An AM1 and PM3 Molecular Orbital and Self-consistent Reaction-field Study of the Aqueous Solvation of Glycine, Alanine and Proline in their Neutral and Zwitterionic Forms

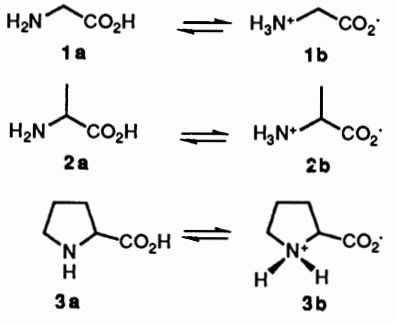
Henry S. Rzepa *,† **and ManYin Yi**
Department of Chemistry, Imperial College of Science Technology and Medicine, London SW7 2AY, UK

Two solvation models based on the semi-empirical AM1 and PM3 SCF–MO Hamiltonians are evaluated for the neutral and zwitterionic forms of the amino acids glycine, alanine and proline. The predicted PM3 solvation energetics for the solvated glycine zwitterion based on a supermolecule comprising up to 15 water molecules are in reasonable agreement with experiment. The calculated PM3 solvent structures agree broadly with previously proposed models, but several new aspects are revealed. The corresponding AM1 model shows the expected preference for bifurcation in the solvent structure and several significant differences to PM3 in the solvation energetics. Energies and geometries obtained using a self-consistent reaction-field (SCRF) model assuming a spherical (isotropic) solvent cavity are compared with the supermolecule approach. Whilst SCRF solvation energies are particularly sensitive to the size of the cavity radius used, geometrical changes involving hydrogen bonding compare well with the supermolecule model. A combined supermolecule/SCRF approach supports the hypothesis that the fifteen-water supermolecules are close to the bulk solvation limit. The failure of both the AM1 and the PM3 solvation models to correctly reproduce the low barrier observed for aqueous intramolecular proton transfer in glycine is discussed.

The potential energy surfaces for a wide range of reactions are thought to be qualitatively modified by the presence of solvent molecules to the extent that, in certain cases, reaction barriers are thought to arise entirely from solvation and hydrogen bonding effects.[1,2] Solvation is particularly important for those reactions which involve the generation of charged intermediates from neutral reactants, *e.g.* many forms of hydrolysis reaction or nucleophilic substitutions. Two approaches to modelling such reactions have been particularly widely applied; the 'supermolecule' approach in which explicit solvent molecules are included,[3] and 'reaction-field' models in which the perturbation of the energy Hamiltonian of the solute by charge, dipole and higher moment terms is evaluated.[4] Each individual approach has both advantages and disadvantages. By explicitly considering solvent molecules, effects such as solute–solvent steric and stereoelectronic interactions can be included, but the solvation sphere is necessarily limited in size by computational requirements. The reaction-field approach is not so limited, but requires a definition of the size and shape of the reaction cavity and a decision regarding the size of the multipolar expansion.

The semi-empirical AM1 and PM3 SCF–MO methods have previously been employed with respect to the hydrogen-bonding features of small water clusters,[5] the supermolecule solvation of small neutral[5] and charged molecules[6] and potential energy surfaces for nucleophilic addition to phosphate groups in the presence of solvent water.[7] AM1 was shown to favour bifurcation in hydrogen bonding to oxygen centres, but gave realistic estimates of water association energies. The structural and vibrational characteristics of hydrogen-bonding interactions predicted by PM3 were generally found to agree with *ab initio* calculations[5] and this semi-empirical method in particular proved capable of qualitatively reproducing *e.g.* polarisation effects, where the strength of a hydrogen bond or electrostatic interaction is modified by the presence of an adjacent water molecule. The polarisation of the solute wavefunction is an inherent feature of the alternative reaction-field model, recently implemented at the AM1 and PM3 level by Karelson and co-workers.[4b] Here the condensed phase environment is considered as a spherical or ellipsoidal cavity of specified relative permittivity and size, and this approach has been shown to be particularly successful in accounting for the relative stability of closely-related systems of moderate polarity in aqueous or polar solution. In this particular implementation however, the perturbation expansion is limited to uni- and di-polar terms,[4b] whereas other approaches have included higher order terms.[4c]

Amino acids are one system for which the well characterised aqueous energetics should serve to calibrate the AM1 and PM3 semi-empirical solvation models for dipolar and ionic systems. In this study, we report AM1 and PM3 results for solvation models of glycine (**1**), alanine (**2**) and proline (**3**) in both their neutral and zwitterionic forms. The proton transfer reaction in glycine also provides a valuable benchmark for considering the effect of various solvation models on a complete potential energy surface.



Computational Procedure.—In establishing a supermolecule model for solvation, preliminary studies revealed that, for the inner solvation sphere at least, the OPLS/A[8] molecular mechanics force field predicts similar structures to PM3 for simple hydrogen-bonding interactions. Initial minimum energy water configurations around the amino acids were therefore

† *E-mail:* RZEPA@UK.AC.IC.CH.VAXA

generated from Monte-Carlo searches using this force field,[8] as implemented in the MacroModel/BatchMin (Versions 2.5 and 3.0) program system[9] and using a constant permittivity (2.5). These geometries were initially expressed in cartesian rather than internal coordinates to avoid problems with co-linearity of adjacent atoms, and served as the starting point for re-optimisation using the MOPAC program (Version 5.0).[10] After initial refinement, problems with atom co-linearity were largely eliminated and geometries could be expressed in terms of bond lengths and angles. Using the PM3[11] Hamiltonian and the PRECISE keyword together with the default BFGS[10] implementation in MOPAC, this approach was found to result in significantly faster final optimisation to final gradient norms $[\Sigma(\partial E/\partial r)^2]^{0.5}$ of 1–2 kcal Å^{-1} for most systems. Subsequently, we found that optimisation based on an eigenvector following implementation (EF)[12] of the BFGS update routinely achieved gradient norms of <0.03, in less than half the processor time taken by the standard MOPAC procedure. This resulted in small but significant improvements in energies ($\leqslant 0.5$ kcal) and changes in bond lengths ($\leqslant 0.005$ Å). Inspection of the optimised structures and hydrogen bonding interactions using real-time rotation and stereo-graphics was found to be essential in revealing subtle structural features. We performed this using an Evans and Sutherland PS390 display system running the MacroModel programs or a Tektronix CAChe workstation (V2.2). Where the number of hydrogen bonds to a specific water molecule appeared low, the geometry was adjusted and re-optimised. Whilst such procedures do not guarantee that the global energy minimum will be located in each case, the final PM3 energies are not expected to be significantly higher than this minimum. There is more uncertainty regarding the AM1 results, since here the final (largely bifurcated) structures are quite dissimilar from the initial guesses provided by OPLS/A or PM3. In each case, a number of different starting geometries were taken, and the lowest energy configuration located is reported here. The structures of pure water clusters were similarly obtained. Solvation energies were obtained by summing the energy of the gas phase solute and the optimised water cluster, and subtracting this quantity from the solute–solvent supermolecule energy. Transition states for intramolecular proton transfers were located initially by minimising the sum of the squared scalar gradients. Use of the eigenvector following (EF) procedure was more recently found to be a more accurate and robust method for the location of transition states.[12] Stationary points so located had the required one negative eigenvalue in the Hessian matrix.

The self-consistent reaction-field model (SCRF) was implemented in MOPAC as reported recently.[4b] We have found, however, that the suggested[4b] geometry optimisation in AMPAC or MOPAC results in unacceptably large final gradient norms and errors in the calculated energy.[12d] We have attributed this[12d] to the approximation used in these programs for calculating derivatives, where the density matrix P is assumed to be invariant to small geometrical changes. If imposed, this approximation can result in significant errors in both geometry and energy. All the calculations reported here were carried out with full SCF iteration for all gradients, using the AMPAC program (V2.1) in which the code for full SCF derivative evaluation is particularly efficient (Keyword DERINU). In all cases, final gradient norms of <0.6 kcal Å^{-1} were achieved. The reaction cavity was approximated as spherical, with a radius calculated from the total molecular volume of the SCF-optimised solute, estimated using the MacroModel program. Molecular electrostatic isopotential maps were generated on the Tektronix CAChe stereographics workstation system using PM3 wavefunctions. Hardcopy of molecular representations was generated using either the Molecule[13] program, or from the CAChe system.

Results and Discussion

i. *Gas Phase Energetics.*—The gas phase energies of **1–3** (Table 1) show reasonable agreement with experimental values where those are available. The energy difference ΔH(**1b** − **1a**) is particularly noteworthy, since a number of *ab initio* calculations of this quantity at various theoretical levels have been reported.[14] The PM3 value of 31.2 kcal mol^{-1} is in better agreement with these *ab initio* results than the AM1 value of 42.3 kcal mol^{-1}. The geometry of **1b** is noteworthy in revealing a strong hydrogen bond between one of the N–H protons and one oxygen (1.719 Å), whereas the corresponding distance with AM1 is significantly longer (1.961 Å, Fig. 1). Such variation between the two methods has been previously noted.[11] For the other amino acids, PM3 predicts ΔH(**3b** − **3a**) to be slightly larger than for glycine, whereas the opposite is found with AM1. This could in fact be due to an error in the PM3 nitrogen parametrisation which we have noted previously,[6] one manifestation of which is that successive alkylation of amines results in a small but systematic error in the calculated enthalpies of formation.

ii. *Solvation Energetics for* **1**.—Glycine has three acidic and two basic centres to which strong hydrogen bonds might form. Gas phase results for NH_4^+ indicate incremental experimental solvation enthalpies of 17.3, 14.1 and 13.4 kcal mol^{-1} for the first three water molecules,[6] whilst the values for $CH_3CO_2^-$ would be expected to be similar for the first two water molecules. Aqueous solvation of the two infinitely separated ions should therefore recover significantly more than 70 kcal mol^{-1} of solvation energy. Estimates[15] of the actual solvation energy, however, are in the range 30–40 kcal mol^{-1}. Three effects might account for the difference. Firstly, the two ionic centres in the zwitterion are not infinitely separated, but are significantly stabilised by their proximity. This is confirmed by the calculated partial charges on the NH_3^+ and CO_2^- fragments (+0.789, −0.59 [PM3], +0.655, −0.629 [AM1]) which are significantly less than unity in each case. Secondly, the strong intramolecular hydrogen bond in *e.g.* **1b**, and hence the structure and conformation, would be disrupted by the formation of intermolecular hydrogen bonds to the solvent. Finally, polarisation effects between the zwitterion and several water molecules would be expected to alter the hydrogen bond strengths of adjacent interactions significantly.[5] Interestingly, the last two effects were neglected in the most recent *ab initio* study of the solvation structure of glycine.[15] This model was obtained by optimising the position of a single water with respect to **1b** and making the approximation that the interaction energies in various configurations were all independent. This procedure resulted in the conclusion that the primary solvation sphere comprises five water molecules with a solvation energy, corrected for basis set superposition errors (BSSE),[16] of *ca.* 40 kcal mol^{-1}. We felt it was particularly important to allow for polarisation and structural re-organisation effects by treating the entire collection of solvent molecules and solute as a single supermolecule. At the *ab initio* level, the congruence of difficulties such as electron correlation corrections, BSSE and geometry optimisation with a large number of variables makes such an approach particularly difficult. Our previous results on the incremental solvation of several small ions suggested that the semi-empirical methods AM1[17] and PM3 are reasonably reliable in this respect, and the latter method in particular reproduces hydrogen bonding structures involving oxygen and polarisation effects at least qualitatively. Since the energies of the hydrogen bonds formed in such ionic systems are significantly greater than kT, a non-dynamic model involving fixed geometrical relationships between the solvent and solute is likely to be valid.[18]

The results for **1–3** are shown in Table 2, presented as the

Table 1 Calculated and experimental heats of formation (kcal mol^{-1}) of **1a–3a** (**1b–3b**)

	Gas phase			SCRF			
	Expt.[11]	PM3	AM1	PM3	AM1		
1	−93.7	−93.8 (−62.5)	−101.5 (−59.2)	−95.5 (−106.7)[b,d]	−97.2 (−161.1)[c]	−101.8 (−113.3)[e]	−101.9 (−157.4)[c]
2	−111.4	−101.1 (−70.5)	−105.0 (−66.2)	−101.5 (−108.9)[b]	−101.8 (−158.9)[c]	−105.4 (−108.3)[e]	−105.8 (−153.6)[c]
3	—	−99.7 (−67.6)	−98.1 (−62.4)	−101.4 (−98.6)[b]	−102.6 (−131.7)[c]	−100.2 (−101.9)[e]	−101.2 (−128.7)[c]

[a] In kcal mol^{-1}. [b] Self-consistent reaction-field model,[4b] using a permittivity of 78.5 and a spherical cavity of radius 3.0, 3.05 and 3.35 Å for **1**, **2** and **3**, respectively. [c] Using a cavity radius of 2.55, 2.60 and 2.90 for **1**, **2** and **3**, respectively. [d] Energy of **1c** = −80.5 kcal mol^{-1}. [e] Self-consistent reaction-field model,[4b] using a permittivity of 78.5 and a spherical cavity of radius 2.9, 2.95 and 3.25 Å for **1**, **2** and **3**, respectively.

Table 2 Calculated heats of formation (kcal mol^{-1}) of the zwitterionic forms **1b–3b** relative to the neutral forms **1a–3a** in the presence of solvating water molecules

	PM3		AM1	
	n = 7	n = 15	n = 7	n = 15
1[a]	4.0 (−505.5)[a]	−0.4 (−983.6)[a]	13.2 (−558.0)[a]	5.0 (−1115.0)[a]
2	9.1 (−512.1)	5.6 (−988.6)	14.8 (−560.2)	1.0 (−1120.0)
3	3.3 (−505.8)	0.4 (−984.1)	2.2 (−559.2)	6.2 (−1109.6)
nH$_2$O[b]	−407.4	−889.5	−464.9	−1012.8

[a] Absolute enthalpy of formation of the neutral amino acid/water cluster (**1a–3a**·nH$_2$O), in kcal mol^{-1}. [b] Energy of the water cluster, in kcal mol^{-1}.

enthalpy difference (**1b**·nH$_2$O − **1a**·nH$_2$O). For glycine at least, this is well-established experimentally[19] as −9.9 kcal mol^{-1} for the condensed phase (n = ∞). Several previous studies have used models constructed with n = 5 or 6 as initial approximations to bulk solvation. We chose to focus on two supermolecule models with either n = 7 or n = 15, representing an adequate primary solvation sphere and an approximate secondary solvation sphere model. Although the inclusion of larger numbers of water molecules (*e.g.* n = 23), is possible within reasonable computational resources, location of the global minimum (or at least of a structure within 1–2 kcal mol^{-1} of this minimum) becomes increasingly difficult using this approach. At the PM3 level, this first model (n = 7) still predicts **1b** to be higher in energy than **1a**, but the relative ordering is now correctly predicted with the larger value of n = 15 (Table 2). The small solvation energy difference between the two sizes of model indicates that the asymptotic limit might be approached within a reasonable number of solvent molecules (<20). The remaining discrepancy of 9.5 kcal mol^{-1} between PM3 and experiment for the energy difference (**1b**·nH$_2$O − **1a**·nH$_2$O) may, at least in part, be due to a specific error in the PM3 nitrogen parametrisation. The incremental solvation enthalpy of NH_4^+ by water is also noticeably underestimated at the PM3 level, and this has been explained in terms of the anomalous U_{ss} and U_{pp} one-centre integrals for nitrogen.[6] The AM1 procedure predicts a positive value for (**1b**·nH$_2$O − **1a**·nH$_2$O), with greater difference between the values for n = 7 and n = 15, indicating that the asymptotic limit has not yet been reached with this number of solvent molecules.

We also evaluated the alternative reaction-field (SCRF) model,[4] where charge and dipole moment perturbations to the solute AM1 or PM3 energies are evaluated for a reaction cavity of specified size and shape. In the present study, we assumed a spherical cavity,[4b,d] although extension to ellipsoidal cavities has been considered.[4b] For **1b**, the calculated molecular volume based on a Van der Waals envelope indicated a spherical cavity radius of 2.55 Å. At this value, both the PM3 and AM1 SCRF models predict the aqueous stability of **1b** relative to **1a** to be well in excess of the experimental value (Table 1). Whilst this method of evaluating the cavity radius gives good estimates of the solvation energy of organic ions containing no nitrogen,[20] the solvation energy of *e.g.* NH_4^+ is also overestimated, possibly owing to the error in the PM3 nitrogen parametrisation noted above. A spherical cavity radius larger by 0.45 Å (3.0 Å) reproduced the experimental relative stability of **1b** and **1a** (Table 1) at the PM3 level, whereas a slightly smaller correction (2.9 Å) was required at the AM1 level. These results indicate that the accuracy of solvation energies obtained using the SCRF model will be limited by the uncertainties in both the atomic parameters and the size and shape of the reaction cavity. For highly anisotropic polar molecules such as **1b**, quadrupole and higher moments may make important contributions to the solvation energy.[4c,21] The PM3 dipolar reaction-field model using a radius value of 3.0 Å also predicts a significant increase in the dipole moment of **1b**, indicating a substantial polarisation of the wavefunction to allow greater charge separation.[4] The calculated charges on the NH_3^+ and CO_2^- fragments for the supermolecule with n = 7 (+0.833, −0.691 [PM3], +0.711, −0.751 [AM1]) and SCRF (+0.933, −0.793 [PM3, A0 = 3.00 A], +0.740, −0.722 [AM1, A0 = 2.90 A]) also show increased charge separation compared to the gas phase result.

iii. *Solvation Structures for* **1**.—The calculated supermolecule structures show interesting features (Fig. 1). All the PM3 low energy configurations located for **1b**·7H$_2$O show the gas phase intramolecular N–H ··· O hydrogen bond to be effectively broken ($r_{H\cdots O}$ 2.728 Å), with an accompanying change in the orientation of the CO_2^- group. This result is in agreement with the PM3 geometry predicted for **1b** using the reaction-field model, which shows $r_{H\cdots O}$ 2.999 Å. The lowest energy PM3 supermolecule structure located reveals that three water bridges are formed between the nitrogen and oxygen centres (Fig. 1), emphasising the importance of cooperativity and polarisation effects, and the treatment of the cluster as a supermolecule. The relationship between each of the three N–H bonds and the nearest carboxyl oxygen are quite different. The closest NH ··· O interaction has only a single water bridge, with an NH ··· O hydrogen bond length (Fig. 1) significantly shorter than that calculated for NH_4^+·3H$_2$O. This again emphasises the polarising effect of the carboxyl anion in strengthening the adjacent NH ··· O hydrogen bond. The next nearest NH ··· O interaction is bridged by two water molecules, whilst the furthest is bridged by three water molecules. The overall effect is a primary water solvation sphere of six molecules. Previous estimates[15] of this number have ranged from 5–7. Each of the NH ··· O or the C–O ··· H bonds is polarised by one or more

adjacent groups and this arrangement therefore maximises the hydrogen-bonding energies. The model constructed for glycine by Kokpol *et al.*[15] predicts that the carboxylate group is strongly bonded by a single water molecule to form a six-membered ring in the first solvation shell of **1b**. However, this result was obtained, as noted above, by optimising the position of a single water with respect to **1b**, and hence did not allow for polarising interactions by other water molecules, or the change in structure of **1b** as a result of such solvation. Our own observations using PM3 are that this ring structure is rather fragile and is easily disrupted by an intervening water molecule.

The PM3 structure with $n = 15$ (Fig. 1) reveals several new features. Three environments for water can be distinguished; primary interactions involving the formation of largely linear hydrogen bonds of length *ca.* 1.7–1.8 Å directly between the solute and water; secondary direct interactions involving the formation of more unusual weak, bifurcated NH · · · O · · · HN bonds of length *ca.* 2.8 Å; and secondary indirect interactions involving linear hydrogen bonds bridging across the primary solvent molecules. There are nine primary-type interactions, involving tetrahedral coordination of both carboxyl oxygens and the three NH groups. Such coordination of the carboxyl oxygen and the nearly linear NH · · · OH bonds is at least qualitatively correct. However, the increase in primary solvation from $n = 6$ to $n = 9$ does not result in much additional stabilisation of the zwitterion. Clearly, the additional solute–solvent hydrogen bonds in the region of the carboxyl group are similar in energy to alternative solvent–solvent interactions, suggesting that the solvation shell of the carboxyl group is less well defined than that of the NH_3^+ group.[15] The stereographics technique enabled a weaker interaction with the solute to be identified, two waters each having bifurcating interactions across two NH bonds (Fig. 1), in addition to forming hydrogen bonds to other water molecules. We have found other examples of such interactions predicted by the PM3 method, as in the solvation of $Cl^-{\cdot}4H_2O$ by a further one or two water molecules.[22] Such interactions are relatively weak, and we are currently conducting investigations to see if these structures can be reproduced at *ab initio* levels of theory.

The bulk water solvation simulations for **1b** with AM1 are significantly different (Fig. 2). The bifurcated hydrogen bonds do not give an optimal lattice packing arrangement, and in both the $n = 7$ and $n = 15$ models at least one of the carbonyl oxygens is found to be penta-coordinated. Furthermore, the balance between the solute–solvent and solvent–solvent interactions appears weighted in favour of the latter, resulting in apparently excessive clustering of the water molecules to one side of the solute.

iv. *Solvation Energetics and Structures for* **2** *and* **3**.—The PM3-predicted hydration structure of **2b**·$7H_2O$ reveals an apparent steric effect (Fig. 3), in which the methyl group disrupts the three-water bridge present in **1b**, resulting in a smaller solvation energy (Table 2). This effect is also reproduced with the reaction-field models (Table 1), which show a smaller energy difference between the zwitterion and the neutral form compared to glycine itself. The AM1 method fails to show this steric effect at the supermolecule level (Table 2), but in this case it is attributable to excessive mutual clustering of the water molecules to one side of **1b**, resulting in no bridging water structure that can be disrupted by the methyl group (Fig. 2).

The solvation of **3b**·$7H_2O$ also shows some significant differences compared with **1b** (Fig. 3). The PM3 solvation energy is rather less than that for **1b**, possibly because there is one NH · · · O hydrogen bond fewer. At the PM3 level for $n = 7$, all the lowest energy configurations for **3b** show that the intramolecular hydrogen bond (1.948 Å) is not completely removed, and both carboxyl oxygens form two hydrogen bonds each, again resulting in a primary solvation shell of six water molecules. With 15 water molecules, the intramolecular bond is now completely broken (2.590 Å). These structural modifications compared to glycine are attributed to the greater rotational rigidity of proline. Other features are also sensibly reproduced by PM3, such as the relatively large hydrophobic methyl group in **2b** and the back ring in **3b** which essentially exclude the approach of water molecules in this region. The PM3 SCRF energetics also indicate less solvation for **3b**, and the calculated geometry shows no intramolecular hydrogen bond (2.47 Å), entirely consistent with the supermolecule result for $n = 15$.

For all three zwitterions, the supermolecule solvation energy for $n = 15$ is similar to that for $n = 7$ (Table 3). Indeed it has even been suggested that solvation energies of ions for small values of n could be larger than those for solvation by bulk water,[2] since the outer-shell water molecules may be randomised by interacting with bulk solvent. As a consequence, the water orientations may not be optimal and their dipoles may be partially neutralised. This however is a dynamic effect which would not be expected to be reproduced in the supermolecule study involving fixed geometrical relationships,[18] but is of course implicit in the reaction-field model. It does appear however that the polarising influence of the solute NH or O^- groups on hydrogen bonds formed by water molecules attenuates rapidly with distance, and that from a mechanistic viewpoint, a supermolecule including seven water molecules may reproduce most significant solvation effects. These conclusions are reinforced by the results of the reaction-field model, which predicts similar solvation enthalpies to the supermolecule results (Table 1). Nevertheless, for reactions where the reaction cavity may change significantly during the course of a reaction, or in comparison of the solvation of different sized species such as **1–3**, the results obtained are clearly sensitive to the size of the cavity used and so may not be particularly accurate. The exclusion of quadrupolar and higher order terms may also contribute significantly to the errors.

v. *Combined Supermolecule/SCRF Approach.*—The supermolecule approach is limited by the number of water molecules included and the problem in locating global minima, whilst the reaction-field model suffers from uncertainties in the size and shape of the reaction cavity. However, a model which combines the supermolecule and SCRF perturbation approaches should in principle more accurately predict the condensed-phase solvation energy, whilst the errors due to cavity shape and radius may be reduced if the cluster is more approximately spherical than the original solute. The SCRF calculations were performed without geometry optimisation, since this results in cleavage of the water–solute hydrogen bonds, and hence in large changes in the cavity radius during geometry optimisation, resulting in instability. The PM3 results of such a calculation on the seven water clusters (Table 4) indicate the additional SCRF bulk solvation energy of this cluster to be *ca.* 6 kcal mol^{-1}, whereas the analogous increment obtained by explicit solvation with a further eight water molecules was in the range 4–5 kcal mol^{-1}. This result suggests that the fifteen-water model was indeed close to convergence, at least at the PM3 level. This combined approach also indicates **1b** to be more stable than **1a**, and emphasises that the remaining difference between experiment and theory is probably due to the defect in the PM3 nitrogen parameters discussed above. It is noteworthy that these results are obtained with an 'unadjusted' cavity radius, where the correction of 0.45 Å noted above has not been applied. This suggests that the solvation energy using this model is less sensitive to this radius and to errors in the parameters associated with the solute molecule. The SCRF/supermolecule

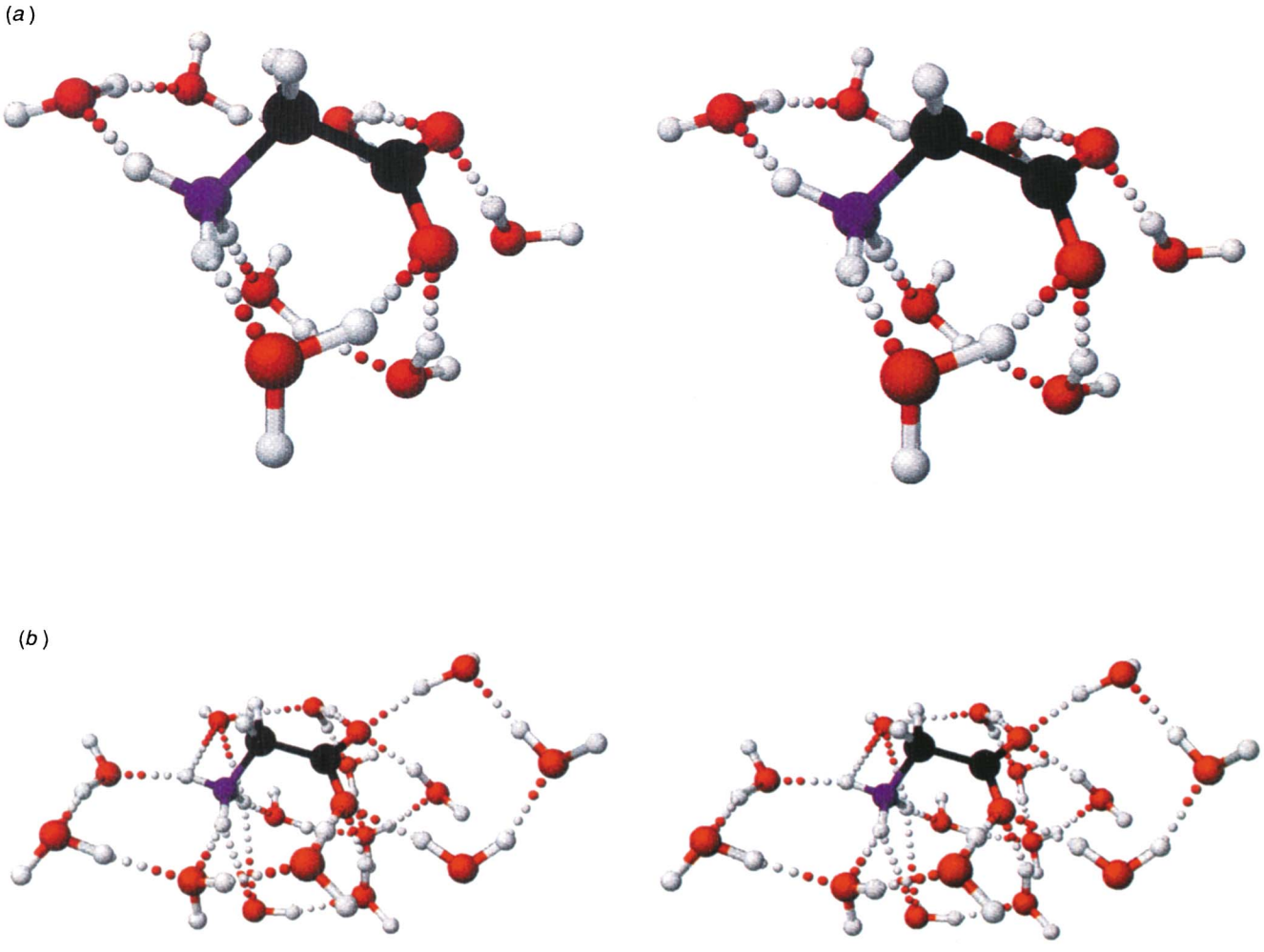


Fig. 1 Stereo-view of the PM3 solvation structure for (*a*) **1b**·7H_2O. Hydrogen bond lengths from the singly-, doubly- and triply-bridged waters to the N–H groups are 1.778, 1.772 and 1.809 Å, respectively, compared to 1.794 Å for NH_4^+·4H_2O. Bond lengths to the carboxyl oxygens are 1.760 and 1.761 Å from the bridging waters and 1.774 Å from the unbridged water, compared with 1.763–1.779 Å for $CH_3CO_2^-$·4H_2O. (*b*) **1b**·15H_2O. Hydrogen bond lengths to the N–H groups are 1.786, 1.793 and 1.812 Å.

(Facing p. 534)

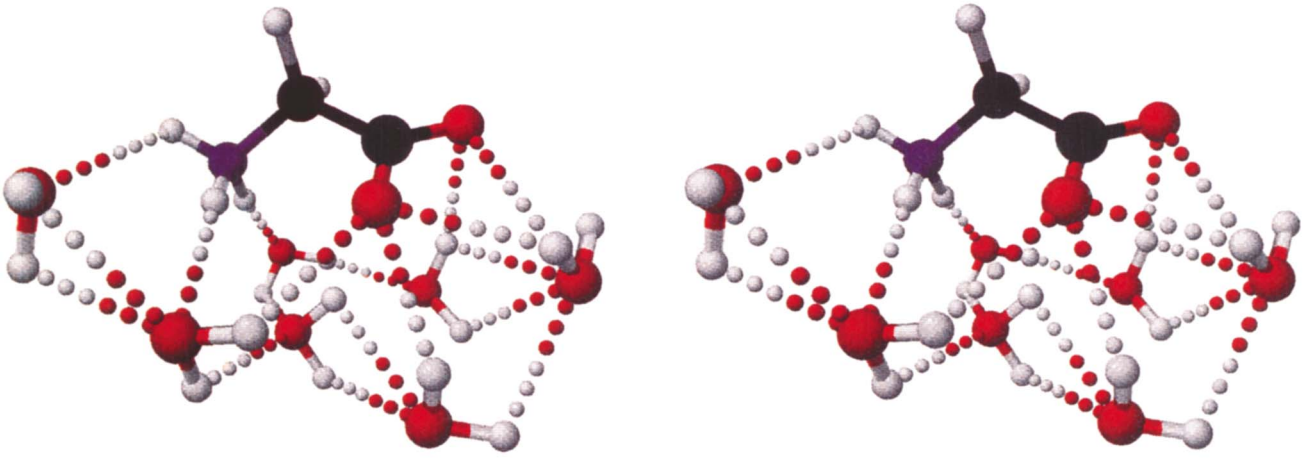


Fig. 2 Stereo-view of the AM1 solvation structure of **1a**, showing the highly bifurcated hydrogen bonding and asymmetric clustering around the solute

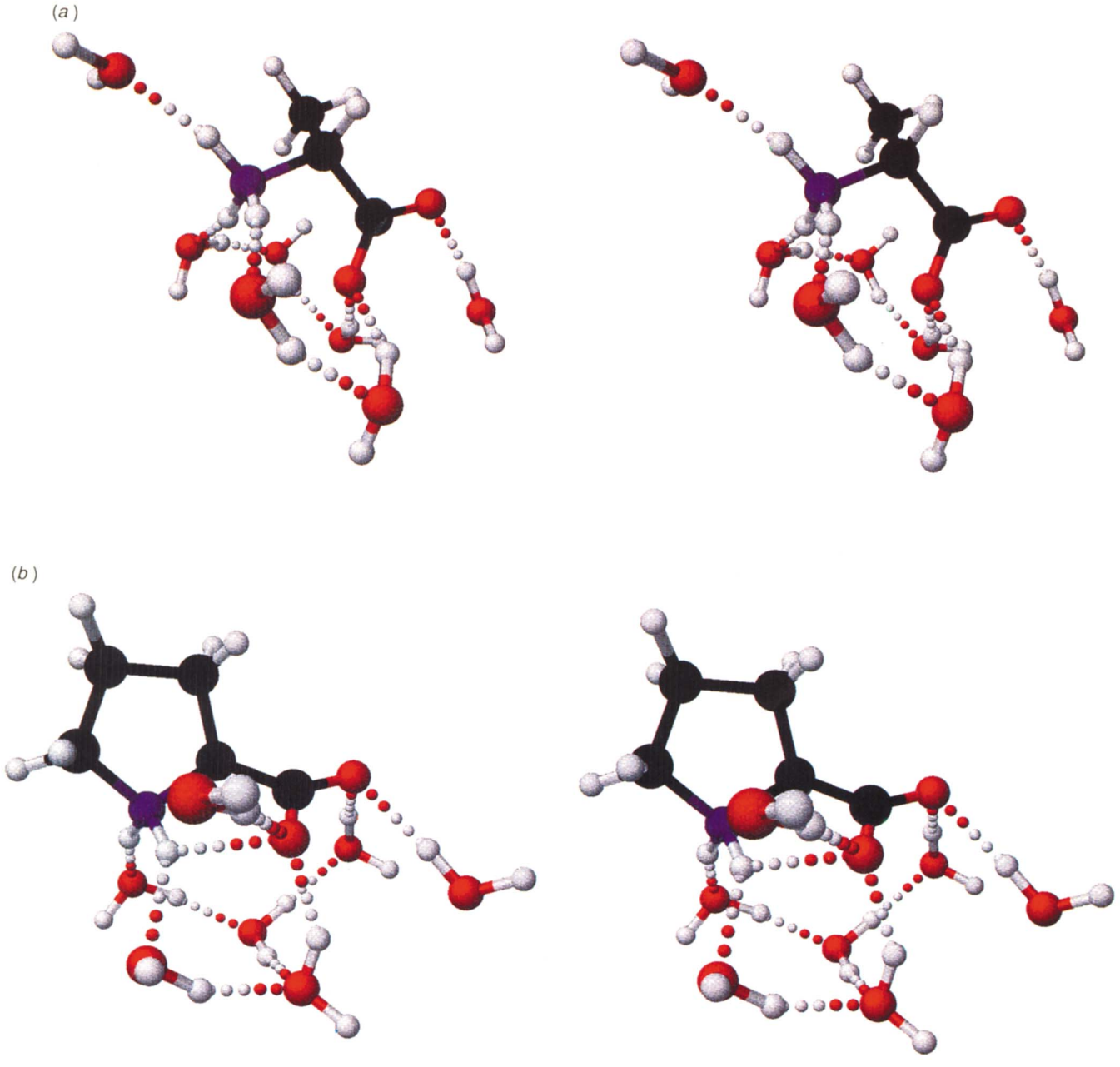


Fig. 3 Stereo-views of the PM3 solvation structure of **2b** and **3b**

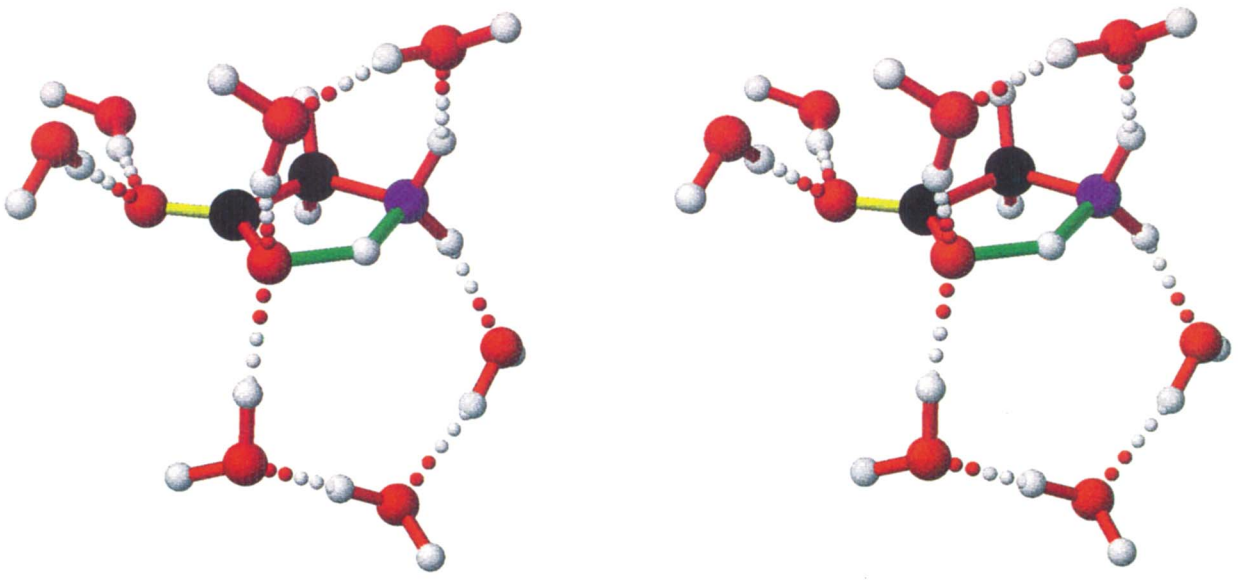


Fig. 5 Stereo-view of the transition state **1c**·$7H_2O$, revealing the absence of hydrogen bonding to the transferring proton

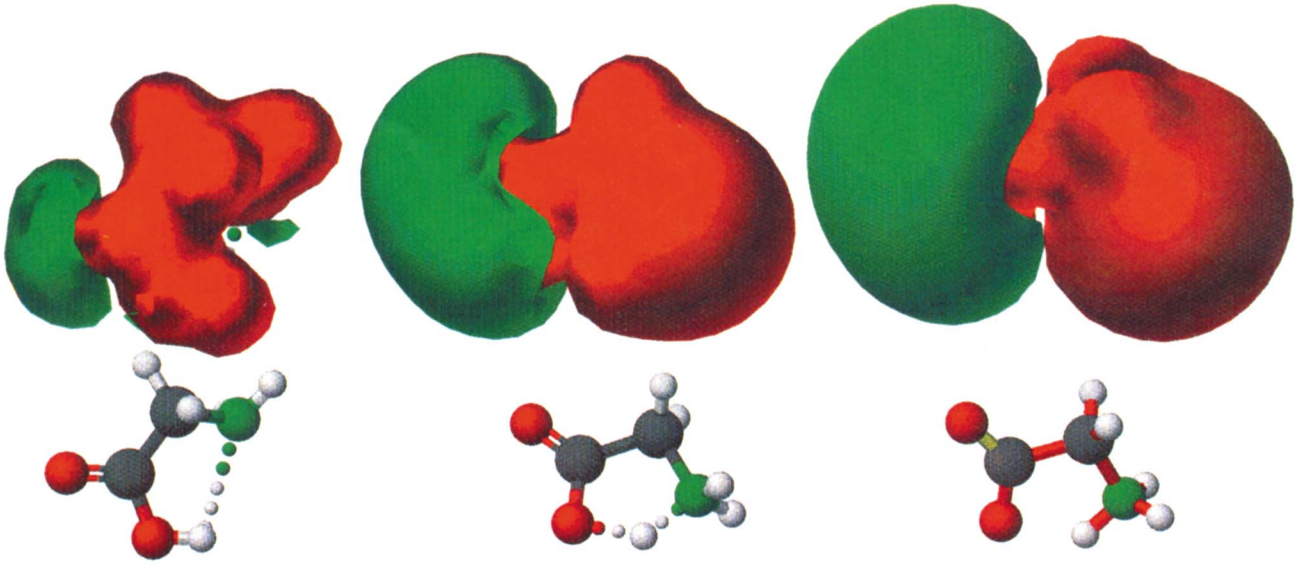


Fig. 6 Molecular electrostatic isopotential maps at 0.05 Hartree for **1a–c**. The green and red areas are, respectively, attractive and repulsive to a proton.

Table 3 Solvation enthalpies of **1b–3b** (**1a–3a**) (kcal mol^{-1})

	PM3		AM1	
	$n = 7$	$n = 15$	$n = 7$	$n = 15$
1	−31.6 (−4.3)	−31.6	−33.9 (−4.8)	−43.0
2	−25.1 (−3.3)	−28.5	−29.1 (−5.1)	−41.0
3	−26.3 (−1.3)	−27.6	−31.9 (−1.7)	−34.0

Table 4 Combined supermolecule/SCRF PM3 energies for **1b–3b**·$7H_2O$ (kcal mol^{-1})

	Zwitterion·$7H_2O$	A0[a]	Neutral·$7H_2O$	A0[a]
1	−507.0	3.46	−505.6	3.49
2	−509.2	3.59	−513.6	3.59
3	−507.6	3.71	−510.2	3.75

[a] Cavity radius/Å, calculated from the total van der Waals molecular volume.

Table 5 Calculated relative energies (kcal mol^{-1}) for the proton transfer reactions for **1**

	AM1	PM3
1b	0	0
1c	4.3	6.4
1c·$1H_2O$	8.2	10.8
1c·$7H_2O$	19.1	24.7
1d	18.8	21.3
1d·$6H_2O$	17.3	26.2
1a	−38.8	−31.3
1a·$1H_2O$	−35.0	−33.3
1a·$7H_2O$	−13.2	−4.4

Table 6 The calculated barriers to proton transfer in **4** and **5** (kcal mol^{-1})

	4			
	X = O	X = N	X = N, O[a]	**5**
PM3	25.6	22.6	28.5	26.9
AM1	21.8	27.0	31.8	—
MNDO	38.8	32.8	37.1	—

[a] From N to O.

energy for **2b** compared to **2a** also follows the same trend as the supermolecule energies alone (Table 4). The anomalous result is for **3a** and **3b**, where the SCRF solvation correction is approximately equal for both the zwitterion and the neutral species. It should be noted that both the water and charge distribution in this species are highly asymmetrical, and the combination of a spherical cavity and the neglect of quadrupolar charge distributions may result in significant errors for this system. In addition, **3b**·$7H_2O$ still has the vestige of an intramolecular hydrogen bond, and hence an SCRF energy at this geometry may also introduce a small error due to the absence of geometry optimisation.

vi. *Proton Transfer Energetics.*—In solution, many reactions which involve the formation of ionic intermediates are also accompanied by rapid intra- or inter-molecular proton transfers. Indeed the kinetics of the equilibrium between **1a** and **1b** have been shown[23] to involve proton transfers with no effective enthalpic barriers, the observed rate constants being almost entirely due to entropic terms. This facile transfer of a proton between two heteroatoms is not confined to solution, since species such as **4** are known to have very small gas-phase

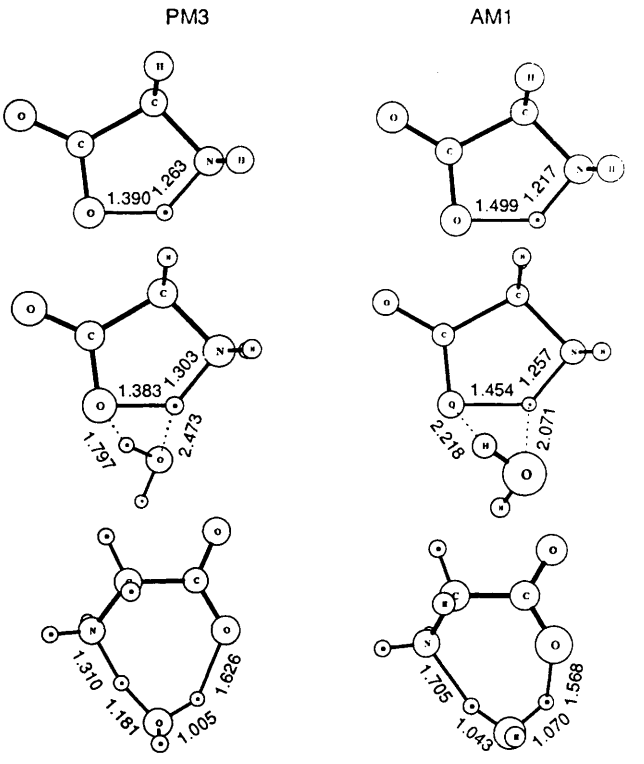
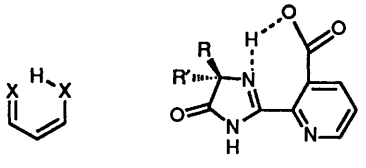


Fig. 4 Calculated transition state structures for **1c**, **1c**·H_2O and **1d**, at the PM3 and AM1 levels. Bond lengths/Å.

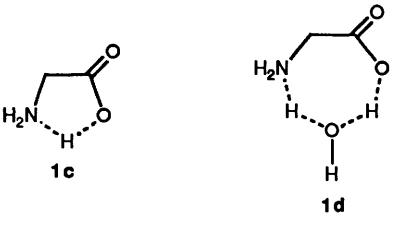
proton-transfer barriers. We have ourselves recently discovered[24] that the proton in **5** is almost equidistant between the nitrogen and oxygen atom in the solid state, suggesting a low barrier for proton transfer between the two heteroatoms.

We chose to study the proton transfer corresponding to



4 **5**

1b ⟶ **1c** ⟶ **1a** to establish the characteristics of the semi-empirical methods for this fundamentally important reaction. In the absence of solvent, a PM3 enthalpic barrier of 6.4 kcal mol^{-1} was calculated for the transition state **1c**, with AM1 giving a similar result (Table 5).



1c **1d**

As expected for a relatively exothermic reaction, the PM3 geometry of **1c** corresponds to an early transition state with a non-linear N–H–O bond angle. This aspect may in itself account for a certain proportion of the barrier, since linear intramolecular proton transfers are the most favourable geometrical arrangement.[25] Structurally, AM1 predicts a less synchronous transition state (Fig. 4). When the enthalpies of reactant and product are more equally balanced, as for proton

transfer in **4** (X = O, N), a different result emerges (Table 6). The barriers are now in the range 22–26 kcal mol^{-1}, all three semi-empirical methods MNDO, AM1 and PM3 giving results which are clearly well in excess of experimental observation. At the X-ray equilibrium geometry[22] for **5**, the enthalpy is predicted to be significantly higher than the ground state (Table 6). Several factors might be responsible for these relatively large errors. One of these may be the limitation of the 1s orbital basis set assigned to hydrogen, and this may also be a case where the neglect of three- and four-centre two-electron repulsion integrals in the formalism cannot be compensated for by any degree of parametrisation. It is also possible that the modification of the core–core repulsion terms in the MNDO, AM1 and PM3 methods, in the latter two by the inclusion of additional attractive and repulsive gaussian terms,[10] leads to anomalous results for proton transfers. The relatively short approach of the two heavy atoms (*ca.* 2.6 Å) may coincide with excessive heavy atom core–core repulsion at this distance.

As the solvation stabilisation of **1b** is increased, so the proton-transfer reaction gradually becomes less exothermic and therefore resembles the case of **4** or **5**. A single passive water in which the solvent is associated with both the O^- and the transferring proton results in a small increase in the PM3 barrier to 10.8 kcal mol^{-1} (Fig. 4). With seven waters, the barrier goes up to 24.7 kcal mol^{-1}, due to greater solvation of **1b** (31.2 kcal mol^{-1}) compared with **1c** (12.9 kcal mol^{-1}). With seven water molecules, the calculated solvent structure shows no significant hydrogen bonds to the transferring proton, with a 'solvation hole' in the region of this atom (Fig. 5). A similar discrimination in solvation energy was obtained using the reaction-field model (18.0 kcal mol^{-1} for **1c** compared to 41.6 for **1b**, using a cavity radius of 2.9 Å for both species, Table 1). To probe the origin of these effects further, we compared the calculated PM3 molecular electrostatic isopotentials (MEPs)[26] at a value of 0.05 Hartree for **1a–c** (Fig. 6). In the gas phase, the positive and negative isopotentials are almost equal in size for **1b**, with the positive region surrounding the ammonium group almost exactly spherical, whilst the negative region associated with the carboxyl group shows more directionality as expected. Proton transfer produces some anisotropy in the positive potential, reducing its magnitude in the region of the transferring proton, but nevertheless the gas-phase electrostatic potentials for **1b** and **1c** remain very similar, and give no indication of the substantially different solvation energies of these two stationary points. Clearly, differing polarisation of the gas-phase wavefunctions by the associated solvent must be in large measure responsible for the differential solvation. This result highlights the difficulty in using properties such as gas-phase electrostatic potentials to predict solvation behaviour, and calls into question the use of *e.g.* point charges derived from such electrostatic potentials in molecular mechanics models of solvent interactions.

In the example of **5**, the surprising symmetry of the N–H–O bond was, at least in part, attributed to the linearity made possible by the planar seven-membered ring formed by this system, and indeed Houk[23] has indicated that this is the optimal ring size for the intramolecular proton transfer. It is therefore possible that the involvement of a bridging water molecule, as in **1d**, may provide a lower energy reaction path for proton transfer *via* a seven-membered ring. However, the PM3 transition state (Fig. 3) has a higher barrier (21.3 kcal mol^{-1}) than **1c** or the isomeric **1c**·H_2O and further passive solvation (*i.e.* **1d**·$6H_2O$) increases the barrier to 26.5 kcal mol^{-1}. The AM1 results differ in several respects. The calculated bond lengths for **1d** show a less synchronous and more product-like structure, and energetically **1d**·$6H_2O$ is now slightly favoured over **1c**·$7H_2O$ with the prospect that further solvation might increase this difference (Table 5). Since, however, a bimolecular reaction is always entropically disfavoured over its unimolecular counterpart,[2] the AM1 and PM3 calculations on balance indicate that the interconversion of the zwitterionic and neutral forms is an intra-molecular process both in the gas phase and in solution, in agreement with experimental evidence.[27] However, the large error in the predicted barriers severely limits the utility of these methods in studying the detail of proton transfer reactions in such systems.

Conclusions

The PM3 supermolecule model for the solvation of amino acid zwitterions appears qualitatively correct in a structural sense, revealing a number of subtle features such as solvent bridges, steric effects due to *e.g.* a methyl group and the disruption of intramolecular hydrogen bonding, which are not reproduced at the AM1 semi-empirical level. These PM3 features are largely reproduced with a modest solvation sphere of *ca.* seven solvent molecules, and there may be no major advantage in going to significantly larger numbers of solvent molecules. Reaction-field models of solvation also reproduce the geometrical changes induced in the solute, but solvation energies are particularly sensitive to the dimensions of the reaction cavity. The combined supermolecule/SCRF model appears to offer a more reliable estimate of the solvation energy. All the parametric semi-empirical Hamiltonians show large errors for proton transfers, and indeed elimination of this type of error may provide valuable indications for the future development and parametrisation of such methods.

Acknowledgements

We thank Shell Research Ltd (Sittingbourne) and the Overseas Research Students (ORS) scheme for financial contributions towards a studentship, and the SERC for grants for computing equipment.

References

1 M. J. S. Dewar and D. M. Storch, *J. Chem. Soc., Perkin Trans. 2*, 1989, 877.
2 A. R. Fersht, *Enzyme Structure and Mechanism*, 2nd edn, W. H. Freeman and Co., New York, 1985.
3 J. D. Madura and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1986, **108**, 2517; P. A. Bash, M. J. Field and M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092; I. H. Williams, *J. Am. Chem. Soc.*, 1987, **109**, 6299.
4 (*a*) For application of the SCRF method using *ab initio* wavefunctions to the glycine system, see R. Bonaccorsi, P. Palla and J. Tomasi, *J. Am. Chem. Soc.*, 1984, **106**, 1945; (*b*) for applications at the semi-empirical level, see M. M. Karelson, T. Tamm, A. R. Katritzky, S. J. Cato and M. C. Zerner, *Tetrahedron Comput. Methodol.*, 1989, **2**, 295; (*c*) J. Bertran, A. Oliva, D. Rinaldi and J. L. Rivail, *Nouv. J. Chim.*, 1979, **4**, 209; (*d*) M. M. Karelson, A. R. Katritzky, M. Szafran and M. C. Zerner, *J. Chem. Soc., Perkin Trans. 2*, 1990, 195; (*e*) for applications in the molecular mechanics area, see W. Hasel, T. F. Hendrickson and W. C. Still, *Tetrahedron Comput. Methodol.*, 1988, **1**, 103.
5 H. S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1990, 943.
6 I. Juranic, H. S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1990, 877.
7 H. S. Rzepa and M. Yi, *J. Chem. Soc., Chem. Commun.*, 1989, 1502.
8 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657.
9 F. Mohamadi, N. J. G. Richards, W. C. Guida, R. Liskamp, M. C. Lipton, M. Caufield, G. Chang, T. Hendrickson and W. C. Still, *J. Comput. Chem.*, 1990, **11**, 440.
10 J. J. P. Stewart, MOPAC, Program 455, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, USA.
11 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209, 221; J. J. P. Stewart, *J. Comp. Aided. Mol. Design*, 1990, **4**, 1.
12 (*a*) J. Baker, *J. Comput. Chem.*, 1986, **7**, 385; (*b*) M. Lehd and F. Jensen, *J. Org. Chem.*, 1990, **55**, 1034; (*c*) J. Baker, F. Jensen, H. S. Rzepa and A. Stebbings, Quantum Chemistry Program Exchange, 1990, **10**, 76; (*d*) H. S. Rzepa and M. Yi, to be submitted.

13 D. K. Agrafiotis, H. S. Rzepa and A. Streitwieser, Molecule, Program 583, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, USA, with local modifications by M. Yi.
14 Y.-C. Tse, M. D. Newton, S. Vishveshwara and J. A. Pople, *J. Am. Chem. Soc.*, 1978, **100**, 4329.
15 S. U. Kokpol, P. B. Doungdee, S. V. Hannongbua and B. M. Rode, *J. Chem. Soc., Faraday Trans. 2*, 1988, **84**, 1789. For a recent study at the MNDO semi-empirical level see O. Kikuchi, T. Natsui and T. Kozaki, *J. Mol. Struct. (Theochem.)*, 1990, **66**, 103.
16 For a recent discussion of basis set superposition errors on the intermolecular interaction potential in water dimer, see I. Mayer and P. R. Surjan, *Int. J. Quantum Chem.*, 1989, **36**, 225.
17 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902.
18 For applications of the molecular dynamics method to the properties of water, see P. L. Moore Plummer and T. S. Chem., *J. Chem. Phys.*, 1987, **86**, 7149. We are grateful to Professor Plummer for a preprint describing the application of the PM3 method in this area.
19 P. Haberfield, *J. Chem. Educ.*, 1980, **57**, 346.
20 H. S. Rzepa and M. Yi, unpublished results.
21 Personal communication from G. P. Ford.
22 H. S. Rzepa and M. Yi, unpublished results. Many examples of bifurcated hydrogen bonded systems are characterised experimentally; R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1983, **105**, 5761. For experimental studies relating to the structure of solute–solvent complexes, see M. C. R. Symons, *Pure Appl. Chem.*, 1986, **58**, 1121; G. Eaton and M. C. R. Symons, *J. Chem. Soc., Faraday Trans. 1*, 1988, **84**, 3459.
23 M. A. Slifkin and S. M. Ali, *J. Mole. Liq.*, 1984, **28**, 215.
24 P. Camilleri, C. A. Marby, B. Odell, H. S. Rzepa, R. N. Sheppard, J. J. P. Stewart and D. J. Williams, *J. Chem. Soc., Chem. Commun.*, 1989, 1722.
25 E. A. Dorigo and K. N. Houk, *J. Am. Chem. Soc.*, 1987, **109**, 2195.
26 For evaluations of electrostatic potentials calculated using semi-empirical methods, see F. J. Luque, F. Illas and M. Orozco, *J. Comput. Chem.*, 1990, **11**, 416; H. Besler, K. M. Merz and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431.
27 G. Wada, E. Tamura, M. Okina and M. Nakamura, *Bull. Chem. Soc. Jpn.*, 1982, **55**, 3064.

Paper 0/03994B
Received 3rd *September* 1990
Accepted 12th *November* 1990