Accurate Prediction of the Solvation of Nucleotide Base Pairs using an *ab initio* Continuum Model

Phillip E. Young, Ian H. Hillier and Ian R. Gould
Chemistry Department, University of Manchester, Manchester, UK M13 9PL

An *ab initio* continuum model successfully predicts the contribution of solvation to the enthalpy of formation of A–T and C–G base pairs in chloroform.

In view of their importance in the structure of DNA, there have been many studies, both theoretical[1-5] and experimental,[6-9] of the interactions in nucleotide base pairs, both in the gas phase and in the condensed phase. We here describe the use of the self-consistent reaction field (SCRF) method to predict the effect of solvation on the association energy of the guanine–cytosine (G•C), and adenine–thymine (A•T) base pairs. Naturally, it is the condensed phase studies that are important from a biological viewpoint, and here the role of solvation is crucial in determining the association energies. The dramatic effect that solvation has on base pair interactions is illustrated by experimental data on the G•C pair, where ΔH is reduced from the gas phase value[8] of -88 kJ mol^{-1} to a value in chloroform of -45 and -24 kJ mol^{-1}, respectively, from spectroscopic[6,7] and calorimetric[9] data. To interpret such changes, there have been many studies concerned with the prediction of the solvation energies of individual bases and base pairs. These have used mainly simulation methods, particularly free energy perturbation methods within a molecular dynamics (MD)[3] or Monte Carlo (MC)[4,5] framework. However, there is some controversy, even for hydration, as to the accuracy of such procedures.[10] An alternative approach for modelling solvation is to employ a reaction field continuum model, where the solvent is treated as a polarisable continuum, rather than as an assembly of discrete molecules as employed in MD and MC methods. The implementation of a SCRF model, within an *ab initio* molecular orbital framework, has been described by a number of workers,[11-13] and has been successfully used to describe the effect of solvation on a range of molecular properties.[14]

The calculations were carried out using a 6-31G** basis and the program GAUSSIAN 92.[15] From structures of the individual bases, optimised at the SCF level using this basis, structures of the base pairs (in the Watson–Crick motif) were obtained by optimisation of the intermolecular hydrogen bonding distances. [Although in the gas phase the Hoogsteen orientation is predicted to be more stable (by <4 kJ mol^{-1})[16] than the Watson–Crick, we have found that solvation favours the latter.]

These dimer and individual monomer structures were then solvated using the SCRF method of Rivail *et al.*[13] Here, an ellipsoidal solvent cavity was used, with dimensions based upon the van der Waals surface of the solute. The volume of the cavity was 432 and 441 Å^3 for the G•C and A•T dimers, respectively. The extension of this model to include a cavity defined by the molecular surface has been described by Dillet *et al.*,[17] who found that for formamide, the ellipsoidal approximation was adequate. The charge distribution of the solute was represented by a multipole expansion up to $l = 6$. The final term in the expansion contributed up to 4 kJ mol^{-1}, giving confidence that convergence had been achieved. This model allows for polarisation of the solute molecules, but here does not include geometry optimisation of the solute in the reaction field. We have found that this latter effect does not influence the conclusions significantly. The calculations were carried out using a dielectric constant of 4.81 for chloroform. The results of the calculations of the electrostatic contribution to solvation are shown in Table 1. Estimates of the contribution of the cavitation and dispersion energies to the change in solvation energies show that the electrostatic term is dominant. It can be seen that for solvation in chloroform the solvation energy is less for the dimer than for the individual bases, by 59.2 and 28.5 kJ mol^{-1} for G–C and A–T respectively. These quantities are formally free energies within the SCRF model. However, the only temperature dependent term in this model is the dielectric constant. Estimates of ΔS based upon the temperature dependence of the dielectric constant show that it is negligible. We may thus compare our solvation energies with those derived from experimental estimates of interaction enthalpies,[6,7,9] as shown in Table 2. Our calculations correctly predict the greater influence of solvation on the G–C compared with the A–T interactions. For the A–T pair, the calculated destabilisation of the interaction, due to solvation in chloroform, is essentially the same as the experimental value. For the G–C pair there is a good agreement between our calculations and the calorimetric data,[9] but not with the spectroscopic results.[6,7] Thus, our calculations support the contention[9] that the calorimetric data are more reliable.

These comparisons with experiment, and with corresponding calculations based upon simulation methods[4,5] show the value of *ab initio* continuum methods for modelling solvation of these biologically important systems. We are thus confident of the

Table 1 Electrostatic contributions (kJ mol^{-1}) to solvation energies in chloroform[a]

1-Me-cytosine (C)	−47.9	(−431.677 99)
1-Me-thymine (T)	−31.4	(−490.562 44)
9-Me-adenine (A)	−23.6	(−503.576 36)
9-Me-guanine (G)	−59.7	(−578.465 66)
A–T	−26.5	(−994.144 36)
G–C	−48.4	(−1 010.160 72)

[a] Absolute energies (au) are given in parentheses.

Table 2 Enthalpies (kJ mol^{-1}) of interaction in gas phase and in chloroform

			Solvation contribution	
	Gas phase[a]	Chloroform	Expt.	Calc.[b]
A–T	−54.3	−25.9 ± 2.5[c]	~28.0	28.5
G–C	−87.8	−45.0 ± 3.2[c]	~43.0 }	59.2
		−24.1 ± 0.6[d]	~64.0 }	

[a] Ref. 8. [b] From Table 1. [c] Refs. 6, 7. [d] Ref. 9.

value of continuum methods for modelling other solvents, such as water, and other bases, including those which may extend the genetic alphabet.[18]

Acknowledgements

We thank SERC (UK) for support of this research.

References

1 B. Pullman, P. Claverie and J. Caillet, *Proc. Natl. Acad. Sci. USA*, 1966, **55**, 904.
2 P. Hobza and C. Sandorfy, *J. Am. Chem. Soc.*, 1987, **109**, 1302.
3 P. Cieplak and P. A. Kollman, *J. Am. Chem. Soc.*, 1988, **110**, 3734.
4 J. Pranata, S. G. Wierschke and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810.
5 A. Pohorille, S. K. Burt and R. D. MacElroy, *J. Am. Chem. Soc.*, 1984, **106**, 402.
6 Y. Kyogoku, R. C. Lord and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1967, **57**, 250.
7 Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta*, 1969, **179**, 10.
8 I. K. Yanson, A. B. Teplitsky and L. F. Sukhodub, *Biopolymers*, 1979, **18**, 1149.
9 L. D. Williams, B. Chawla and B. R. Shaw, *Biopolymers*, 1987, **26**, 591.
10 A. H. Elcock and W. G. Richards, *J. Am. Chem. Soc.*, 1993, **115**, 7930.
11 O. G. Parchment, I. H. Hillier and D. V. S. Green, *J. Chem. Soc., Perkin Trans. 2*, 1991, 799.
12 M. W. Wong, K. B. Wiberg and M. J. Frisch, *J. Am. Chem. Soc.*, 1992, **114**, 1645.
13 J. L. Rivail and B. Terryn, *J. Chim. Phys.*, 1982, **79**, 2; D. Rinaldi, *Comput. Chem.*, 1982, **6**, 155; D. Rinaldi, J. L. Rivail and N. Rguini, *J. Comput. Chem.*, 1992, **13**, 675.
14 P. E. Young, D. V. S. Green, I. H. Hillier and N. A. Burton, *Mol. Phys.*, 1993, **80**, 503.
15 M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. DeFrees, J. Baker, J. J. P. Stewart and J. A. Pople, GAUSSIAN 92, Revision A, Gaussian Inc., Pittsburgh, 1992.
16 I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, 1994, **116**, 2493.
17 V. Dillet, D. Rinaldi and J.-L. Rivail, *J. Phys. Chem.*, 1994, **98**, 5034.
18 J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature (London)*, 1990, **343**, 33.

Paper 4/03357D
Received 6*th June* 1994
Accepted 30*th June* 1994