ELSEVIER

Biophysical Chemistry 78 (1999) 147–155

Biophysical Chemistry

Application of a universal solvation model to nucleic acid bases: Comparison of semiempirical molecular orbital theory, ab initio Hartree–Fock theory, and density functional theory

Jiabo Li, Christopher J. Cramer*, Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431 USA

Received 13 October 1998; received in revised form 23 November 1998; accepted 23 November 1998

Abstract

The free energies of solvation of six nucleic acid bases (adenine, cytosine, hypoxanthine, guanine, thymine, and uracil) in water and chloroform are calculated using CM2 class IV charges and SM5.42R atomic surface tensions. Using any of three approximations to the electronic wave function (AM1, Hartree–Fock, or DFT), we obtain good agreement with experiment for five cases where the experimental results are known for the partition coefficients between the two solvents. Decomposition of the solvation effects into bulk electrostatic contributions and first-solvation-shell effects shows that the partitioning is dominated by the former, and this illustrates the importance of using accurate partial atomic charges for modeling these molecules in aqueous solution.

Keywords: Partition coefficient; Free energy of solvation; Continuum solvent; Polarizable solute; Partial atomic charge; Dipole moment; Electrostatics

1. Introduction

Implicit solvent models have developed rapidly in recent years, and a large number of self-consistent reaction field (SCRF) models have been proposed for calculating the electrostatic component of free energies of solvation [1–7]. Some attention has also been paid to including explicit dispersion and/or cavitation energies [1–3] in the first solvation shell without introducing explicit solvent molecules. Our own approach has involved combining an SCRF treatment of electrostatic effects with a semiempirical approach for including all first-solvation-shell terms, in an average way, without introducing explicit solvent molecules [5–7]. The implied average is both an equilibrium average of solvent configurations and a chemical average over specific molecular instances of a given functional group. We have used

* Corresponding authors.

0301-4622/99/$ - see front matter
PII: S0301-4622(98)00228-2

this approach for a series of solvation models called Solvation Model 1 (SM1) [8], Solvation Model 2 (SM2) [9], etc., or, in general, SMx.

In the SMx models, the standard-state free energy of solvation for the case where the standard-state concentration is the same (e.g. 1 mol/l) in the gas phase and in the liquid-phase solution is given by

$$\Delta G_S^0 = \Delta G_{ENP} + G_{CDS} \quad (1)$$

where ΔG_{ENP} is called the electrostatic term and G_{CDS} is called the cavity-dispersion-solvent-structure term. The latter term accounts for first-solvation-shell contributions, which are modeled by an expression of the form

$$G_{CDS} = \sum_k \sigma_k A_k \quad (2)$$

where σ_k is the atomic surface tension of atom k, and A_k is the exposed surface area of atom k. In SMx models, the exposed surface area is equated to the solvent-accessible surface area (SASA) [10,11] ($x = 1$–4 [8,9,12,13] or 5.4 [14–17]) or to the exposed van der Waals area ($x = 5.0$ [18,19], 5.2 [20], 5.42 [21–23]); the latter is a special case of SASA in which the effective solvent radius is modeled as vanishingly small in order to increase the sensitivity of A_k to molecular structure. The physical motivation for Eq. (1) is that the SASA is proportional to the area of a surface passing through the first solvation shell and thus, for a continuum model of the solvent, it is proportional to the average number of solvent molecules in the first solvation shell (for small solvents) or to the average number of solvent monomer units in the first solvation shell (for example, the average number of $-CH_2-$ units interacting with or perturbed by the solute in the case of n-hexadecane solvent). Then the atomic surface tension σ_k is a constant of proportionality between this average number of interacting or perturbed solvent molecules (or solvent monomer units) and a resulting free energy change attributable to that atom of the solute. This is intrinsically approximate because entropies and free energies cannot be decomposed into atomic contributions in any rigorous fashion, but it accords well with intuitive chemical concepts. In our most recent models ($x \geq 5$ [14–23]), the atomic surface tensions are written as a linear combination with a small number (~ 1–4) of geometry-dependent coefficients:

$$\sigma_k = \sum_\eta \tilde{\sigma}_\eta(Z_k) T_{k\eta}(\mathbf{R}) \quad (3)$$

where $T_{k\eta}(\boldsymbol{R})$ is a function of solute geometry ($\boldsymbol{R}$), and $\tilde{\sigma}_\eta(Z_k)$ depends on the atomic number of Z_k of atom k. For example, if atom k is a hydrogen atom ($Z_k = 1$), one of the $T_{k\eta}$ functions may change from zero for H far from any nitrogen atom to approximately unity for H at a typical N–H bond distance from a nitrogen atom. Thus the geometry-dependent functions serve to automatically assign molecular mechanics 'types' to the individual atoms (but, unlike molecular mechanics types, they are continuous and differentiable functions of molecular geometry and are well defined for all structures including, for instance, transition states). Furthermore, in our universal solvation models [15,17,19–23], we write

$$\tilde{\sigma}_\eta(Z_k) = \sum_\nu \tilde{\sigma}_{\eta\nu}(Z_k)\nu \quad (4)$$

where $\tilde{\sigma}_{\eta\nu}(Z_k)$ is a surface tension coefficient, and ν is a solvent descriptor (for example, ν may be the index of refraction n, the hydrogen-bond acidity α, the hydrogen-bond basicity β, or the macroscopic surface tension γ).

The surface tension coefficients are determined semiempirically by regressing Eqs. (2)–(4) against the difference between experimental free energies of solvation and a bulk electrostatic estimate obtained by SCRF theory. Thus, inherently, the G_{CDS} term models all deviations of the true free energy of solvation from the bulk electrostatics. To emphasize this we use the acronym CDS to remind us of three leading contributions: cavitation (C), dispersion (D) interactions (both solute–solvent dispersion interactions and the change in solvent–solvent dispersion interactions upon insertion of the solute), and non-electrostatic aspects of solvent structural (S) changes (such as those induced by solute–solvent hydro-

gen bonding and the changes in solvent–solvent hydrogen bonding upon insertion of the solute — the latter being an important contributor to the hydrophobic effect). To the extent that hydrogen bonding is partially modeled by bulk-solvent electrostatics, it is included in the SCRF term, so G_{CDS} has only the non-electrostatic (or, more accurately, non-*bulk*-electrostatic) component. The C, D and S effects are not mutually exclusive, and in fact other contributions to G_{CDS} are also inseparable from the C, D and S effects. These other contributions include *any* deviation from bulk electrostatics, both those due to approximations in the SCRF calculation (such as an inaccurate solute charge distribution due to an incomplete basis set) and those due to inherent limitations of the SCRF model (such as the necessity to define a solute–solvent boundary in terms of solute atomic radii or the fact that the solvent interacts with the average rather than the instantaneous solute charge distribution). Clearly the ability of the G_{CDS} term to correct for any inadequacies in our bulk electrostatic term and for the uncertainties in the atomic radii (or, more generally, in the treatment of the solute–solvent boundary) is most believable if the electrostatic treatment is not too far from the correct electrostatics, and this is one of the reasons we developed the class IV charge models.

The fact that the semiempirical atomic surface tensions attempt to model *all* deviations from the consistent electrostatic term is the greatest strength of SM*x* models. Most implicit-solvent models have very significant uncertainties in their quantitative predictions due to uncertainties in the treatment of the atomic boundary for the electrostatic part of the calculation, whereas the predictions of SM*x* models are relatively insensitive to fairly large variations in the atomic radii, at least for neutral solutes, provided the atomic surface tensions are reoptimized for a given choice of radii.

Although the G_{CDS} terms can compensate for some of the error in the SCRF electrostatic computation, such compensation can never be perfect. Because the compensation is reasonably successful, one finds that relatively inexpensive SCRF models are often adequate, and we have developed inexpensive models like SM5.4/AM1 [14–17] and SM5.2R/MNDO/d [20], which are based on the AM1 [24] and MNDO/d [25] parameterizations of semiempirical molecular orbital theory, and which are very affordable even for systems as large and complicated as metalloproteins. On the other hand, we have recently developed practical SM*x* models based on ab initio Hartree–Fock (HF) theory [22,23], density functional theory (DFT) [21,23], and hybrid HF-DFT [23] in order to provide presumably more reliable predictions when such calculations are affordable.

A critical element in the electrostatic part of the calculation is the accuracy of the partial atomic charges. We have used two complementary routes to make the calculation of accurate partial atomic charges more affordable for large solute molecules. First we developed the MIDI! basis set [26] which is specifically designed to yield accurate geometries and partial atomic charges at the Hartree–Fock level without polarization basis functions on C or H atoms. (This may be compared, for example, to the 6-31G* basis set [27] which has polarization functions on all atoms except H atoms.) Second is the CM2 class IV charge model [28] which provides very accurate partial atomic charges from semiempirical, ab initio HF, DFT, or HF-DFT wave functions by a semiempirical mapping from the density matrix. These partial charges are employed in the SM5.42R solvation model [21–23], which was parameterized using gas-phase HF/MIDI! geometries. In order to compensate as well as possible for any systematic errors in the partial atomic charges, a separate set of surface tension coefficients was optimized for each combination of semiempirical or ab initio Fock operator, Kohn–Sham operator, or hybrid Fock–Kohn–Sham operator with a particular one-electron basis set [21–23]. In the present article we will focus on three such combinations [23], namely SM 5.42/AM1, SM5.42/HF/6-31G*, and SM5.42R/BPW91/MIDI!, where the latter is based on DFT with the BPW91 density functional [29,30] and the MIDI! basis set. We will also compare the performance of these calculations to SM5.42R/PM3 [23] calculations, which are based

on the PM3 parameterization [31] of semiempirical molecular orbital theory.

For comparison we will also give results obtained with the SM5.4/AM1 method [14–17], which is based on the earlier CM1 class IV charge model [32].

All methods are tested by experiment for the chloroform/water partition coefficients of the five most common nucleic acid bases, namely, adenine, cytosine, guanine, thymine and uracil.

2. Calculations

The standard-state free energy of solvation is given by Eq. (1) where we use the same standard state in the gas phase and the liquid solution (1 mol/l), and where the electrostatic component has two terms:

$$\Delta G_{\mathrm{ENP}} = \Delta E_{\mathrm{EN}} + G_{\mathrm{P}} \tag{5}$$

Here ΔE_{EN} is the distortion cost, which is the change in the solute's internal energy [electronic (E) plus nuclear (N)] upon insertion in the solvent, and G_{P} is the electric polarization free energy, which includes solute–solvent electrostatic interactions (whose net effect is favorable, i.e. negative) minus the cost of achieving this by raising the internal free energy of the solvent. The latter is evaluated by linear response theory [7].

The SM5.42R model is designed to predict accurate free energies of solvation from calculations with rigid gas-phase geometries. For the six molecules studied here, we used gas-phase geometries optimized by Johnson et al. at the MP2/6-31G(d,p) [33] level. Such a level should provide reasonably accurate geometries, although we note that tests for a wide variety of solutes have shown that the results are not overly sensitive to reasonable variations in geometry [21,23]. An SCF calculation is carried out in the gas phase, and a separate SCRF calculation is carried out for each solvent. Since the nuclei are not explicitly relaxed in solution, the nuclear component of the solute distortion cost is zero, and ΔG_{ENP} is written as ΔG_{EP} for these methods.

The SM5.4 calculations involve geometry optimization in both the gas phase, by AM1, and in solution, with separate optimization for each solvent. An SCF calculation is carried out at the gas-phase geometry, and then SCRF calculations are carried out for each solvent. We previously presented SM5.4/AM1 and SM5.42R/BPW91/MIDI! results for bases methylated at the 1 and 9 positions for pyrimidines and purines, respectively, and for thymine [21,34,35]. In the present paper, in order to make the comparison with SM5.42R models more clear, we also report SM5.4 calculations on the unmethylated bases using the MP2/6-31G(d,p) gas-phase geometries.

2.1. Software

The SM5.42R/HF and SM5.42/BPW91 calculations were carried out with MN-GSM [36], which augments GAUSSIAN [37]. The SM5.42R/AM1 calculations were carried out with GAMESOL [38], which augments GAMESS [39]. The SM5.4/AM1 calculations were carried out with AMSOL [40].

3. Results

Table 1 gives the free energies of solvation in water (w) and chloroform (c). From these we can calculate the partition coefficient $P_{\mathrm{c/w}}$ for transfer of the solute from water to chloroform by

$$P_{\mathrm{c/w}} = \exp\left\{\left[-\Delta G_{\mathrm{S}}^{0}(\mathrm{c}) + \Delta G_{\mathrm{S}}^{0}(\mathrm{w})\right]/RT\right\} \tag{6}$$

The common logarithm of $P_{\mathrm{c/w}}$ is tabulated in Table 2, where it is compared to experiment [41,42].

Table 3 shows the gas-phase and solution dipole moments for two of the models. For the same two models, Table 4 shows a breakdown of the solvation energies for the aqueous case into the two components of Eq. (1).

4. Discussion

There are no experimental results available for the absolute solvation energies in Table 1, so comparison to the experimental partition coefficients is critical for the evaluation of the model.

Table 1
Solvation free energy in aqueous solution and chloroform

Molecule	ΔG_S^0 (kcal/mol)			
	SM5.42R/HF/6-31G*	SM5.42R/BPW91/MIDI!	SM5.42R/AM1	SM5.4/AM1[a]
Adenine	−16.2[b]	−15.7	−15.8	−19.4 (−19.4)
	−14.8[c]	−14.6	−14.8	−14.5 (−14.3)
Cytosine	−21.0	−19.0	−20.2	−3.5 (−22.5)
	−15.4	−13.7	−14.9	−17.4 (−15.7)
Guanine	−24.3	−22.5	−22.2	−26.5 (−24.3)
	−20.3	−18.9	−18.8	−17.7 (−16.4)
Hypoxanthine	−21.3	−18.8	−20.7	−24.3 (−22.1)
	−18.0	−16.2	−17.8	−15.8 (−14.7)
Thymine	−15.0	−13.8	−14.4	−12.6 (−10.6)
	−11.4	−9.9	−11.2	−9.7 (−8.4)
Uracil	−16.1	−14.6	−15.4	−13.6 (−11.5)
	−11.5	−9.8	−11.3	−9.6 (−8.1)

[a] Values in parentheses: single-point calculations at MP2/6-31G(d,p) gas-phase geometries.
[b] Upper entry: water.
[c] Lower entry: chloroform.

Wolfenden and Shih [43] have noted that water/chloroform partition coefficients are very difficult to measure for cytosine and guanine because the water affinity of these two bases are so high that the amount in the chloroform phase is hard to detect under conditions where the molecules remain monomeric. Leo [44] has similarly noted that chloroform/water partition coefficients in the range $\log_{10} P < -3$ are not always dependable because often they are not measured with the proper solvent ratio, e.g. cloroform/water (1000:1). With these caveats in mind, we tabulate the best available experimental values of the base-10 logarithm of $P_{c/w}$ in Table 2.

Table 2 shows that the present model reduces the mean unsigned error in $\log_{10} P_{c/w}$ from 1.6 log units in SM5.4 to 0.6–0.7 log units in SM5.42R, depending on the level of the SCRF calculation. Since the error is reduced by more than a factor of two even when the semiempirical AM1 method is used for both SM5.4 and SM5.42R, and since the error is not reduced very much when we use MP2 geometries with SM5.4/AM1, we attribute the improvement primarily to the improved qual-

Table 2
$\text{Log}_{10}\ P_{c/w}$

Molecule	SM5.42R/HF/6-31G*	SM5.42R/BPW91/MIDI!	SM5.42R/AM1	SM5.4/AM1[a]	Expt.
Adenine	−1.0	−0.8	−0.7	−3.6 (−3.6)	−2.4 [40]
Cytosine	−2.9	−3.9	−3.9	−6.0 (−5.0)	−3.5 [40]
Guanine	−4.1	−3.6	−2.5	−7.1 (−5.8)	−3.2 [40]
Hypoxanthine	−2.4	−1.9	−2.1	−6.2 (−5.6)	–[b]
Thymine	−2.6	−2.8	−2.3	−2.1 (−1.6)	−2.3 [41]
Uracil	−3.4	−3.5	−3.0	−3.0 (−2.5)	−3.0 [41]
MUD[c]	0.7	0.7	0.6	1.6 (1.3)	–

[a] Numbers in parentheses are single-point calculations at MP2/6-31G(d,p) gas-phase geometries.
[b] Unavailable.
[c] Mean unsigned deviation over five cases for which experimental result is available.

Table 3
SM5.42R/CM2 dipole moments (debye) in the gas phase and in aqueous solutions

Molecule	HF/6-31G*			BPW91/MIDI!		
	Gas phase	Water	Chloroform	Gas phase	Water	Chloroform
Adenine	2.35	3.11	2.90	2.20	2.87	2.70
Cytosine	6.53	8.45	7.97	5.89	7.44	7.30
Hypoxanthine	5.31	7.13	6.67	4.52	5.74	5.43
Guanine	6.42	8.21	7.77	5.82	7.28	6.92
Thymine	4.44	5.98	5.57	3.99	5.15	4.84
Uracil	4.49	6.04	5.64	4.01	5.20	4.89

Table 4
Components of the solvation free energy (kcal/mol) in SM5.42R aqueous calculations

Molecule	HF/6-31G*		BPW91/MIDI!	
	ΔG_{EP}	G_{CDS}	ΔG_{EP}	G_{CDS}
Adenine	−7.9	−8.3	−8.7	−7.0
Cytosine	−14.4	−6.6	−11.5	−7.5
Guanine	−15.6	−8.7	−12.8	−9.7
Hypoxanthine	−15.0	−6.3	−11.8	−7.0
Thymine	−10.9	−4.1	−9.0	−4.9
Uracil	−11.7	−4.5	−9.4	−5.2

ity of the partial atomic charges provided by the CM2 model.

The high-quality of class IV charges for biological molecules was emphasized in a previous paper [35]. This is reconfirmed here by comparison of the gas-phase dipole moments to the large-basis-set B3LYP/cc-pVTZ calculations of Kwiatkowski and Leszczynski [45]. For example, Kwiatkowski and Leszczynski calculated 6.41 D for cytosine and 6.57 D for guanine whereas the present CM2/HF/6-31G* gas-phase calculations yield 6.53 and 6.42 D, respectively. However, allowing the electronic charge cloud to relax in aqueous solution by the SM5.42R/HF/6-31G* model raises these dipole moments to 8.44 and 8.21 D, respectively. Figs. 1 and 2 show that these dipole moment changes result from changes in the partial atomic charges that can be as large as 0.08 in

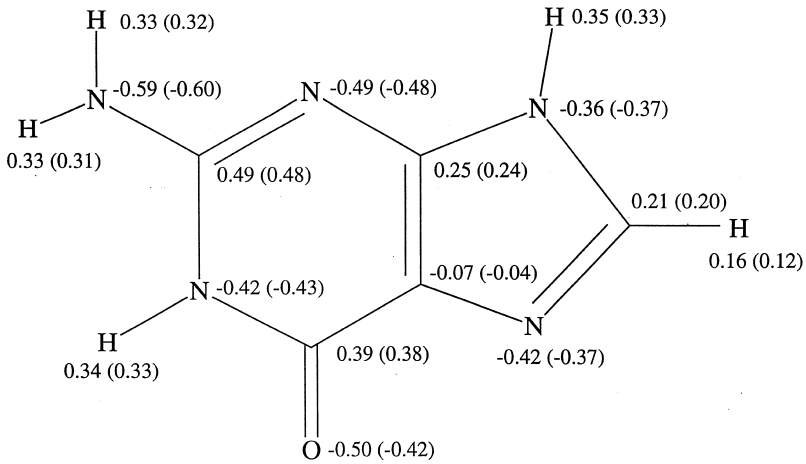

Fig. 1. CM2 charges of guanine in aqueous solution (gas phase) obtained from SM5.42R/HF/6-31G*//MP2/6-31G(d,p) calculations in water and from CM2/HF/6-31G*//MP2/6-31G(d,p) calculations in the gas phase.

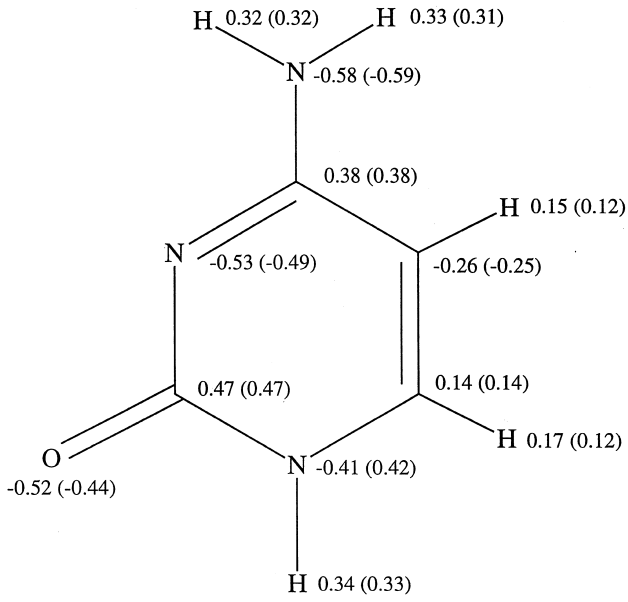

Fig. 2. CM2 charges of cytosine in aqueous solution (gas phase) obtained from SM5.42R/HF/6-31G*//MP2/6-31G(d,p) calculations in water and from CM2/HF/6-31G*//MP2/6-31G(d,p) calculations in the gas phase.

both cytosine and guanine. The largest change occurs at the oxo position in both cases, with the oxygen atom becoming more negative in solution.

Table 3 shows that the HF/6-31G* dipole moments increase by a median amount of 1.25 D in chloroform and 1.67 D in water. The dipole moments and their solvent-induced increases in the dipole moments are both smaller at the BPW91/MIDI! level than at the HF/6-31G* level. Although the CDS terms make up, to some extent, for systematic deficiencies in electrostatics, Table 1 shows that the predicted free energies of solvation are smaller at the BPW91/MIDI! level, as might be expected from the smaller polarities. Since the gas-phase CM2/HF/6-31G* dipole moments show better agreement with experiment [46,47] than do the CM2/BPW91/MIDI! dipole moments, we tentatively accept the SM5.42R/HF/6-31G* results as providing the better representation of the nucleic acid bases.

Table 4 shows that only about two-thirds of the aqueous free energy of solvation comes from the electrostatic contributions. Similarly both contributions are large for chloroform as a solvent. For example, for BPW91/MIDI! calculations on cytosine, ΔG_{EP} and G_{CDS} are, respectively, -9.1 kcal/mol and -4.6 kcal/mol in chloroform.

One aspect of Table 4 is particularly remarkable. In all six cases, whichever method (HF/6-31G* or BPW91/MIDI!) has the smaller value of ΔG_{EP} has the larger value of G_{CDS}. This provides a striking illustration of the fact that G_{CDS} not only models true first-solvation-shell effects, but also, as we have often stressed, it makes up to some extent for systematic deficiencies, if any, in the calculated electrostatic term.

Table 5 shows the values that would be predicted for $\log_{10} P_{c/w}$ if G_{CDS} were neglected. Agreement with experiment is about the same as for the full model for SM5.42R, on the average, and actually slightly better for SM5.4/AM1. It is not clear what to make of this situation. On the one hand the similarity of the mean errors in Table 5 to those in Table 2 might be indicating that there is room for improvement if one would regress the atomic surface tensions against experimental data for a representative set of biological molecules like the present bases or any set of molecules in which the aromatic nitrogen heterocycle functionality is better represented than it

Table 5
$\text{Log}_{10}\ P_{c/w}$ when G_{CDS} is neglected

Molecule	SM5.42R/HF/6-31G*	SM5.42R/BPW91/MIDI	SM5.42R/AM1	SM5.4/AM1[a]	Expt.
Adenine	−1.2	−1.6	−3.1	−2.4 (2.3)	−2.4 [40]
Cytosine	−2.5	−1.8	−3.2	−5.0 (−4.2)	−2.7 [40]
Guanine	−2.7	−2.9	−3.4	−4.7 (−3.9)	−3.2 [40]
Hypoxanthine	−2.6	−2.0	−3.5	−4.9 (−4.0)	–[b]
Thymine	−1.9	−1.5	−2.5	−2.0 (−1.8)	−2.3 [41]
Uracil	−3.2	−1.5	−2.6	−2.2 (−1.8)	−3.0 [41]
MUD[c]	0.7	1.0	0.4	0.8 (0.6)	–

[a] Numbers in parentheses are single-point calculations at MP2/6-31G(d,p) gas-phase geometries.
[b] Unavailable.
[c] Mean unsigned deviation over five cases for which experimental result is available.

was in our training set. On the other hand, the similarity might result from the fact that since the electrostatics predict the differential solvation energies so well for this set of five molecules, there is not much room, in these cases, for improvement by CDS terms. In either event we can conclude from Table 5 that differential solvation of the nucleic acid bases is dominated by bulk electrostatic effects, and the use of realistic partial atomic charges is critical for their accurate modeling. Another conclusion from Table 5, when considered in conjunction with Tables 2 and 4, is that good agreement with log P data does not necessarily imply that the absolute free energies of solvation are reliable. Clearly the ΔG_{EP} values are much smaller than the ΔG_s^0 values, but they lead to predictions of similarity quality for log P. On the basis of comparisons to over 2000 absolute free energies of solvation [21–23], we know that the calculated ΔG_s^0 are much more accurate than ΔG_{EP} in general. Thus the value of testing solvation models against absolute free energies for a large data set (rather than just relative free energies for a small number of cases) is illustrated very dramatically by the fact that these comparisons allow us to draw this conclusion.

Although we have not presented a full set of results for the SM5.42R/PM3 [23] parameterization, we note that it performs quite similarly to SM5.42R/AM1. In particular the mean unsigned deviation from experiment is 0.6 kcal/mol (which may be compared to 0.6 kcal/mol for SM5.42R/AM1 in Table 2), and this decreases to 0.3 kcal/mol when G_{CDS} is neglected (as compared to 0.4 kcal/mol for SM5.42R/AM1 in Table 5). The relatively uniform behavior of the SM5.42R model across four different parameterizations is encouraging.

5. Concluding remarks

The nucleic acid bases provide a very sensitive test set for continuum solvation models because of their complex functionality. This functionality contributes to large local multipole moments and corresponding solvent polarization, and it would also be expected to lead to specific interactions with the first solvation shell that are large in magnitude. Our models, which are optimized to handle consistently both of these components of the solvation free energy, are able to reproduce experimental partition coefficients well. Such performance lends confidence that any subsequent analysis of the partitioning and or solvation free energies (e.g. to determine contributions from specific functional groups [34], to estimate changes from functional group modifications not yet experimentally surveyed, etc.) will be on a solid footing.

Comparison of different models does indicate, however, that observables other than the free energy of solvation may be required in order to further evaluate which models provide the most accurate description of solute electronic structures in solution. For instance, reliable solution dipole moments would be useful in such an evalu-

ation [which amounts in some sense to shedding light on the most accurate way to partition ΔG_{EP} and G_{CDS}], as would work correlating changes in electronic spectra as a function of solvent (solvatochromism). Further development of continuum solvent models will benefit from such comparisons.

Acknowledgements

The authors are grateful to Paul Winget for helpful discussions and assistance and to Albert J. Leo, Richard Wolfenden, and Phoebe Shih for helpful correspondence. This work was supported in part by the National Science Foundation.

References

[1] J.-L Rivail, D. Rinaldi, M.F. Ruiz-Lopez, NATO ASI Ser. C 339 (1991) 79.
[2] J.-L. Rivail, D. Rinaldi, Comp. Chem. Rev. Curr. Trends 1 (1996) 139.
[3] J. Tomasi, M. Persico, Chem. Rev. 94 (1994) 2027.
[4] M. Orozco, C. Alhambra, X. Barril, J.M. Lopéz, M.A. Busquest, F.J. Luque, J. Mol. Biol. 2 (1996) 1.
[5] C.J. Cramer, D.G. Truhlar, Rev. Comp. Chem. 6 (1995) 1.
[6] C.J. Cramer, D.G. Truhlar, Theor. Comp. Chem. 1 (1994) 9.
[7] C.J. Cramer, D.G. Truhlar, in: O. Tapia, J. Bertrán (Eds.), Solvent Effects and Chemical Reactivity, Kluwer, Dordrecht, 1996, p. 1.
[8] C.J. Cramer, D.G. Truhlar, J. Am. Chem. Soc. 113 (1991) 8305.
[9] C.J. Cramer, D.G. Truhlar, Science 256 (1992) 213.
[10] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379.
[11] R.B. Hermann, J. Phys. Chem. 76 (1972) 2754.
[12] C.J. Cramer, D.G. Truhlar, J. Comput. Chem. 13 (1992) 1089.
[13] D.J. Giesen, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. 99 (1995) 7137.
[14] C.C. Chambers, G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. 100 (1996) 16385.
[15] D.J. Giesen, M.Z. Gu, C.J. Cramer, D.G. Truhlar, J. Org. Chem. 61 (1996) 8720.
[16] D.J. Giesen, C.C. Chambers, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. 101 (1997) 2061.
[17] D.J. Giesen, G.D. Hawkins, D.A. Liotard, C.J. Cramer, D.G. Truhlar, Theor. Chem. Acc. 98 (1997) 85.
[18] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. B 101 (1997) 7147.
[19] G.D. Hawkins, D.A. Liotard, C.J. Cramer, D.G. Truhlar, J. Org. Chem. 63 (1998) 4305.
[20] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. B 102 (1998) 3257.
[21] T. Zhu, J. Li, G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Chem. Phys. 109 (1998) 9117.
[22] J. Li, G.D. Hawkins, C.J. Cramer, D.G. Truhlar, Chem. Phys. Lett. 288 (1998) 293.
[23] J. Li, T. Zhu, G.D. Hawkins, D.A. Liotard, C.J. Cramer, D.G. Truhlar, Theor. Chem. Acc., in press.
[24] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[25] W. Thiel, A.A. Voityuk, J. Phys. Chem. 100 (1996) 616.
[26] R.E. Easton, D.J. Giesen, A. Welch, C.J. Cramer, D.G. Truhlar, Theor. Chim. Acta 93 (1996) 281.
[27] M.M. Francl, W.J. Pietro, W.J. Hehre, et al., J. Chem. Phys. 77 (1982) 3654.
[28] J. Li, T. Zhu, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. A 102 (1998) 1820.
[29] A. Becke, Phys. Rev. A 38 (1988) 3098.
[30] J.P. Perdew, K. Burke, Y. Wang, Phys. Rev. B 54 (1996) 16533.
[31] J.J.P. Stewart, Rev. Comp. Chem. 1 (1990) 45.
[32] J.W. Storer, D.J. Giesen, C.J. Cramer, D.G. Truhlar, J. Comput.-Aided Mol. Des. 9 (1995) 87.
[33] R.C. Johnson, T.D. Power, J.S. Holt, B. Immarapora, J.E. Monat, A.A. Sissoko, M.M. Yanik, A.V. Zagorodug, S.M. Cybulski, J. Phys. Chem. 100 (1996) 18875.
[34] D.J. Giesen, C.C. Chambers, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. B 101 (1997) 5084 and erratum to be published.
[35] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Chim. Phys. -Chim. Biol. 94 (1997) 1448.
[36] J. Li, G.D. Hawkins, D.A. Liotard, C.J. Cramer, D.G. Truhlar, MN-GSM-version 98.2.3, University of Minnesota, Minneapolis, MN 55455-0431.
[37] M.J. Frisch, G.W. Trucks, H.B. Schlegel et al., GAUSSIAN 94, Gaussian Inc., Pittsburgh, 1995.
[38] J. Li, T. Zhu, G.D. Hawkins, D.A. Liotard, D. Rinaldi, C.J. Cramer, D.G. Truhlar, GAMESOL-version 2.0, University of Minnesota, Minneapolis, MN 55455-0431. http://comp.chem.umn.edu/gamesol
[39] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, et al., J. Comput. Chem. 14 (1993) 347.
[40] G.D. Hawkins, D.J. Giesen, G.C. Lynch, C.C. Chambers, I. Rossi, J.W. Storer, J. Li, T. Zhu, D.A. Liotard, C.J. Cramer, D.G. Truhlar, AMSOL-version 6.5.3, University of Minnesota, Minneapolis, MN 55455-0431. http://comp.chem.umn.edu/amsol
[41] R. Vasanwala, C. Hansch, C. Strong, C. Hansch, average of measurements, unpublished data, in: A.J. Leo (Ed.), MedChem Software Masterfile, BioByte Corp., Claremont, CA, 1994.
[42] P. Shih, L.G. Pedersen, P.R. Gibbs, R. Wolfenden, J. Mol. Biol. 280 (1998) 421.
[43] R. Wolfenden, P. Shih, personal communication.
[44] A.J. Leo, personal communication.
[45] J.S. Kwiatkowski, J. Leszczyski, J. Phys. Chem. 100 (1996) 941.
[46] H. DeVoe, I. Tinoco, Jr., J. Mol. Biol. 4 (1962) 500.
[47] I. Kulakaowska, M. Geller, B. Lesyng, K. Bolewska, K.L. Wierzchowski, Biochim. Biophys. Acta 407 (1970) 420.