophobicity scales exist for refined estimates.[77] We chose a very simple scheme.

Results and Discussion

Zwitterion solvation

The electrostatic contribution to the free energy of solvation of all the 20 amino acids in their zwitterionic forms, *i.e.* at pH 7, calculated by the FDPB method are given in Table 1. All the zwitterion free energies of hydration fall within the range of −69.51 to −82.95 kcal mol^{-1} for amino acids with neutral side chains, while these values are almost double for amino acids with charged side chains, *i.e.* for arginine(+), lysine(+), aspartate(−) and glutamate(−). These results can be justified if we consider the reaction field approach due to Onsager for the solvation energy of a dipole embedded in a spherical cavity[67,68]

$$\Delta A_{sol} = \left(\frac{1-\varepsilon}{1+2\varepsilon}\right)\frac{\mu^2}{a_0^3}$$

Here ΔA_{sol} is the solvation energy, ε is the relative permittivity of the solvent water, μ is the dipole moment associated with the molecule, and a_0 is the diameter of the low dielectric spherical cavity containing the dipole. All the amino acids in their zwitterionic form may be considered as dipoles embedded in a spherical cavity surrounded by the solvent. This equation gives a quick numerical estimate of the Helmholtz energy of dipolar solvation. The solvation energy for glycine is −59 kcal mol^{-1} (for $r = 3.22$ Å, $a_0 = 3.07$ Å, $\varepsilon = 80$). We have calculated this by assuming that glycine is embedded in a spherical cavity of diameter a_0 which is estimated by taking the average of all diagonal distances of glycine. The dipole moment is evaluated as $\mu = er$, where r is the dipole length, here taken to be the distance between the positive and negative charge centres in the molecule. Since μ is almost the same for all amino acids, it is only the a_0 factor which causes variations in the A_{sol} values. For glycine, a_0 is the smallest, and is expected to give the maximum value (−82.95 kcal mol^{-1}, Table 1), while others show variation due to different a_0 values. In the FDPB method, molecules are not considered to be embedded in a spherical cavity, but are embedded in a cavity formed by their solvent accessible surface. This is one reason for the larger values (Table 1) than expected from

Table 1 Electrostatic contribution to the solvation free energy (in kcal mol^{-1}) of amino acid zwitterions

	ΔA_{sol}(elec)
ALA	−77.40
ARG	−157.49
ASN	−73.62
ASP	−119.47
CYS	−74.42
GLN	−82.02
GLU	−112.74
GLY	−82.95
HID	−76.74
HIE	−71.80
HIP	−140.83
ILE	−69.51
LEU	−81.13
LYS	−132.27
MET	−72.85
PHE	−72.71
PRO	−75.18
SER	−77.15
THR	−74.43
TRP	−74.11
TYR	−73.40
VAL	−71.30

Onsager's simple analytical theory. Overall, trends expected from Onsager's theory and those obtained from the FDPB studies are qualitatively similar. The FDPB method of course involves rigorous numerical calculations, so improved estimates are expected from the latter.

Structures obtained employing solvent conditions during minimization were also used to calculate the solvation energies to capture the effects of structural relaxation in solvent. The results obtained for these structures are *ca.* 15% more negative with the formal charge model than with structures corresponding to vacuum conditions. The solvation energies of amino acids with charged side chains are even more negative (*ca.* 35%). A set of FDPB calculations for the structures obtained both in vacuum and solvent was also performed with the charges and radii corresponding to the CVFF force field. Though the calculated values with these parameters are less negative than those obtained with the formal charge model, the trends in the solvation energies of the structures obtained in vacuum and water remains the same as above.

Solvation energy calculations of zwitterions can be configured to estimate the desolvation contribution in the peptide bond formation. The computed solvation energy of diglycine zwitterion is around -131 kcal mol^{-1}. This, taken together with the glycine zwitterion solvation energy in Table 1 (-83 kcal mol^{-1}) implies that the desolvation of individual zwitterions resulting in a peptide bond costs *ca.* $+35$ kcal mol^{-1}. Quantitatively more accurate estimates are expected with better charge distributions employed together with a consideration of the entropic contribution of the released water molecules.

pK_a shifts

pK_a shifts yield valuable information on intramolecular interactions affecting the ionization equilibrium of the functional group of interest. Calculating the pK_a values of small molecules has been a problem of long-standing interest. Kirkwood, Westheimer and their co-workers[70,72] used classical electrostatics to approach the problem. Lack of a proper geometric description of the molecular system and uncertainties with regard to the spatial distribution of the charges limited the accuracy of the results that were obtained. The availability of improved structural data on amino acids and numerical techniques to solve for electrostatic properties, facilitate a reinvestigation of this problem. The calculated pK_a shifts for amino acids are given in Table 2. No direct experimental values are available for an assessment of the results. This problem can be circumvented by choosing some appropriate common reference system for all the 20 amino acids. Here, amino acids are considered as substituted acetic acids in which the two Hs on the CH_3 group have been replaced by $-NH_3^+$ and a side chain characteristic of each amino acid. The effect of this positively charged $-NH_3^+$ group on the acidity of acetic acid is calculated. Acetic acid has a pK_a of 4.8.[78] By pK_a shift is meant here how different substituents on the $-CH_3$ group in acetic acid influence its intrinsic acidity and shift its pK_a value.

For the reaction $HA \rightarrow H^+ + A^-$

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

$$\Delta G^0 = -RT \ln K_a = 2.303RT\mathrm{p}K_a$$

Suppose that the ionization reaction is coupled to some other process or interaction[75] as in this case, interaction of an α-NH_3^+ group with an α-CO_2^-. The Gibbs energy change involving this interaction is given by ΔG_c, where subscript c refers to coupling. The apparent pK_a is now

$$\mathrm{p}K_a = \frac{\Delta G^0 + \Delta G_c}{2.303RT}$$

The effect of this ΔG_c is to shift the pK_a. Depending on whether ΔG_c is negative or positive, pK_as are shifted to lower or higher values, respectively. In our case, ΔG_c is negative, owing to the interaction of the positive charge on $-NH_3^+$ with the negative charge on the $-CO_2^-$ which facilitates the release of H^+ from the $-CO_2H$ group, resulting in a decrease in pK_a values. We have calculated this electrostatic contribution (Table 2) and results are generally in good agreement with the pK_a shifts derived from the reported experimental values.[75] Deviations, if any, may be due to the choice of the reference system and small numerical errors associated with a grid representation of the molecule.

Note that the quality of the results obtained with simpler alternatives to FDPB method such as a uniform dielectric model with $\varepsilon = 80$ everywhere is rather poor. The expected

Table 2 Calculated pK shifts

	α-CO_2H[a]	α-NH_3^+[a]		
	experiment		$\Delta\mathrm{p}K_1(CO_2H \rightarrow CO_2^-)$[b]	
	pK_1	pK_2	expected	calculated (FDPB)
ALA	2.3	9.9	2.5	2.7
ARG	1.8	9.0	3.0	3.3 (2.9)[c]
ASN	2.0	8.8	2.8	2.5
ASP	2.0	10.0	2.8	1.9 (2.6)
CYS	1.8	10.8	3.0	2.9
GLN	2.2	9.1	2.6	2.0
GLU	2.2	9.7	2.6	1.6 (2.2)
GLY	2.4	9.8	2.4	1.9
HIS	1.8	9.2	3.0	2.8
ILE	2.4	9.7	2.4	3.0
LEU	2.4	9.6	2.4	2.0
LYS	2.2	9.2	2.6	9.5 (2.3)
MET	2.3	9.2	2.5	2.8
PHE	1.8	9.1	3.0	2.8
PRO	2.0	10.6	2.8	4.4
SER	2.1	9.2	2.7	2.3
THR	2.6	10.4	2.2	2.7
TRP	2.4	9.4	2.4	2.7
TYR	2.2	9.1	2.6	2.7
VAL	2.3	9.6	2.5	2.8

[a] Ref. 75; [b] $\Delta\mathrm{p}K_1$ expected = pK_a of acetic acid (4.8) − pK_1 of the amino acid, the magnitudes of the $\Delta\mathrm{p}K_2$ expected values are to be computed as 8.0 − pK_2 of amino acid; [c] values in brackets correspond to calculations with uncharged side chains.

Table 3 Solvation free energies of amino acid side chains with AMBER parameters (in kcal mol^{-1})

	molecule	$\Delta A_{elec(FDPB)}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−19.07	2.04	−17.03	−9.70	−7.33
CYS	methylthiol	−4.00	1.92	−2.08	−1.24	−0.84
GLN	propionamide	−20.65	2.20	−18.45	−9.38	−9.07
HID	methylimidazole	−13.91	2.25	−11.66	−10.27	−1.39
HIE	methylimidazole	−11.62	2.25	−9.37	−10.27	+0.90
SER	methanol	−8.80	1.83	−6.97	−5.06	−1.91
THR	ethanol	−9.00	2.02	−6.98	−4.88	−2.10
TRP	methylindole	−10.58	2.62	−7.96	−5.88	−2.08
TYR	*p*-cresol	−9.31	2.46	−6.85	−6.11	−0.74
ARG	*N*-propylguanidinium ion	−88.07	2.49	−85.58		
ASP	acetate ion	−79.36	2.01	−77.35	−80.65	+3.30
GLU	propionate ion	−77.46	2.18	−75.28	−79.12	+3.84
HIP	methylimidazolium ion	−74.20	2.31	−71.89	−64.13	−7.76
LYS	*N*-butyl ammonium ion	−83.31	2.34	−80.97	−69.24	−11.73
ALA	methane	−0.11	1.79	1.68	1.94	−0.26
ILE	butane	−0.25	2.28	2.03	2.15	−0.12
LEU	isobutane	−0.45	2.27	1.82	2.28	−0.46
MET	methyl ethyl sulfide	−1.87	2.28	0.41	−1.48	+1.89
PHE	toluene	−2.47	2.42	−0.05	−0.76	+0.71
VAL	propane	−0.24	2.14	1.90	1.99	−0.09
					mean unsigned error	2.97

[a] $\Delta A_{h\phi} = (0.004\,83 \pm 0.000\,61)S_a + (1.044\,31 \pm 0.179\,94)$; [b] ref. 82, 29.

pK_a shift in glycine with a distance of charge separation of *ca.* 3.2 Å and $\varepsilon = 80$ is 1.3 units which is off by 1.1 units from the corresponding experimental value. This once again indicates that a good description of the electrostatics requires proper treatment of the shape of the molecule and dielectric inhomogeneities in the system[1] at the continuum solvent level.

The effect of the carboxylate group on the dissociation of the $-NH_3^+$ group has also been investigated. The calculated pK_a shifts for the α-NH_3^+ dissociation are of course identical to the computed pK_a shifts for the α-CO_2H dissociation. To compare the results with experiment or, more specifically, to arrive at the expected p*K*a shifts from experiment, one needs a suitable reference system as above. Analogous to acetic acid, methyl amine suggests itself as a reference system here. Coulombic interaction between the charged terminals in a zwitterion must favour the existence of the amino acid in the zwitterion form and hence shift the pK_a of the amino terminal to a value greater than 10.6, which is the pK_a of methyl amine.[79] Experimental pK_as for *N*-terminal dissociation of the amino acids (Table 2) indicate a different effect. The pK_as are actually less than 10.6. Obviously methyl amine constitutes a poor choice for the reference system. Another possible reference system is α-carbonyl substituted methyl amine[75] for which the pK_a is around 8.0. The estimated pK_a shifts with this reference system also deviate considerably from the expected values but are better than those with methyl amine as the reference. The good agreement that was noticed for the $-CO_2H$ dissociation is not paralleled by the $-NH_3^+$ group dissociation. A suitable reference system is not available for the latter. An interesting feature of the experimental pK_a values is their spread. While the total spread among the 20 amino acids for the $-CO_2H$ dissociation is only 0.6 units, it is 2.0 units for the $-NH_3^+$ dissociation. Thus, any single reference system for the $-NH_3^+$ dissociation is bound to show large deviations between the calculated and expected values.

The pK_a shift calculations were also repeated with structures corresponding to solvent conditions. The magnitude of the calculated pK_a shifts is smaller by *ca.* 35% for the $-CO_2H$ dissociation and the structures corresponding to vacuum conditions give better results. For the $-NH_3^+$ dissociation, on the other hand, solvent structures yield relatively better results. Subsequently, these calculations were repeated once again with the charge and radii parameters from the CVFF force field instead of the formal charge model. For the $-CO_2H$ dissociation, the formal charge model, together with the structures minimized in vacuum, constitutes the best model. For the $-NH_3^+$ dissociation also, the formal charge model with the structure in water yields better results in most cases. This anomalous behaviour probably needs to be explained on the basis of explicit solvent structure.[80,81]

These results, although calculated for independent amino acids, are very significant in that they provide basic model studies which can be extended to proteins and enzymes, to appreciate contextual effects. The control of residual pK_a values may provide a way to regulate the behaviour of charged residues at the active site of the enzymes, *i.e.* reactivity of certain significant residues may be controlled by replacing other groups, which increase or decrease the acidity *via* coupling effects. A simple analytical theory in the Tanford–Kirkwood style[71] was employed earlier by Westheimer and Shookhoff[72] to determine the charge separation in some amino acids and peptides using the experimentally determined pK_a shifts. The estimated distances of charge separation of glycine and alanine zwitterions were 4.05 and 3.85 Å.[70] The true values are closer to 3.22 and 3.03 Å.[25,70] The current work, though employing similar concepts, gives much more precise results on electrostatic effects, indicating the strength of the numerical technique involved.

Side chain solvation

In proteins, the zwitterionic character of amino acids is lost, and hence a study of the side chain solvation energies becomes important. The solvation energies for amino acid side chains, calculated using different force-field parameters *via* the FDPB method are given in Tables 3 to 9. Experimental results from dynamic vapour pressure distribution studies[82,83] are available for comparison. These studies of Wolfenden and coworkers on the solvation energies of amino acids have been reference points for several subsequent theoretical investigations. Though hydrophobic contributions have been included for an effective comparison with experimental values, these results are a reflection of the importance of electrostatic contribution to the total solvation free energy.

Overall, all the force fields show good correlation between the expected electrostatic contribution and the experimental values yielding correlation coefficients above 0.99 in all cases, making a choice of the 'best' force field for modelling electrostatics on a statistical basis difficult. This reiterates the common experience that most of the force fields work well

Table 4 Solvation free energies of amino acid side chains with CHARMM parameters (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−11.42	2.06	−9.36	−9.70	+0.34
CYS	methylthiol	−3.05	1.94	−1.11	−1.24	+0.13
GLN	propionamide	−11.29	2.22	−9.07	−9.38	+0.31
HID	methylimidazole	−14.06	2.24	−11.82	−10.27	−1.55
HIE	methylimidazole	−13.19	2.24	−11.15	−10.27	−0.88
SER	methanol	−8.59	1.86	−6.73	−5.06	−1.67
THR	ethanol	−8.77	2.03	−6.74	−4.88	−1.86
TRP	methylindole	−7.07	2.62	−4.45	−5.88	+1.43
TYR	*p*-cresol	−8.77	2.53	−6.24	−6.11	−0.13
ARG	*N*-propylguanidinium ion	−82.83	2.50	−80.33		
ASP	acetate ion	−82.29	2.02	−80.27	−80.65	+0.38
GLU	propionate ion	−82.27	2.19	−80.08	−79.12	−0.96
HIP	methylimidazolium ion	−73.18	2.21	−70.97	−64.13	−6.84
LYS	*N*-butyl ammonium ion	−85.47	2.36	−83.11	−69.24	−13.87
ALA	methane	−0.29	1.78	1.49	1.94	−0.45
ILE	butane	−0.57	2.26	1.69	2.15	−0.46
LEU	isobutane	−0.52	2.25	1.73	2.28	−0.55
MET	methyl ethyl sulfide	−1.16	2.29	1.13	−1.48	+2.61
PHE	toluene	−2.00	2.42	0.42	−0.76	+1.18
VAL	propane	−0.46	2.13	1.67	1.99	−0.32
					mean unsigned error	1.89

[a] $\Delta A_{h\phi} = (0.004\,92 \pm 0.000\,61)S_a + (1.059\,23 \pm 0.177\,07)$; [b] ref. 82, 29.

with amino acids and proteins. To help the fine-tuning exercises in subsequent versions of the force fields, we point out some systems that show significant deviations. Our observations are of course limited by the accuracies of the hydrophobicity estimates based on the accessible surface area model, especially when the electrostatic contributions are small. As already indicated, the calculated electrostatic contributions refer to Helmholtz energies, while the hydrophobicity estimates and experimental solvation energies refer to the Gibbs energies. The PV correction has been neglected in the electrostatic contributions when comparing the calculated solvation energies with the experimental values. This is not expected to alter qualitatively the conclusions on the relative performance of the force fields. Also a certain amount of error is always associated with any molecular model and the numerical technique employed, although a conscious attempt has been made to minimize such errors.

The calculated solvation energies of the non-polar amino acids with the AMBER force-field parameters (Table 3) are fairly accurate except for Met and Phe. The estimated electrostatic contributions for some polar and charged amino acid side chains are in excess of the expected values based upon experiment. Notable among these are the electrostatic contributions of Asn, Gln, Hip and Lys. Also the electrostatic contributions of Hid and Hie are quite different though they are expected to have the same solvation energies.

The unsigned errors in the calculated solvation energies are much less with the CHARMM force-field parameters (Table 4), but here too the electrostatic contributions of Lys and Hip are exaggerated while those of Met and Phe may be considered as underestimates. The other non-polar amino acids have strong electrostatic contributions.

The *ab initio*-derived charges provide fairly accurate estimates of the solvation energies (Table 5) of the non-polar amino acids. Met and Phe are better modelled with these charges. The electrostatic contributions for all the polar and charged amino acids are overestimated. Here too Asn, Gln, Hip and Lys have very large electrostatic contributions.

The electrostatics of charged amino acid side chains are well modelled by the CVFF force-field parameters (Table 6). Prominent among the results with these parameters are the electrostatic contributions calculated for Hid and Hie, which

Table 5 Solvation free energies of amino acid side chains with *ab initio*-derived charges and AMBER radii (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−19.51	2.04	−17.47	−9.70	−7.77
CYS	methylthiol	−5.25	1.92	−3.33	−1.24	−2.09
GLN	propionamide	−19.94	2.20	−17.74	−9.38	−8.36
HID	methylimidazole	−14.34	2.25	−12.09	−10.27	−1.82
HIE	methylimidazole	−14.80	2.25	−12.55	−10.27	−2.28
SER	methanol	−8.76	1.83	−6.93	−5.06	−1.87
THR	ethanol	−9.33	2.02	−7.31	−4.88	−2.43
TRP	methylindole	−12.33	2.62	−9.71	−5.88	−3.83
TYR	*p*-cresol	−10.84	2.46	−8.38	−6.11	−2.27
ARG	*N*-propylguanidinium ion	−89.65	2.49	−87.16		
ASP	acetate ion	−84.58	2.01	−82.57	−80.65	−1.92
GLU	propionate ion	−81.83	2.18	−79.65	−79.12	−0.53
HIP	methylimidazolium ion	−73.18	2.31	−70.87	−64.13	−6.74
LYS	*N*-butyl ammonium ion	−82.19	2.34	−79.85	−69.24	−10.61
ALA	methane	−0.35	1.79	1.44	1.94	−0.50
ILE	butane	−0.08	2.28	2.20	2.15	+0.05
LEU	isobutane	−0.55	2.27	1.72	2.28	−0.56
MET	methyl ethyl sulfide	−3.45	2.28	−1.17	−1.48	+0.31
PHE	toluene	−3.04	2.42	−0.62	−0.76	+0.14
VAL	propane	−0.11	2.14	2.03	1.99	+0.04
					mean unsigned error	2.85

[a] $\Delta A_{h\phi} = (0.004\,83 \pm 0.000\,61)S_a + (1.044\,31 \pm 0.179\,94)$; [b] ref. 82, 29.

Table 6 Solvation free energies of amino acid side chains with CVFF parameters (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−6.89	2.06	−4.83	−9.70	+4.87
CYS	methylthiol	−1.53	1.92	0.39	−1.24	+1.63
GLN	propionamide	−15.36	2.22	−13.14	−9.38	−3.76
HID	methylimidazole	−3.76	2.28	−1.48	−10.27	+8.79
HIE	methylimidazole	−3.81	2.28	−1.53	−10.27	+8.74
SER	methanol	−8.14	1.84	−6.30	−5.06	−1.24
THR	ethanol	−8.11	2.02	−6.09	−4.88	−1.21
TRP	methylindole	−3.72	2.63	−1.09	−5.88	+4.79
TYR	*p*-cresol	−8.53	2.47	−6.06	−6.11	+0.05
ARG	*N*-propylguanidinium ion	−67.41	2.54	−64.87		
ASP	acetate ion	−85.45	2.01	−83.44	−80.65	−2.79
GLU	propionate ion	−84.38	2.18	−82.20	−79.12	−3.08
HIP	methylimidazolium ion	−63.85	2.28	−61.57	−64.13	+2.56
LYS	*N*-butyl ammonium ion	−77.18	2.38	−74.80	−69.24	−5.56
ALA	methane	−0.21	1.79	1.58	1.94	−0.36
ILE	butane	−0.64	2.27	1.63	2.15	−0.52
LEU	isobutane	−0.57	2.26	1.69	2.28	−0.59
MET	methyl ethyl sulfide	−0.53	2.30	1.77	−1.48	+3.25
PHE	toluene	−1.61	2.44	0.83	−0.76	+1.59
VAL	propane	−0.53	2.14	1.61	1.99	−0.38
					mean unsigned error	2.93

[a] $\Delta A_{h\phi} = (0.004\,81 \pm 0.000\,61)S_a + (1.076\,76 \pm 0.177)$; [b] ref. 82, 29.

are highly underestimated. It is interesting to observe that while the electrostatic contribution of Asn is underestimated, that of Gln is overestimated. Met, Phe and Trp have smaller electrostatic contributions relative to experiment. CVFF too overestimates the electrostatics of non-polar amino acid side chains. A new version of this force field, namely CFF'91, was examined by Schmidt and Fine.[61] The mean unsigned error for the solvation energy of uncharged amino acids was reported to be 0.3.

In comparison to the other force fields discussed above, the estimates of the electrostatic contributions for Lys and Hip with GROMOS parameters (Table 7) are better, but Asp and Glu show rather large electrostatic contributions. Unlike the other force fields these parameters underestimate the electrostatics of all the polar amino acid side chains. The force field prescribes no charges to the atoms in the non-polar amino acids.

While the negatively charged amino acids are well modelled with the OPLS parameters (Table 8), the positive side chains namely Hip and Lys have overestimated electrostatic contributions. Here too, the electrostatic contributions of polar side chains are overemphasized. Trp and Phe should have larger electrostatic contributions.

The PARSE parameter set has been developed by a scaling of other charge and radii sets to predict the solvation energies of amino acid side chains accurately (Table 9) and, as might be expected, the results are in good accord with experiment in relation to the other force-field parameters investigated here.

Net unsigned error in the calculated solvation energies is the least with the PARSE parameters. Among the force fields, solvation energy calculations with CHARMM results in the least unsigned errors. The solvation energies obtained with the force fields in increasing order of net unsigned error is as follows

Overall: PARSE, CHARMM, GROMOS, OPLS, *ab initio*, CVFF, AMBER
Polar: PARSE ~ CFF'91, CHARMM, OPLS, GROMOS, AMBER, *ab initio*, CVFF
Charged: PARSE, CVFF, GROMOS, *ab initio*, CHARMM, OPLS, AMBER

Table 7 Solvation free energies of amino acid side chains with GROMOS parameters (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−7.27	2.06	−5.21	−9.70	+4.49
CYS	methylthiol	−1.68	1.94	+0.26	−1.24	+0.98
GLN	propionamide	−7.06	2.23	−4.83	−9.38	+4.55
HID	methylimidazole	−11.42	2.25	−9.17	−10.27	+1.10
HIE	methylimidazole	−11.38	2.25	−9.13	−10.27	+1.14
SER	methanol	−4.47	1.86	−2.67	−5.06	+2.39
THR	ethanol	−4.12	2.03	−2.09	−4.88	+2.79
TRP	methylindole	−5.73	2.63	−3.10	−5.88	+2.78
TYR	*p*-cresol	−5.12	2.48	−2.64	−6.11	+3.47
ARG	*N*-propylguanidinium ion	−60.51	2.57	−57.94		
ASP	acetate ion	−89.20	1.96	−87.24	−80.65	−6.59
GLU	propionate ion	−87.92	2.14	−85.78	−79.12	−6.66
HIP	methylimidazolium ion	−68.19	2.25	−65.94	−64.13	−1.81
LYS	*N*-butyl ammonium ion	−74.62	2.43	−72.19	−69.24	−2.95
ALA	methane	0.00	1.84	1.84	1.94	−0.10
ILE	butane	0.00	2.37	2.37	2.15	+0.22
LEU	isobutane	0.00	2.36	2.36	2.28	+0.08
MET	methyl ethyl sulfide	0.00	2.38	2.38	−1.48	+3.86
PHE	toluene	0.00	2.54	2.54	−0.76	+3.30
VAL	propane	0.00	2.22	2.22	1.99	+0.23
					mean unsigned error	2.60

[a] $\Delta A_{h\phi} = (0.004\,83 \pm 0.000\,60)S_a + (1.043\,04 \pm 0.179\,04)$; [b] ref. 82, 29.

Table 8 Solvation free energies of amino acid side chains with OPLS parameters (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−15.36	2.09	−13.27	−9.70	−3.57
CYS	methylthiol	−6.74	2.06	−4.68	−1.24	−3.44
GLN	propionamide	−15.43	2.24	−13.19	−9.38	−3.81
HID	methylimidazole	−10.31	2.30	−8.01	−10.27	+2.26
SER	methanol	−9.09	1.87	−7.22	−5.06	−2.16
THR	ethanol	−9.07	2.03	−7.04	−4.88	−2.16
TRP	methylindole	−6.62	2.67	−3.95	−5.88	+1.93
TYR	*p*-cresol	−8.88	2.51	−6.37	−6.11	−0.26
ARG	*N*-propylguanidinium ion	−82.18	2.54	−79.64		
ASP	acetate ion	−82.48	2.06	−80.42	−80.65	+0.23
GLU	propionate ion	−81.84	2.22	−79.62	−79.12	−0.50
HIP	methylimidazolium ion	−72.88	2.30	−70.58	−64.13	−6.45
LYS	*N*-butyl ammonium ion	−88.19	2.42	−85.77	−69.24	−16.53
ALA	methane	0.00	1.80	1.80	1.94	−0.14
ILE	butane	0.00	2.26	2.26	2.15	+0.11
LEU	isobutane	0.00	2.25	2.25	2.28	−0.03
MET	methyl ethyl sulfide	−3.34	2.12	−1.22	−1.48	+0.26
PHE	toluene	0.00	2.44	2.44	−0.76	+3.20
VAL	propane	0.00	2.12	2.12	1.99	+0.13
					mean unsigned error	2.62

[a] $\Delta A_{h\phi} = (0.005\,31 \pm 0.000\,65)S_a + (1.046\,23 \pm 0.175\,08)$; [b] ref. 82, 29.

Table 9 Solvation free energies of amino acid side chains with PARSE parameters (in kcal mol^{-1})

	molecule	$\Delta A_{\text{elec(FDPB)}}$	$\Delta A_{h\phi(S_a)}$ [a]	ΔA_{tot}	ΔA_{expt} [b]	error
ASN	acetamide	−12.29	2.07	−10.22	−9.70	−0.52
CYS	methylthiol	−3.37	1.98	−1.39	−1.24	−0.15
GLN	propionamide	−12.02	2.23	−9.79	−9.38	−0.41
HID	methylimidazole	−12.84	2.26	−10.58	−10.27	−0.31
SER	methanol	−7.45	1.86	−5.59	−5.06	−0.53
THR	ethanol	−7.16	2.01	−5.15	−4.88	−0.27
TRP	methylindole	−8.76	2.63	−6.13	−5.88	−0.25
TYR	*p*-cresol	−8.91	2.47	−6.44	−6.11	−0.33
ARG	*N*-propylguanidinium ion	−69.45	2.54	−66.91		
ASP	acetate ion	−83.14	2.02	−81.12	−80.65	−0.47
GLU	propionate ion	−81.84	2.08	−79.76	−79.12	−0.64
HIP	methylimidazolium ion	−66.99	2.29	−64.70	−64.13	−0.57
LYS	*N*-butyl ammonium ion	−73.49	2.42	−71.07	−69.24	−1.83
ALA	methane	0.00	1.79	1.79	1.94	−0.15
ILE	butane	0.00	2.26	2.26	2.15	+0.11
LEU	isobutane	0.00	2.25	2.25	2.28	−0.03
MET	methyl ethyl sulfide	−3.77	2.11	−1.66	−1.48	−0.18
PHE	toluene	−3.24	2.40	−0.84	−0.76	−0.08
VAL	propane	0.00	2.12	2.12	1.99	+0.13
					mean unsigned error	0.39

[a] $\Delta A_{h\phi} = (0.005\,24 \pm 0.000\,64)S_a + (1.033\,54 \pm 0.176\,83)$; [b] ref. 82, 29.

Non-polar: PARSE, *ab initio*, AMBER, CFF'91, OPLS, CHARMM, CVFF, GROMOS

The above force-field comparisons have two immediate implications. First, the parameters for the 'deviant' amino acids can be fine-tuned in each force field. Secondly, the method employed seems to slightly overestimate the electrostatics in most cases. A higher solute relative permittivity could be employed in the electrostatic contribution calculations as suggested by Antosiewicz *et al.*,[60] for quantitative results.

Conclusions

The electrostatic contribution to the solvation energies of zwitterions of all the 20 amino acids has been calculated. The results are consistent with the expectations based on a simple analytical theory due to Onsager. Electrostatic coupling interactions causing the pK_a shift of the CO_2H group in all 20 amino acid zwitterions have been studied with the pK_a of acetic acid as a reference, with satisfactory results. Electrostatic contributions to the transfer free energies of the side chains were estimated with charges and radii from AMBER, CHARMM, CVFF, GROMOS, OPLS, PARSE and an *ab initio*-derived charge set. Unsigned errors in the calculated solvation free energies are the least with the PARSE parameter set, followed by the CHARMM force-field parameters. In summary, the results are reflective of the role of electrostatics in solvation and ionization equilibria of amino acids and also the power and utility of numerical procedures in evaluating the thermodynamic properties of molecular systems.

Financial support received from the Department of Science and Technology, India, is gratefully acknowledged.

References

1 B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144.
2 S. Harvey, *Proteins: Struct., Funct., Genet.*, 1989, **5**, 78.
3 K. Sharp and B. Honig, *Annu. Rev. Biophys. Chem.*, 1990, **19**, 301.
4 M. E. Davis and J. A. McCammon, *Chem. Rev.*, 1990, **90**, 509.
5 N. M. Allewell and H. Oberoi, *Meth. Enzym.*, 1991, **202**, 3.
6 J. B. Matthew, *Annu. Rev. Biophys. Biophys. Chem.*, 1985, **14**, 387.
7 N. K. Rogers, *Prog. Biophys. Mol. Biol.*, 1986, **48**, 37.
8 D. Bashford, *Curr. Opin. Struct. Biol.*, 1992, **2**, 40.
9 M. K. Gilson, *Curr. Opin. Struct. Biol.*, 1995, **5**, 216.

10 A. Warshel and S. Russell, *Q. Rev. Biophys.*, 1984, **17**, 283.
11 C. J. Cramer and D. G. Truhlar, *Science*, 1992, **256**, 213.
12 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027.
13 S. L. Price, *J. Chem. Soc., Faraday Trans.*, 1996, **92**, 2997.
14 J. Warwicker and H. C. Watson, *J. Mol. Biol.*, 1982, **157**, 671.
15 I. Klapper, R. Hagstrom, R. Fine, K. Sharp and B. Honig, *Proteins: Struct., Funct., Genet.*, 1986, **1**, 47.
16 M. K. Gilson, K. A. Sharp and B. Honig, *J. Comput. Chem.*, 1988, **9**, 327.
17 M. K. Gilson and B. Honig, *Proteins: Struct., Funct., Genet.*, 1988, **4**, 7.
18 B. Jayaram, K. A. Sharp and B. Honig, *Biopolymers*, 1989, **28**, 975.
19 B. A. Luty, M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, 1992, **13**, 1114.
20 H. Oberoi and N. Allewell, *Biophys. J.*, 1993, **65**, 48.
21 M. Holst and F. Saied, *J. Comput. Chem.*, 1993, **14**, 105.
22 M. Holst, R. E. Kozack, F. Saied and S. Subramanian, *Proteins: Struct., Funct., Genet.*, 1994, **18**, 231.
23 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
24 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187. and A. D. MacKerell Jr., CHARMM parameter set for proteins Version 22, 1993.
25 Insight II, Version 2.3.0 and Delphi, Version 2.5, Biosym Technologies, San Diego, 1993.
26 S. Lifson, A. T. Hagler and P. Dauber, *J. Am. Chem. Soc.*, 1979, **101**, 5111.
27 W. F. van Gunsteren and H. J. C. Berendsen, Groningen Molecular Simulation (GROMOS) System, University of Groningen, The Netherlands, 1987 and Version 43A1 parameters from W. F. van Gunsteren, GROMOS '95 Manual, *Intensive Course On Molecular Modelling and Drug Design*, C-MMACS, Banglore, India, 1995.
28 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657.
29 D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1994, **98**, 1978.
30 C. Chipot, B. Maigret, J. I. Rivail and H. A. Scheraga, *J. Phys. Chem.*, 1992, **96**, 10276.
31 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087.
32 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
33 H. L. Friedman, *A Course in Statistical Mechanics*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
34 Z. I. Hodes, G. Nemethy and H. A. Scheraga, *Biopolymers*, 1979, **18**, 1565.
35 A. J. Hopfinger, *Conformational Properties of Macromolecules*, Academic Press, New York, 1973, p. 70.
36 B. Jayaram, G. Ravishankar and D. L. Beveridge, *J. Phys. Chem.*, 1988, **92**, 1032.
37 C. Chothia, *Nature* (*London*), 1975, **254**, 304.
38 D. Eisenberg and A. D. Mclachlan, *Nature* (*London*), 1986, **319**, 199.
39 B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379.
40 A. Shrake and J. A. Rupley, *J. Mol. Biol.*, 1973, **79**, 351.
41 M. L. Connolly, *J. Mol. Graphics*, 1986, **4**, 3.
42 P. J. Rossky and M. Karplus, *J. Am. Chem. Soc.*, 1979, **101**, 1913.
43 M. Mezei, P. K. Mehrotra and D. L. Beveridge, *J. Am. Chem. Soc.*, 1985, **107**, 2239.
44 A. Anderson and J. Hermans, *Proteins: Struct., Funct., Genet.*, 1988, **3**, 262.
45 D. J. Tobias and C. L. Brooks III, *J. Phys. Chem.*, 1992, **96**, 3864.
46 T. J. Marrone, M. K. Gilson and J. A. McCammon, *J. Phys. Chem.*, 1996, **100**, 1439.
47 K. Osapay, W. S. Young, D. Bashford, C. L. Brooks III and D. A. Case, *J. Phys. Chem.*, 1996, **100**, 2698.
48 P. Bash, U. C. Singh, R. Langridge and P. Kollman, *Science*, 1987, **236**, 564.
49 A. Ben-Naim, K-L. Ting and R. L. Jernigan, *Biopolymers*, 1989, **28**, 1327.
50 A. Ben-Naim, in *Protein–Solvent Interactions*, ed. R. B. Gregory, Marcel Dekker, New York, 1995, ch. 9.
51 M. Vasquez, G. Nemethy and H. A. Scheraga, *Chem. Rev.*, 1994, **94**, 2183.
52 C. L. Brooks, M. Karplus and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, (Adv. Chem. Phys., vol. 71), Wiley, New York, 1988, ch. 9.
53 P. E. Smith and B. M. Pettitt, *J. Phys. Chem.*, 1994, **98**, 9700.
54 M. J. Potter, M. K. Gilson and J. A. McCammon, *J. Am. Chem. Soc.*, 1994, **116**, 10298.
55 E. Rajasekaran, B. Jayaram and B. Honig, *J. Am. Chem. Soc.*, 1994, **116**, 8238.
56 A-S. Yang and B. Honig, *J. Mol. Biol.*, 1993, **231**, 459.
57 C. Tanford and R. Roxby, *Biochemistry*, 1972, **11**, 2192.
58 A. Krashikoff, *Protein Eng.*, 1995, **8**, 243.
59 M. K. Gilson and B. Honig, *Nature* (*London*), 1987, **330**, 84.
60 J. Antosiewicz, J. A. McCammon and M. K. Gilson, *J. Mol. Biol.*, 1994, **238**, 415.
61 A. B. Schmidt and R. M. Fine, *Mol. Simul.*, 1994, **13**, 347.
62 T. Simonson and A. T. Brunger, *J. Phys. Chem.*, 1994, **98**, 4683.
63 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127.
64 C. Lim, D. Bashford and M. Karplus, *J. Phys. Chem.*, 1991, **95**, 5610.
65 T. Ooi, M. Oobatake, G. Nemethy and H. A. Scheraga, *Proc. Natl. Acad. Sci., USA*, 1987, **84**, 3086.
66 J. E. Sternberg, R. F. Haynes, A. J. Russel, G. Thomas and A. Fersht, *Nature* (*London*), 1987, **330**, 86.
67 L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486.
68 B. Jayaram and D. L. Beveridge, *Biopolymers*, 1988, **27**, 617.
69 J. G. Kirkwood, *J. Chem. Phys.*, 1934, **2**, 351.
70 J. G. Kirkwood, in *John Gamble Kirkwood Collected Works: Proteins*, ed. G. Scatchard, Golden and Breach, New York, 1967, p. 53.
71 C. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, 1961, ch. 7.
72 F. H. Westheimer and M. W. Schookoff, *J. Am. Chem. Soc.*, 1939, **61**, 555.
73 F. H. Westheimer and J. G. Kirkwood, *J. Chem. Phys.*, 1938, **6**, 513.
74 P. Claverie, J. P. Daudey, J. Langlet, B. Pullman, D. Piazzola and M. J. Huron, *J. Phys. Chem.*, 1978, **82**, 405.
75 C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, W. H. Freeman and Co., San Francisco, 1980, part 1, p. 49.
76 A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, 1994, **81**, 2016.
77 P. K. Ponnuswamy and M. M. Gromiha, *J. Theor. Biol.*, 1994, **166**, 6374.
78 J. March, *Advanced Organic Chemistry, Reactions, Mechanisms and Structure*, Wiley Eastern Limited, New Delhi, 1992, p. 230.
79 H. K. Hall Jr., *J. Am. Chem. Soc.*, 1957, **79**, 5441.
80 F. Shahidi and P. G. Farrell, *J. Chem. Soc., Faraday Trans. 1*, 1978, **74**, 858.
81 J. F. Reading and G. R. Hedwig, *J. Chem. Soc., Faraday Trans.*, 1990, **86**, 3117.
82 R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849.
83 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664.

Paper 6/03913H; *Received* 4*th June*, 1996