PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 10908–10922

www.rsc.org/pccp

PAPER

The solvation, partitioning, hydrogen bonding, and dimerization of nucleotide bases: a multifaceted challenge for quantum chemistry†

Raphael F. Ribeiro, Aleksandr V. Marenich, Christopher J. Cramer* and Donald G. Truhlar*

Received 5th December 2010, Accepted 21st March 2011
DOI: 10.1039/c0cp02784g

We present M06-2X density functional calculations of the chloroform/water partition coefficients of cytosine, thymine, uracil, adenine, and guanine and calculations of the free energies of association of selected unsubstituted and alkylated nucleotide base pairs in chloroform and water. Both hydrogen bonding and π–π stacking interactions are considered. Solvation effects are treated using the continuum solvent models SM8, SM8AD, and SMD, including geometry optimization in solution. Comparison of theoretical results with available experimental data indicates that all three of these solvation models predict the chloroform–water partition coefficients for the studied nucleobases qualitatively well, with mean unsigned errors in the range of 0.4–1.3 log units. All three models correctly predict the preference for hydrogen bonding over stacking for nucleobase pairs solvated in chloroform, and SM8, SM8AD, and SMD show similar accuracy in predicting the corresponding free energies of association. The agreement between theory and experiment for the association free energies of the dimers in water is more difficult to assess, as the relevant experimental data are indirect. Theory predicts that the stacking interaction of nucleobases in water is more favorable than hydrogen bonding for only two out of three tested hetero-dimers.

1. Introduction

An accurate theoretical description of noncovalent interactions is vital for the understanding of complex biological systems.[1–6] Chemical interactions between the components of biopolymers determine their three-dimensional structures and their functions. Furthermore, solvent effects and ionic interactions modulate the behavior of nucleotides, nucleic acids, proteins, and carbohydrates.[7,8] Nucleotides carry genetic information in DNA and RNA, and they serve as energy transport and cellular signaling molecules in the cytosol. The nucleotide bases (which will be called nucleobases for brevity) are the units that carry genetic information, and their pairing is fundamental for storing and transmitting that information. Therefore, it is essential to understand the interactions of nucleobases with each other in various environments.

Gas-phase experiments demonstrate that, in vacuum, the hydrogen bonding interaction within dimers of the nucleobases is more favorable than the π–π stacking interaction.[9] (Note, however, that for trimers of nucleobases in the gas phase, stacking interactions do become important.)[10] The situation is the same in organic solvents; hydrogen-bonded structures of nucleobases have been observed in chloroform and DMSO solutions by infrared spectroscopy and nuclear magnetic resonance experiments.[11–15] Aqueous-phase experiments have been less informative because proton exchange between the solvent and the bases confounds the interpretation of the NMR spectrum, and the signal noise generated by the solvent makes it difficult to distinguish the molecular structures of the nucleobases in the IR spectrum. Nevertheless, NMR experiments provide indirect evidence for the association of nucleobases in water by π–π stacking,[16,17] and further indirect evidence is provided by solubilities,[18,19] ultracentrifugation, ultraviolet spectroscopy, and the osmotic pressure[16,18,20] of aqueous solutions containing mixtures of nucleosides. The association processes, however, need not be limited to the formation of a heterobase dimer; they may generate a complex speciation of homo- and hetero-dimers and higher polymers, thereby complicating any attempts to compare preferences for hydrogen bonding to those for π–π stacking.

Most of the experimental findings described above have been confirmed by classical molecular dynamics, Monte Carlo simulations[21–32] and free energy calculations of nucleobase pairs in systems containing explicit solvent molecules by using nonpolarizable molecular mechanics force fields, and this kind of treatment has also been used to study the hydration patterns

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455-0431, USA. E-mail: cramer@umn.edu, truhlar@umn.edu

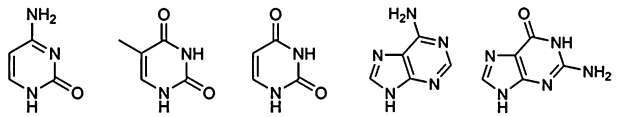
† Electronic supplementary information (ESI) available: Solvation free energies of nucleobase dimers; polarization, electronic relaxation, and cavity–dispersion–solvent-structure components of the computed aqueous solvation free energies; Cartesian coordinates of alkylated and natural nucleobases dimers. See DOI: 10.1039/c0cp02784g

in the first solvation shell of DNA nucleobase dimers.[28–32] Florian *et al.* studied the interactions of nucleobases in water by using a modified Langevin dipoles solvation model[33,34] that included empirically scaled entropies of binding so that the computed free energies of association would agree with equilibrium constants obtained experimentally for substituted nucleosides of DNA and RNA. Danilov and van Mourik studied the hydration of adenine–thymine hydrogen-bonded and stacked dimers in a medium with 200 water molecules with the semiempirical molecular orbital method PM6 and found that within this model stacking was more favorable than hydrogen bonding for the mentioned dimer, due to enhanced water–base interactions in the stacking dimer compared to the hydrogen bonded one.[96] Furthermore, *ab initio* and density functional calculations with explicit water molecules have been performed,[30,32,35–38] but not on systems containing enough solvent molecules to describe bulk-solvent effects. Nevertheless, the consensus in the literature is that aqueous solvent effects specific to the first solvation shell are responsible for the preference of stacked over hydrogen-bonded structures.[30,32,35–39] However, multiscale models (in which the solute or supersolute monomers or dimers are treated by quantum mechanics, and the solvent is treated as a continuum), which are the most practical way to combine high-level treatments of dimerization with first-solvation-shell effects and bulk electrostatic effects corresponding to the experimental dielectric constant, have been employed in only a few studies to compute the stabilization energies and molecular geometries of nucleobase pairs in solution,[40–42] and—even in those studies—they have not been employed with reliable quantum mechanical treatments of the nucleobase dimers. In the present article, we shall employ continuum models that include effects specific to the first solvation shell so that the calculations simultaneously include solute–solvent polarization, the effect of the bulk solvent, and effects of the first solvation shell. The solutes are treated by a density functional.

An accurate theoretical study of the noncovalent interactions of nucleobases in solution could greatly facilitate our understanding of the chemistry of these important compounds. However, until recently the quantum mechanical computation of noncovalent interactions was challenging for large systems because reliable wave function methods were too expensive for most applications, and most density functionals provided inaccurate results for many cases involving noncovalent interactions.[43–46] However, the recent Minnesota density functionals (such as M05-2X and M06-2X)[47–50] allow for more accurate description of the kinds of systems of interest here, as described elsewhere.[49,50]

In the present study, we have investigated both natural and alkylated nucleobases, namely, cytosine (C), thymine (T), uracil (U), adenine (A), guanine (G), 9-hexyladenine, 1-ethyladenine, 1-hexylthymine, and 1-cyclohexyluracil; the tautomeric structures of the natural bases employed are shown in Scheme 1. Analogous tautomeric structures were employed to study the alkylated nucleobases.

We also studied natural and substituted nucleobase dimers, formed by the bases listed above, in the gas phase and in solution. We computed properties of natural nucleobases (*i.e.*, unsubstituted A, C, G, T, and U) along with the A–T, A–U, C–G stacking and



Scheme 1 Tautomeric structures of nucleobases.

Watson–Crick pairs, and the A–A and U–U minimum energy conformers. In addition, we also investigated structural properties of substituted nucleobases and their hydrogen bonding dimers in various conformations, in particular, 9-hexyladenine–9-hexyladenine, 1-hexylthymine–1-hexylthymine, 9-hexyladenine–1-hexylthymine, 9-ethyladenine–1-cyclohexyluracil, and 1-cyclohexyluracil–1-cyclohexyluracil. The natural nucleobases are studied for their fundamental interest, whereas the substituted nucleobases are studied because experimental data are available mainly for nucleobase derivatives, in particular nucleosides and alkylated nucleobases.[51,52] We will distinguish the results provided by alkylated bases from the ones provided by their natural counterparts by always referring to a chemical system by its appropriate name (for example, uracil, or U, will always refer to the natural uracil monomer, whereas 1-cyclohexyluracil, or 1-cHexU, refers to the substituted monomer). It is a potential source of confusion in the literature that some workers refer to substituted nucleobases by the names of their unsubstituted analogs.

We employed the M06-2X[48,49] density functional in all computations. We employed this density functional in conjunction with the Minnesota solvation models to describe the solvent effects on the stabilization energies and molecular geometries of the selected systems in chloroform and water. Namely, we use Solvation Model 8 (SM8),[53] Solvation Model 8 with Asymmetric Descreening (SM8AD),[54] and Solvation Model based on Density (SMD).[55] These solvation models were originally tested over a set of 2892 solvation data including the solvation free energies and transfer free energies of neutral solutes in water and in 90 nonaqueous solvents as well as the free energies of aqueous and nonaqueous mono-charged ions.[54] The mean unsigned error averaged over 26 combinations of various basis sets and density functionals for neutral solutes is 0.7, 0.6, and 0.8 kcal mol^{-1} for SM8, SM8AD, and SMD, respectively, and the mean unsigned errors for ions are 4.4 (SM8), 4.0 (SM8AD), and 4.3 kcal mol^{-1} (SMD).[54] These errors are among the lowest ones provided by current continuum solvent models.[56,57]

2. Computational methods

The first part of the present work involves a study of the partitioning of monomeric cytosine, thymine, uracil, adenine, and guanine between chloroform and water. The corresponding partition coefficient is defined as

$$P_{c/w} = \frac{[\text{solute}]_c}{[\text{solute}]_w} \quad (1)$$

where $[\text{solute}]_c$ is the equilibrium concentration of the solute in chloroform, and $[\text{solute}]_w$ is the equilibrium concentration

of the solute in water. The common logarithm of $P_{c/w}$ is calculated as

$$\log P_{c/w} = \frac{\Delta G_S^o(w) - \Delta G_S^o(c)}{2.303RT} \quad (2)$$

where ΔG_S^o is the standard-state solvation free energy of a given solute in chloroform (c) or in water (w) at $T = 298$ K. The standard-state solvation free energy is defined as the free energy of transfer from the gas phase to the condensed phase under standard state conditions, and it may be calculated by

$$\Delta G_S^o = \Delta G_S^* + \Delta G_{conc}^o \quad (3)$$

where the "*" superscript refers to the fixed-concentration transfer free energy, *i.e.*, that for transfer from an ideal gas at a concentration of 1 mol L^{-1} to an ideal solution at a liquid-phase concentration of 1 mol L^{-1}. The "o" superscript refers to the standard-state solvation free energy defined for the gas-phase solute having a standard-state partial pressure of 1 atm to an ideal solution at a liquid-phase concentration of 1 mol L^{-1}. The last term in eqn (3) accounts for the concentration change, and it equals 1.89 kcal mol^{-1} at 298 K. Since theoretical solvation free energy calculations usually focus on ΔG_S^* (and not on ΔG_S^o) we will refer to ΔG_S^* in the rest of this article as the free energy of solvation.

The second part of the present work involves a calculation of the standard-state free energies of association for selected nucleobase pairs (ΔG_{int}^o) in solution as follows

$$\Delta G_{int}^o = G_{AB}^o - G_A^o - G_B^o \quad (4)$$

where G_X^o denotes the total free energies of the monomers A, B and the dimer A–B in liquid solution at 298 K and is defined as

$$G_X^o = E_{tot,gas} + \delta G_{T,gas}^o + \Delta G_S^* + \Delta G_{conc}^o \quad (5)$$

where X is A, B, or AB, $E_{tot,gas}$ is the total energy of the gas-phase solute molecule from a quantum-mechanical electronic structure calculation (it includes the total electronic and nuclear-repulsion energies), $\delta G_{T,gas}^o$ is a thermal correction to $E_{tot,gas}$ for the gas-phase solute at $T = 298$ K under a standard-state partial pressure of 1 atm calculated based on statistical thermodynamics, and the other quantities are defined above.

The fixed-concentration solvation free energy ΔG_S^* can be further expressed as

$$\Delta G_S^* = \Delta E_E + \Delta E_N + G_P + G_{CDS} + \Delta G_{vib-rot} \quad (6)$$

where ΔE_E is the change in the solute's internal electronic (E) energy in moving from the gas phase to the liquid phase at the gas-phase molecular geometry, ΔE_N is the change in the internal energy (electronic, including nuclear repulsion) of the solute due to changes in its equilibrium nuclear (N) configuration, G_P is the polarization free energy associated with the dissolution process, G_{CDS} is the non-bulk-electrostatic component of the free energy that is nominally associated with cavitation, dispersion, and solvent structure (CDS), and $\Delta G_{vib-rot}$ is the change in the vibrational and rotational free energies of the solute in moving from the gas-phase to the liquid-phase. We do not explicitly compute $\Delta G_{vib-rot}$.

The CDS term is given by

$$G_{CDS} = \sum_{k=1}^{N} A_k(\mathbf{R})\sigma_k(\mathbf{R}) \quad (7)$$

where A_k is the solvent-accessible surface area of atom k, which depends on the molecular geometry $\mathbf{R}$, N is the number of atoms in the molecule, and σ_k is the atomic surface tension of atom k given by

$$\sigma_k = \sum_{m=1}^{M}\left[\tilde{\sigma}_{Z_k m} + \sum_{k'=1}^{N}\sum_{j=1}^{J(Z_k,Z_{k'})} T_{Z_k Z_{k'}}^{(j)}(\mathbf{R})\tilde{\sigma}_{Z_k Z_{k'} m}^{(j)}\right]\omega_m \quad (8)$$

where $\tilde{\sigma}_{Z_k m}$ and $\tilde{\sigma}_{Z_k Z_{k'} m}^{(j)}$ are atomic surface tension coefficients, which are empirical parameters in each solvation model and which depend on the identities of atoms k and k' only through their atomic numbers Z_k and $Z_{k'}$ (some $\tilde{\sigma}_{Z_k m}$ are independent of atomic numbers); $T_{Z_k Z_{k'}}^{(j)}$ is a geometric factor depending on molecular geometry $\mathbf{R}$; ω_m is a solvent descriptor (such as the solvent's hydrogen bond acidity or basicity); J is the number of geometry factors for a given pair of Z_k, and M is the number of ω_m. Since the solvent-accessible surface runs through the middle of the first solvation shell, its area is a continuous function of geometry that is approximately proportional to the number of solvent molecules in the first solvation shell, and therefore A_k is a measure of the number of solvent molecules at a short distance from atom k. Therefore, the CDS terms account for solvent effects specific to the first solvation shell. Note that since the CDS term is parameterized in each solvation model, it also includes (in an empirical way) a correction for the uncertainty in location and finite width of the solute–solvent boundary in the electrostatic part of the calculation and a correction for the neglected change in solute vibrational–rotational energy in passing from the gas-phase to a particular solution.

The solvation free energy to be used for comparison to experiment is calculated as a difference between the solute molecule's free energy in solution at the geometry optimized in solution and the gas phase free energy at the geometry optimized in the gas phase. However, for interpretive purpose we carry out calculations both with and without solution-phase optimization, and we always state which choice is made for a given set of results. Note that the ΔE_N term vanishes in the case when the solute's molecular geometry is fixed at the gas-phase one, and G_P and G_{CDS} change.

The free energies of solvation were calculated using three continuum solvation models: SM8,[53] SM8AD,[54] and SMD.[55] The SM8 and SM8AD models utilize the generalized Born (GB) approximation[58–61] based on self-consistent polarized class IV partial atomic charges[62–64] to calculate the bulk-electrostatic contribution to the free energy of solvation (the first three terms in eqn (6)). The SM8 and SM8AD models treat dielectric descreening effects in terms of the so-called Born radii of individual atoms in the solute molecule. The SM8 model employs Born radii based on the Coulomb field approximation of Still *et al.*,[60] but with empirically improved d parameter. The SM8AD model improves on the earlier SM8 model by using a new asymmetric descreening algorithm with a modified formula for the Born radius suggested by Grycuk.[65]

Unlike the GB-based SM8 and SM8AD models, the SMD model solves the nonhomogeneous-dielectric Poisson equation for bulk electrostatics in terms of the continuous charge density; for this purpose it uses the Integral-Equation-Formalism Polarizable Continuum Model (IEF-PCM).[66–69] Although SMD employs a more complete description of the solute charge distribution than SM8 or SM8AD, it is not necessarily more accurate because application of the Poisson equation to a molecule in solution, especially with a discontinuous change of dielectric constant at a postulated solute–solvent boundary, is as approximate as invocation of the Generalized Born equation.[70]

The protocol described above was applied for the computation of free energies of association for dimers characterized by a single conformer. In this work we apply this single-conformer strategy to study the formation of Watson–Crick base pairs and stacked structures of *unsubstituted* natural nucleobases. However, in order to make a more precise comparison with association free energies of the *alkylated* nucleobases provided by experiment,[51,52] we went beyond the single-conformation approach and included conformational averaging over multiple dimeric structures. We included 3–4 low-energy structures for each alkylated pair in the calculation of Boltzmann weighted average dimer free energies in solution according to

$$\langle G^{\circ}_{\mathrm{AB}} \rangle = -RT \ln \sum_{i \in \mathrm{AB}} e^{-G^{\circ}_i / RT} \qquad (9)$$

where G°_i is the standard-state liquid-phase free energy of conformer i of dimer AB, and T = 298 K. Free energies of association were then obtained by replacement of G°_{AB} in eqn (4) by $\langle G^{\circ}_{\mathrm{AB}} \rangle$.

All calculations were carried out with a locally modified version[71,72] of the *Gaussian 03* electronic structure program suite[73] using the M06-2X[48,49] density functional with the 6-31G(d,p) and 6-31+G** basis sets.[74] Thermal contributions $\delta G^{\circ}_{T,\mathrm{gas}}$ are calculated using the M06-2X/6-31G(d,p) or M06-2X/6-31+G** gas-phase harmonic frequencies at the corresponding minima. The frequencies were scaled by a factor of 0.97 in order to account for systematic errors in the density functional and for anharmonicity,[75,76] and all frequencies below 50 cm^{-1} were raised to 50 cm^{-1} to account for inadequacies in the harmonic oscillator approximation. The SMx/M06-2X/6-31G(d,p) and SMx/M06-2X/6-31+G** solvation free energies $\Delta G^{*}_{\mathrm{S}}$ (where x is 8, 8AD, and D) were calculated using both the gas-phase geometry and the liquid-phase geometry optimized in chloroform and water with SMx/M06-2X/6-31G(d,p) and SMx/M06-2X/6-31+G**. Molecular geometries were optimized using the default convergence-related settings in *Gaussian 03* for all molecular structures considered in the present article, except for a few dimers (for instance, stacked A–T, A–U, and C–G in solution) for which we applied less strict convergence criteria and still obtained reasonably converged structures. All the optimized geometries for the substituted and unsubstituted dimers are given in ESI.†

Solvent descriptors used in the SM8, SM8AD, and SMD calculations are taken from the Minnesota Solvent Descriptor Database.[77]

3. Results and discussion

3.1 Chloroform–water partitioning

Table 1 shows the free energies of solvation for cytosine, thymine, uracil, adenine, and guanine in chloroform and water calculated as calculated in the present work and compares them with the results of some previous[41] calculations. The table shows both results calculated at geometries optimized in the gas-phase (called "unrelaxed" in the table) and results calculated at geometries optimized in the liquid phase (called "relaxed"). Table 2 shows calculated common logarithms of the corresponding chloroform–water partition coefficients, calculated by eqn (1) and (2) in comparison with available experimental data.[78–80]

Table 1 Fixed-concentration (1M → 1M) free energies of solvation (kcal mol^{-1}) for nucleobases in chloroform and water[a]

	Chloroform		Water	
Base	Unrelaxed[b]	Relaxed[c]	Unrelaxed[b]	Relaxed[c]
SM5.4/AM1[d]				
Cytosine	−15.7	−17.4	−22.5	−23.5
Thymine	−8.4	−9.8	−10.6	−12.6
Uracil	−8.1	−9.6	−11.5	−13.6
Adenine	−14.3	−14.5	−19.4	−19.4
Guanine	−16.4	−17.7	−24.3	−26.5
SM5.42R/AM1[d]				
Cytosine	−14.9		−20.2	
Thymine	−11.2		−14.4	
Uracil	−11.3		−15.4	
Adenine	−14.8		−15.8	
Guanine	−18.8		−22.2	
SM5.42R/HF/6-31G*[d]				
Cytosine	−15.4		−21.0	
Thymine	−11.4		−15.0	
Uracil	−11.5		−16.1	
Adenine	−14.8		−16.2	
Guanine	−20.3		−24.3	
SM5.42R/BPW91/MIDI![d]				
Cytosine	−13.7		−19.0	
Thymine	−9.9		−13.8	
Uracil	−9.9		−14.6	
Adenine	−14.6		−15.7	
Guanine	−18.9		−22.5	
SM8/M06-2X/6-31G(d,p)				
Cytosine	−15.5	−15.6	−21.0	−22.8
Thymine	−11.3	−11.5	−14.3	−15.1
Uracil	−11.3	−11.5	−15.0	−15.9
Adenine	−14.9	−15.2	−16.5	−16.9
Guanine	−20.3	−20.6	−24.8	−26.4
SM8/M06-2X/6-31+G**				
Cytosine	−15.0	−15.2	−19.9	−21.9
Thymine	−10.2	−10.4	−11.7	−12.4
Uracil	−10.2	−10.5	−12.6	−13.5
Adenine	−13.8	−14.1	−15.2	−15.7
Guanine	−18.8	−19.2	−22.6	−24.7
SM8AD/M06-2X/6-31G(d,p)				
Cytosine	−15.5	−15.6	−20.4	−22.5
Thymine	−11.2	−11.5	−13.9	−15.1
Uracil	−11.2	−11.4	−14.4	−15.5
Adenine	−15.1	−15.4	−16.3	−16.7
Guanine	−20.7	−21.0	−24.6	−26.7
SM8AD/M06-2X/6-31+G**				
Cytosine	−14.9	−15.1	−19.0	−21.3
Thymine	−10.1	−10.4	−10.7	−11.7
Uracil	−10.1	−10.4	−11.4	−12.6
Adenine	−14.2	−14.5	−15.2	−15.6
Guanine	−19.4	−19.8	−23.0	−25.3

Table 1 (*continued*)

	Chloroform		Water	
Base	Unrelaxed[b]	Relaxed[c]	Unrelaxed[b]	Relaxed[c]
SMD/M06-2X/6-31G(d,p)				
Cytosine	−14.9	−15.1	−19.4	−20.5
Thymine	−10.9	−11.0	−12.5	−13.0
Uracil	−10.9	−11.0	−13.3	−13.8
Adenine	−16.1	−16.0	−16.6	−16.7
Guanine	−20.9	−21.3	−24.4	−25.6
SMD/M06-2X/6-31+G**				
Cytosine	−16.2	−16.7	−21.1	−22.9
Thymine	−12.2	−12.4	−14.2	−15.1
Uracil	−12.3	−12.5	−15.0	−15.9
Adenine	−16.8	−16.8	−17.8	−17.9
Guanine	−22.2	−22.8	−26.1	−27.8

[a] Fixed-concentration solvation free energies ΔG^*_S (1M → 1M). [b] Without nuclear relaxation. For the SM8, SM8AD, and SMD results, the solvation free energies were calculated as a difference between the SM*x*/M06-2X/*B* total energy (*x* = 8, 8AD, D; *B* = 6-31G(d,p), 6-31+G**) in solution at the M06-2X/6-31+G** gas-phase geometry and the M06-2X/*B* total energy in the gas phase at the same M06-2X/6-31+G** gas-phase geometry. [c] With nuclear relaxation. For the SM8, SM8AD, and SMD results, the solvation free energies were calculated as a difference between the SM*x*/M06-2X/*B* total energy (*x* = 8, 8AD, D; *B* = 6-31G(d,p), 6-31+G**) in solution at the SM*x*/M06-2X/*B* liquid-phase geometry and the M06-2X/*B* total energy in the gas phase at the M06-2X/*B* gas-phase geometry. [d] Ref. 41.

Table 1 shows that the SM8, SM8AD, and SMD solvation models are usually—but not always—in good agreement with each other. The models studied here also agree reasonably well with the results of previous[41] calculations. However, the latter use less advanced solvation models and electronic structure methods.

Table 1 indicates that geometry optimization in water leads to 0.1–2.3 kcal mol^{-1} more negative solvation free energies. The nuclear relaxation effect in chloroform is smaller because of its lower dielectric permittivity. The difference between the solvation free energies calculated with and without geometry relaxation in chloroform range from −0.1 to −0.6 kcal mol^{-1}. Cytosine and guanine have the largest geometry relaxation effects. The larger nuclear relaxation effects in water lead to more negative log $P_{c/w}$ values when relaxed geometries are used (Table 2).

Table 2 shows that all the three models (SM8, SM8AD, and SMD) predict the available experimental values[78–80] of log $P_{c/w}$ equally well. Their mean unsigned errors are in the range of 0.4–1.3 log units or in the range of 0.5–1.8 kcal mol^{-1}, in energy units. These errors are somewhat larger than typical errors (< 0.8 kcal mol^{-1}) expected for SM8, SM8AD, or SMD.[54] However, there may be some uncertainties in the experimental data used for this comparison. For example, the experimental values of log $P_{c/w}$ for uracil found in the literature are quite different: −1.7 (ref. 78 and 79) and −2.99 (ref. 80). Furthermore, the mean errors are raised by the unusually large errors for adenine.

Table 2 shows noteworthy general agreement between the successful SMD model and results[41] obtained with the older SM5.42 model. This is particularly interesting because the SM5.42 model has also been compared to experiment[81] for methylated nucleic acid bases, and the agreement was encouragingly good.[82,83]

3.2 Dimerization

3.2.1 Gas-phase dimerization. The molecular structures of hydrogen-bonded and π–π stacked natural nucleobase dimers considered in the present study are shown in Fig. 1 and 2, while the structures of the alkylated base pairs here studied are provided in Fig. 3. For each pair, the considered structures include the apparent global minimum on the potential energy surface, except for unsubstituted hydrogen bonded A–T, A–U and C–G, in which preference is given to the Watson–Crick (WC) structures for their fundamental interest. We have also added a higher energy U–U conformation to Table 3 (U–U HB4) because it is the unsubstituted equivalent of the apparent global minimum of the 1-cyclohexyluracil–1-cyclohexyluracil dimer that was studied experimentally in chloroform, and will be discussed later. We note that the conformation of the global minimum of the natural U–U dimer is not available

Table 2 Log $P_{c/w}$ values for nucleobases partitioned between chloroform (c) and water (w)

Base	SM5.4[a]	SM5.42[b]	SM5.42[c]	SM5.42[d]	SM8[e]	SM8[f]	SM8AD[g]	SM8AD[h]	SMD[i]	SMD[j]	Exp
Without nuclear relaxation[k]											
Cytosine	−6.0	−3.9	−2.9	−3.9	−4.0	−3.6	−3.6	−3.0	−3.3	−3.6	−3.52[l,m]
Thymine	−2.1	−2.3	−2.6	−2.8	−2.2	−1.1	−2.0	−0.5	−1.2	−1.5	−2.26[n]
Uracil	−3.0	−3.0	−3.4	−3.5	−2.7	−1.7	−2.3	−1.0	−1.7	−2.0	−2.99[n,o]
Adenine	−3.6	−0.7	−1.0	−0.8	−1.2	−1.1	−0.8	−0.7	−0.4	−0.7	−2.48[l]
Guanine	−7.1	−2.5	−4.1	−3.6	−3.2	−2.8	−2.8	−2.6	−2.5	−2.9	−3.25[l]
MUE[p]	1.5	0.6	0.7	0.7	0.4	0.9	0.6	1.3	1.0	0.8	
With nuclear relaxation[k]											
Cytosine	−5.0				−5.3	−4.9	−5.1	−4.5	−4.0	−4.6	−3.52[l,m]
Thymine	−1.6				−2.6	−1.5	−2.6	−1.0	−1.5	−2.0	−2.26[n]
Uracil	−2.5				−3.2	−2.2	−3.0	−1.6	−2.0	−2.5	−2.99[n,o]
Adenine	−3.6				−1.3	−1.1	−1.0	−0.8	−0.5	−0.8	−2.48[l]
Guanine	−5.8				−4.2	−4.0	−4.2	−4.1	−3.2	−3.7	−3.25[l]
MUE[p]	1.3				0.9	1.0	0.9	1.3	0.8	0.8	

[a] SM5.4/AM1 calculation (ref. 41). [b] SM5.42R/AM1 calculation (ref. 41). [c] SM5.42R/HF/6-31G* calculation (ref. 41). [d] SM5.42R/BPW91/MIDI! calculation (ref. 41). [e] SM8/M06-2X/6-31G(d,p). [f] SM8/M06-2X/6-31+G**. [g] SM8AD/M06-2X/6-31G(d,p). [h] SM8AD/M06-2X/6-31+G**. [i] SMD/M06-2X/6-31G(d,p). [j] SMD/M06-2X/6-31+G**. [k] See footnotes *b* and *c* in Table 1. [l] Ref. 78. [m] Ref. 79. [n] Ref. 80. [o] Additional value: −1.7 (refs. 78 and 79). [p] Mean unsigned error with respect to experiment.

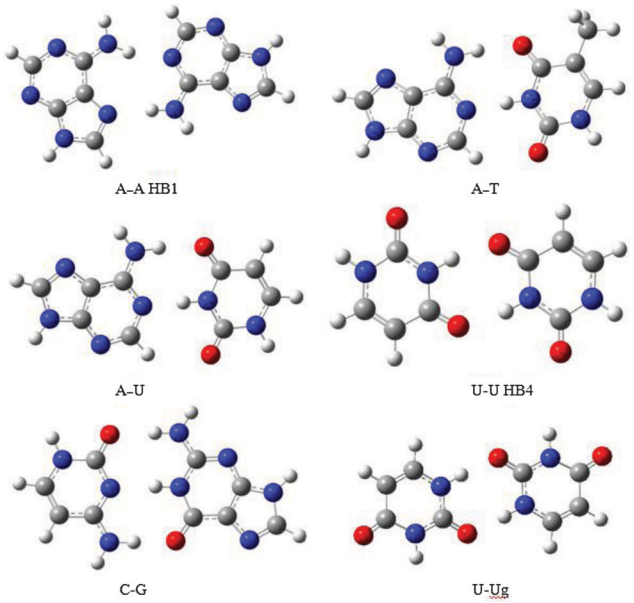


Fig. 1 Molecular structures of hydrogen-bonded nucleobase pairs.

Fig. 2 Molecular structures of π–π stacked nucleobase pairs.

for the alkylated structures considered later, as the alkyl groups creates steric hindrance that impedes the substituted dimer to adopt the lowest energy conformation of the natural one.

Table 3 contains gas-phase Born–Oppenheimer association energies for the natural dimers (ΔE_{int}) obtained using the M06-2X/6-31G(d,p) and M06-2X/6-31+G** total energies at consistently optimized geometries. These are Born–Oppenheimer interaction energies, *i.e.*, they include electronic energies and nuclear repulsion energies but not vibrational energies. Table 4 presents the corresponding standard-state (1 atm) gas-phase free energies of association ΔG^{o}_{int}(298 K).

Table 3 also contains the benchmark gas-phase electronic interaction energies[84,85] for several structures of hydrogen-bonded and π–π stacked nucleobase dimers computed using high-order quantum mechanics (HOQM; derived from composite models designed to compensate for single-particle basis set incompleteness and electron correlation). The mean unsigned deviation of our results from the benchmark results presented does not exceed 2 kcal mol^{-1}, *i.e.*, it is small enough to validate our gas-phase treatment within this uncertainty range.

In Table 5 we present gas-phase electronic interaction energies ΔE_{int} and free energies of association ΔG^{o}_{int}(298 K) calculated for different conformations of alkylated derivatives of nucleobase dimers. The alkylated derivatives that we studied are the ones investigated experimentally by Kyogoku *et al.*[51] and Sartorius and Schneider.[52] The ESI† contains the optimized Cartesian coordinates of the alkylated dimers investigated experimentally in ref. 51 and 52, and Fig. 3 shows the conformations that we utilized in this work. The conformations of alkylated nucleobase dimers employed for this study correspond to the apparent lowest energy dimers of the potential energy surface of each system considered. Most of the alkyl chains in these dimers (Fig. 3) are extended and there could be an entropic factor due to other possible conformations of the chains that were not considered here. However, we estimate this effect to be small for the dimers since the minimum energy structures of those that have non-extended alkyl chains are in most cases highly energetic compared to the case in which the alkyl groups are extended. The difference in the association free energy of the 1-hexylthymine–1-hexylthymine HB1 dimer, in which the alkyl chains are extended, and the HB2 or HB3 dimers, in which the chains are not extended, already illustrates this trend, as one can see from Table 5. Therefore, dimers in which the monomers have non-extended alkyl chains will probably be irrelevant to the calculation of the average free energy of the dimers, as illustrated by the 1-hexylthymine–1-hexylthymine case. However, it is possible that conformations of the monomers with non-extended alkyl chains play a more important role in their description than they do for the dimers, which would contribute to a destabilization of the dimers of alkylated nucleobases studied here.

In Table 6 we give estimations of the free energies of association of the 1-hexyladenine–1-hexyladenine dimers that include the HOQM corrections computed for the unsubstituted dimers. In order to get a reasonable approximation to the HOQM hydrogen bonding energy for the alkylated nucleobase dimer, we added to its M06-2X/6-31+G(d,p) or M06-2X/6-31G(d,p) electronic and free energy of association a correction corresponding to the difference between the HOQM and M06-2X association energy for the unsubstituted (natural) analog of this dimer given in Table 3. This method is defined according to the equations below

$$\Delta E_{AHOQM(alkylated)} = \Delta E_{M06\text{-}2X/B(alkylated)} + (\Delta E_{HOQM(natural)} - \Delta E_{M06\text{-}2X/B(natural)}) \quad (10)$$

$$\Delta G^{o}_{AHOQM(alkylated)} = \Delta G^{o}_{M06\text{-}2X/B(alkylated)} + (\Delta E_{HOQM(natural)} - \Delta E_{M06\text{-}2X/B(natural)}) \quad (11)$$

where "natural" refers to the unsubstituted base pair, "alkylated" refers to the substituted ones, and B represents the basis set employed for the density functional calculations.

We denote these methods as AHOQM/M06-2X/6-31G(d,p) and AHOQM/M06-2X/6-31+G**, where AHOQM denotes approximate HOQM to indicate the additional approximation of assuming that the high-order quantum mechanical correction for the alkylated cases is the same as what we computed for natural nucleobases. We only employed the AHOQM method for the 1-hexyladenine–1-hexyladenine base pair, since this correction is only reasonable for this system.

3.2.2 Liquid-phase dimerization

3.2.2.1 Liquid-phase dimerization with gas-phase geometries. In this section we will discuss results of solvation energy calculations using the molecular geometries optimized in the gas-phase at the M06-2X/6-31+G** level and gas-phase

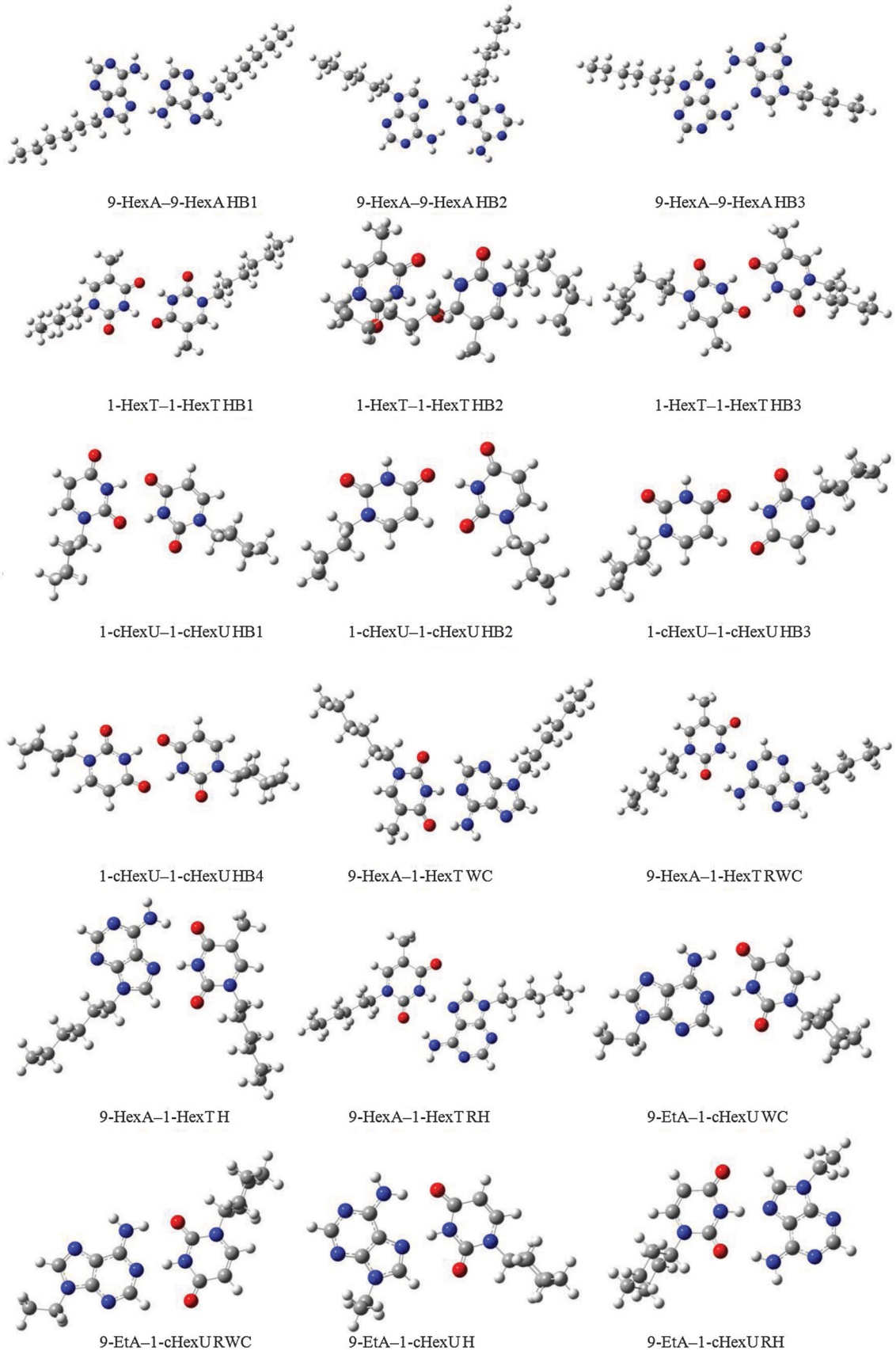


Fig. 3 Molecular structures of alkylated nucleobase pairs.

Table 3 Gas-phase Born–Oppenheimer association energies (kcal mol^{-1})

Base pair	M06-2X/6-31G(d,p)	M06-2X/6-31 + G**	HOQM
Hydrogen bonding			
A–T WC	−17.69	−14.97	−16.74[a]
A–U WC	−17.78	−15.08	
C–G WC	−31.59	−28.57	−32.06[b]
A–A HB1	−13.91	−11.65	−14.50[b]
U–Ug[c]	−20.32	−18.62	−20.69[a]
U–U HB4[d]	−14.83	−12.72	−13.70[b]
π–π stacking			
A–T	−14.59	−13.52	−11.66[a]
A–U	−13.43	−12.55	
C–G	−20.77	−19.32	−19.02[b]
Mean unsigned deviation from HOQM			
	1.17	1.90	

[a] High-order quantum mechanics (HOQM) calculation (ref. 85). [b] High-order quantum mechanics (HOQM) calculation (ref. 84). [c] Apparent global minimum of the uracil–uracil hydrogen-bonded dimer (unavailable for the alkylated U–U dimer considered in the next tables). [d] Higher energy structure of the uracil–uracil hydrogen-bonded dimer available to the alkylated U–U dimer considered in the next tables.

Table 4 Standard-state (1 atm) gas-phase association free energies (kcal mol^{-1}) at 298 K

Base pair	M06-2X/6-31G(d,p)	M06-2X/6-31 + G**
Hydrogen bonding		
A–T WC	−4.50	−1.44
A–U WC	−4.99	−1.97
C–G WC	−18.11	−14.94
A–A HB1	−0.49	2.79
U–Ug[a]	−7.32	−5.65
U–U HB4[b]	−2.40	−0.59
π–π stacking		
A–T	−0.42	1.46
A–U	0.31	1.71
C–G	−6.32	−4.66

[a] Apparent global minimum of the uracil–uracil hydrogen-bonded dimer (unavailable for the alkylated U–U dimer considered in the next tables). [b] Higher energy structure of the uracil–uracil hydrogen-bonded dimer available to the alkylated U–U dimer considered in the next tables.

free energies of association provided by M06-2X/6-31G(d,p), M06-2X/6-31 + G**, AHOQM/M06-2X/6-31G(d,p), and AHOQM/M06-2X/6-31 + G**.

Free energies of association labeled as SM*x*/M06-2X/B are always based on gas-phase free energies given by either M06-2X/B or the respective AHOQM method from Table 6. Which of these gas-phase results is used will always be specified in tables. The solvation free energies of all solutes (monomers and dimers) are given in the ESI.†

Table 7 gives free energies of association of natural unsubstituted nucleobase dimers in chloroform based on a single conformation as indicated, while Table 8 presents free energies of association of alkylated nucleotide base dimers based on both single (SC) and multiple conformations (MC) in chloroform. We note that Table 8 also contains results for free energies of association that include nuclear relaxation in solution, but discussion of those results is deferred until Section 3.2.2.2. The free energies of association were calculated according to eqn (4) and (5).

The calculated association free energies in chloroform are compared in Table 8 and throughout the paper with the corresponding energies derived from the results of experimental measurements of hydrogen bonding association constants in chloroform or $CDCl_3$ by Kyogoku *et al.*[51] and Sartorius and Schneider.[52] In particular, we compare our results to data from the experimental work of Sartorius and Schneider,[52] which is more recent than the work of Kyogoku *et al.*[51] (the latter used older and less accurate experimental techniques as discussed in ref. 52) for all nucleobase pairs, except A–U and U–U; experimental data for the latter are only available in ref. 51.

All the SM*x* models (x = 8, 8AD, and D) in Table 7 predict that in chloroform hydrogen bonding is much more favorable than π–π stacking for all of the natural dimers, in agreement with previous experimental[12–14] and theoretical studies.[25,26,32] However, while SM*x*/M06-2X/6-31G(d,p) predicts a negative standard-state free energy of hydrogen bonding association between nucleotide base dimers in chloroform solution, SM*x*/M06-2X/6-31+G(d,p) gives slightly positive standard-state free energies of hydrogen bonding for A–T and A–U.

Table 8 provides a comparison between theoretical free energies of association of alkylated nucleotide base dimers and experiment. In the first part of Table 8 M06-2X/6-31+G(d,p) and M06-2X/6-31G(d,p) gas-phase association free energies are used for the computation of the liquid-phase free energy of association. We notice that there is a tendency for theory to underestimate the interaction between the alkylated bases. Mean unsigned errors given by the methods that consider only the apparent global minimum of each level of theory to calculate the free energy of association (SC//g column) are in the range between 1.4–1.9 kcal mol^{-1}. On the other hand, by employing multiple conformations (MC//g column) the accuracy of the predicted free energies of association is improved by roughly 0.2 kcal mol^{-1}, with a MUE of 1.2 kcal mol^{-1} for SMD, 1.4 kcal mol^{-1} for SM8AD, and 1.7 kcal mol^{-1} for SM8. Therefore, conformational averaging is a significant issue in the description of the studied systems when gas-phase geometries and density-functional gas-phase results are utilized. Mean unsigned errors are smaller for SMD when the 6-31G(d,p) basis set is used, while SM8AD gives closer agreement with experiment when 6-31+G(d,p) is used.

We note that the global minimum generated by each of the methods for the alkylated A–T and A–U base pairs might differ, since most methods predict that the free energies of association of individual conformations of these dimers differ by only ~1 kcal mol^{-1}. The last three rows of Table 8, where we combine our solvation free energies with our best estimates (AHOQM) of the gas-phase values, includes our best theoretical estimates of the free energy of association of 9-hexyladenine–9-hexyladenine in chloroform for comparison to experiment. Comparison of the predictions in the last three rows of Table 8 to experiment shows that when nuclear relaxation in solution is not taken into account, all three solvation models predict the corresponding experimental data within 0.9 kcal mol^{-1}. However, they systematically underestimate the association free energy of the 9-hexyladenine–9-hexyladenine hydrogen-bonded dimer. The //liq columns in Table 8 will be discussed in Section 3.2.2.2.

Table 5 Gas-phase hydrogen bonding Born–Oppenheimer and standard-state (1 atm) free energies (kcal mol^{-1}) of association for alkylated nucleobases

	M06-2X/6-31G(d,p)		M06-2X/6-31+G**	
Base pair	ΔE_{int}	ΔG^{o}_{int}(298 K)	ΔE_{int}	ΔG^{o}_{int}(298 K)
9-HexA–9-HexA HB1[a]	−17.18	−1.63	−14.70	−0.52
9-HexA–9-HexA HB2	−12.66	2.11	−10.74	2.72
9-HexA–9-HexA HB3	−15.95	−0.32	−13.60	0.81
9-HexA–1-HexT WC[b]	−19.23	−3.59	−16.35	−2.29
9-HexA–1-HexT RWC	−18.78	−4.19	−15.99	−1.90
9-HexA–1-HexT H	−19.68	−3.93	−16.98	−2.60
9-HexA–1-HexT RH	−19.40	−3.61	−16.85	−1.94
9-EtA–1-cHexU WC[c]	−17.86	−3.04	−15.14	−0.91
9-EtA–1-cHexU RWC	−17.16	−2.83	−14.55	−0.76
9-EtA–1-cHexU H	−19.79	−5.06	−17.12	−2.99
9-EtA–1-cHexU RH	−19.40	−4.42	−16.85	−2.65
1-HexT–1-HexT HB1[d]	−14.58	−0.30	−12.39	1.24
1-HexT–1-HexT HB2	−12.93	6.11	−9.36	8.42
1-HexT–1-HexT HB3	−12.39	3.55	−9.78	5.60
1-cHexU–1-cHexU HB1[e]	−14.04	−0.35	−12.22	1.50
1-cHexU–1-cHexU HB2	−10.24	0.00	−8.63	4.87
1-cHexU–1-cHexU HB3	−11.11	2.64	−9.42	4.36
1-cHexU–1-cHexU HB4	−14.74	−0.64	−12.67	1.53

[a] The alkylated analog of adenine is 9-hexyladenine. HB1, HB2 and HB3 refer to different conformers of the homodimers presented in the table. [b] The alkylated analogs of adenine and thymine are 9-hexyladenine and 1-hexylthymine. WC corresponds to Watson–Crick base pairs; RWC corresponds to Reverse Watson–Crick base pairs; H represents Hoogsteen base pairs, and RH identifies reverse Hoogsteen base pairing. [c] The alkylated analogs of adenine and uracil are 9-ethyladenine and 1-cyclohexyluracil. [d] The alkylated analog of thymine is 1-hexylthymine. 1, 2 and 3 refer to different conformers of the 1-hexylthymine homodimer. [e] The alkylated analog of uracil is 1-cyclohexyluracil. 1, 2, 3 and 4 refer to different conformers of the 1-cyclohexyluracil homodimer.

Table 6 Approximate HOQM (AHOQM) standard-state (1 atm) gas-phase hydrogen bonding energies (kcal mol^{-1}) for alkylated nucleobases

	AHOQM/M06-2X/6-31G(d,p)[a]		AHOQM/M06-2X/6-31+G**[b]	
Base pair	ΔE_{int}	ΔG^{o}_{int}(298 K)	ΔE_{int}	ΔG^{o}_{int}(298 K)
9-HexA–9-HexA HB1[c]	−17.77	−2.22	−17.55	−3.37
9-HexA–9-HexA HB2	−13.25	1.52	−13.59	−0.13
9-HexA–9-HexA HB3	−16.54	−0.91	−16.45	−2.04

[a] Calculated with eqn (10) and (11). [b] Calculated with eqn (10) and (11). [c] The alkylated analog of adenine is 9-hexyladenine.

Now we discuss the results of calculations in water, which are presented in Table 9. It is generally accepted in the literature[16,18] that π–π stacked nucleobase dimers in water are more stable than the corresponding hydrogen-bonded dimers. However, the SM8 and SM8AD models predict the opposite for all studied dimers, whereas the SMD model predicts that the π–π stacked nucleobase dimers in water are more stable than their corresponding WC pairs in two out of three cases (namely, A–T and A–U). The role of geometry relaxation effects on the stability of the corresponding dimers will be discussed in the next section.

Table 7 Standard-state (1 M) free energies of association (kcal mol^{-1}) of natural nucleotide bases dimers in chloroform calculated using gas-phase geometries[a]

	SMx/M06-2X/6-31G(d,p)			SMx/M06-2X/6-31+G**		
Base pair	SM8	SM8AD	SMD	SM8	SM8AD	SMD
Hydrogen bonding						
Natural bases						
A–T WC	−0.78	−1.24	−1.43	0.78	0.26	2.27
A–U WC	−1.25	−1.74	−1.81	0.81	0.28	1.85
C–G WC	−9.19	−9.90	−10.70	−7.86	−8.69	−6.52
π–π stacking						
A–T	6.48	4.97	4.32	8.46	6.83	6.70
A–U	6.89	5.30	4.85	8.45	6.84	6.75
C–G	5.94	4.17	3.56	7.34	5.45	6.04

[a] Calculated at the SMx/M06-2X/6-31G(d,p) and SMx/M06-2X/6-31+G** levels of theory (x = 8, 8AD, and D) using M06-2X/6-31+G** gas-phase optimized geometries (without account for nuclear relaxation in solution); gas-phase free energies of association obtained by M06-2X/6-31G(d,p) and M06-2X/6-31+G(d,p), respectively.

The decomposition of the aqueous solvation free energies calculated at gas-phase geometries of the studied dimers into constituent polarization (G_P), electronic relaxation (ΔE_E), and cavity–dispersion–solvent-structure (G_{CDS}) components according to eqn (6) is given in Tables S3 and S4 in the ESI† for the SMx/M06-2X/6-31G(d,p) and SMx/M06-2X/6-31+G** methods, respectively. The SM8 model predicts less favorable free energies of solvation for all three stacked nucleobase pairs in comparison with SM8AD and SMD by up to 4 kcal mol^{-1}; this is mostly due to a less negative polarization contribution for the stacked nucleobase dimers. The SMD model provides the most negative solvation free energies of the stacked dimers among the three tested models due primarily to more negative polarization energies, thereby stabilizing the stacked dimers better than the other models. The SM8AD model is deemed to

Table 8 Standard-state (1 M) hydrogen bonding free energies (kcal mol^{-1}) of selected alkylated nucleotide bases in chloroform calculated using gas and liquid-phase geometries[a]

	SM*x*/M06-2X/6-31G(d,p)				SM*x*/M06-2X/6-31+G**		
	SC//g[c]	SC//liq[d]	MC//g[e]	MC//liq[f]	SC//g	MC//g	Exp[b]
				SM8			
9-HexA–9-HexA	1.07	4.21	0.97	4.09	4.22	4.09	−0.5[g]
1-HexT–1-HexT	1.63	1.23	1.63	1.23	0.88	0.88	−0.7[h]
9-EtA–1-cHexU	−0.80	−0.63	−1.04	−0.90	−0.54	−0.86	−2.7[i]
1-cHexU–1-cHexU	1.24	0.99	0.70	0.50	0.08	−0.22	−1.1[j]
9-HexA–1-HexT	−0.71	0.66	−0.91	0.40	1.25	0.67	−2.1[k]
MUE[l]	1.91	2.71	1.70	2.49	2.60	2.34	
				SM8AD			
9-HexA–9-HexA	0.78	3.80	0.67	3.67	3.68	3.56	−0.5[g]
1-HexT–1-HexT	1.68	1.36	1.68	1.36	0.84	0.84	−0.7[h]
9-EtA–1-cHexU	−1.37	−1.14	−1.59	−1.39	−1.19	−1.50	−2.7[i]
1-cHexU–1-cHexU	1.14	1.11	0.61	0.59	−0.09	−0.39	−1.1[j]
9-HexA–1-HexT	−1.03	0.37	−1.34	0.05	0.81	0.17	−2.1[k]
MUE	1.66	2.52	1.44	2.29	2.23	1.97	
				SMD			
9-HexA–9-HexA	0.56	3.82	0.42	3.63	1.98	1.82	−0.5[g]
1-HexT–1-HexT	1.40	0.76	1.40	0.76	3.55	3.55	−0.7[h]
9-EtA–1-cHexU	−1.58	−1.61	−1.83	−1.87	1.03	0.72	−2.7[i]
1-cHexU–1-cHexU	0.79	0.46	0.26	−0.03	3.43	3.07	−1.1[j]
9-HexA–1-HexT	−1.32	−0.14	−1.60	−0.43	1.43	0.78	−2.1[k]
MUE	1.39	2.08	1.21	1.84	3.70	3.42	
				AHOQM[m]			
				SM8			
9-HexA–9-HexA	0.48	3.62	0.38	3.61	1.37	1.24	−0.5[g]
				SM8AD			
9-HexA–9-HexA	0.19	3.21	0.08	3.08	0.83	0.71	−0.5[g]
				SMD			
9-HexA–9-HexA	−0.03	3.85	−0.17	3.04	−0.87	−1.03	−0.5[g]

[a] Free energies of solvation computed by SM*x*/M06-2X/6-31G(d,p) or SM*x*/M06-2X/6-31+G** with gas-phase geometries (SC or MC//g) or liquid-phase optimized ones (SC or MC//liq); gas-phase free energies of association obtained by M06-2X/6-31G(d,p) and M06-2X/6-31+G(d,p) in the first three subsections and by AHOQM in the next three subsections. [b] Experimental ΔG°_{int}(298 K) values for hydrogen-bonded dimers. Unless noted otherwise, these quantities are derived in the present work from the corresponding equilibrium constants K (in L mol^{-1}) at 298 K as $\Delta G^{\circ}_{int} = -RT \ln K$. [c] SC//g refers to calculations that considered a single conformation of a particular dimer to compute its free energy of association, without accounting for nuclear relaxation in solution. [d] SC//liq refers to liquid-phase calculations that allowed dimers and the monomers to relax in solution, and used the lowest minimum energy conformer of a particular dimer to compute the free energy of association. [e] MC//g refers to calculations that employed a Boltzmann weighted average that included multiple conformations (in particular, three or four) for each dimer to compute the free energy of association, without accounting for nuclear relaxation in liquid-phase computations. [f] MC//liq refers to liquid-phase calculations that allowed dimers and monomers to relax in solution, and considered three or four conformations of each dimer in order to compute Boltzmann weighted average free energies of association. [g] Using $K = 2.4 \pm 0.2$ in $CDCl_3$ (ref. 52). [h] Using $K = 3.5 \pm 1.2$ in $CDCl_3$ (ref. 52). [i] Using $K = 100 \pm 20$ in $CHCl_3$ (ref. 51). [j] Using $K = 6.1 \pm 0.6$ in $CHCl_3$ (ref. 51). [k] Using $\Delta G = -2.08$ kcal mol^{-1} given for the Watson–Crick base pair A–T in $CDCl_3$ in ref. 52. [l] Mean unsigned error with respect to experiment. [m] Free energies of solvation computed by SM*x*/M06-2X/6-31G(d,p) and SM*x*/M06-2X/6-31+G** (x = 8, 8AD, and D) using the M06-2X/6-31+G** gas-phase optimized geometries (without account for nuclear relaxation in solution), or SM*x*/M06-2X/6-31G(d,p) liquid-phase geometries (SC or MC//liq), but with gas-phase free energies of association given by the AHOQM values provided in Table 6.

be more accurate than the earlier SM8 model, owing to a better description of the bulk electrostatics. This improvement is important in many cases in which the individual partial atomic charges are asymmetrically situated in the molecule, *i.e.*, located near the dielectric boundary rather than at the center of the molecular cavity. Table 9, S3, and S4 (ESI†) show that the SM8AD model stabilizes the stacked base pairs by 3–4 kcal mol^{-1} relative to SM8, which makes the stability of hydrogen bonding and stacking dimers closer in value when SM8AD is employed (hydrogen bonding is more favorable than stacking in A–T and A–U by ~2 kcal mol^{-1}, whereas this difference is 4–5 kcal mol^{-1} when SM8 is applied).

Based on the results discussed above, we conclude that there is no tested combination of gas-phase and liquid-phase treatment *with the gas-phase optimized geometries* that predicts that π–π stacking is more favorable than hydrogen bonding for the C–G dimer in water. All the methods studied in this subsection indicate that the hydrogen-bonded C–G nucleobase pair is favored relative to the stacked base pair by a few kcal mol^{-1}. Based on calculations with up to six explicit water molecules, Sivanesan *et al.*[38] attributed the stabilization of the stacked C–G dimer over the hydrogen-bonded dimer in water to the better hydration of the stacked structure that can accommodate 5–6 water molecules in its first solvation shell in comparison to the hydrogen-bonded structure that can accommodate only 4–5 water molecules. Kabelác and Hobza[29] carried out molecular dynamics simulations of the nucleobase dimers microsolvated with up to 16 explicit water molecules by using nonpolarizable force fields and concluded that the fraction of stacked structures in water increases at higher concentrations of the solvent. The π–π stacking interaction in the A–T complex becomes favored when only two water molecules are added to the system whereas

Table 9 Standard-state (1 M) free energies of association (kcal mol^{-1}) of natural nucleobases in water calculated using gas-phase geometries[a]

	SM*x*/M06-2X/6-31G(d,p)[b]			SM*x*/M06-2X/6-31 + G**[c]		
Base pair	SM8	SM8AD	SMD	SM8	SM8AD	SMD
Hydrogen bonding						
A–T WC	1.88	1.57	1.35	2.65	2.02	5.23
A–U WC	1.38	0.99	1.01	2.93	2.37	4.84
C–G WC	−3.31	−3.87	−4.12	−2.89	−3.48	0.26
π-π stacking						
A–T	6.09	2.88	0.71	6.93	3.25	3.31
A–U	6.35	3.11	1.23	7.09	3.65	3.36
C–G	7.27	3.49	2.48	7.93	4.30	5.30

[a] Calculated at the SM*x*/M06-2X/6-31G(d,p) and SM*x*/M06-2X/6-31 + G** level of theory (*x* = 8, 8AD, and D) using M06-2X/6-31 + G** gas-phase optimized geometries (without account for nuclear relaxation in solution). [b] Gas-phase association free energies given by M06-2X/6-31G(d,p). [c] Gas-phase association free energies given by M06-2X/6-31 + G**.

Table 10 Standard-state (1 M) free energies (kcal mol^{-1}) of association of natural nucleobases in chloroform calculated with nuclear relaxation[a]

	SM*x*/M06-2X/6-31G(d,p)			SM*x*/M06-2X/6-31 + G**		
Base pair	SM8	SM8AD	SMD	SM8	SM8AD	SMD
Hydrogen bonding						
A–T WC	0.35	−0.15	−1.53	1.86	1.29	2.14
A–U WC	−0.25	−0.78	−1.94	1.79	1.22	1.74
C–G WC	−8.55	−9.37	−10.46	−7.26	−8.17	−5.98
π–π stacking						
A–T	7.77	6.19	4.33	9.17	7.57	6.39
A–U	8.03	6.41	4.77	9.20	7.62	6.35
C–G	6.62	4.62	4.04	7.92	5.91	5.56

[a] Calculated at the SM*x*/M06-2X/6-31G(d,p) and SM*x*/M06-2X/6-31 + G** level of theory (*x* = 8, 8AD, and D) with account for nuclear relaxation in solution as described in the text; gas-phase association free energies obtained by M06-2X/6-31G(d,p) and M06-2X/6-31 + G(d,p), respectively.

for the C–G complex the π–π stacking interaction dominates only when at least eight water molecules are included explicitly.[29] On the other hand, there is only indirect experimental evidence for the association of nucleosides (nucleobases covalently bonded to a ribose or deoxyribose sugar) in water by vertical stacking, and indeed experiments do not distinguish between G–C stacking and, for instance, G–G stacking, or oligomeric stacks that do not require stacked dimers themselves to be stable.[16,18–20]

3.2.2.2 Liquid-phase dimerization with liquid-phase geometries. It is important to separately examine the effect of geometry relaxation because most liquid-phase simulations use gas-phase geometries. Nuclear (geometry) relaxation is driven by G_P in eqn (6) becoming more negative. This makes ΔE_E and ΔE_N more positive and has a small effect in G_{CDS}. The net result is always that ΔG_S^* is more negative. Note that the nuclear relaxation effects in solution can be safely neglected in general unless there are one or more "soft" (*i.e.*, low-frequency) normal modes corresponding to the large-amplitude motion that can lead to substantial changes in the spatial charge distribution. An example of a system with such low-frequency modes, is a π–π stacked complex of two nucleobases.

We performed geometry optimization of the natural nucleobase monomers and dimers in solution using the SM8, SM8AD, and SMD solvation models at the M06-2X/6-31G(d,p) and M06-2X/6-31 + G** levels of theory. Geometry optimization of the alkylated monomers and 3 or 4 minimum-energy structures of various hydrogen-bonded dimers in solution was carried out at the SM*x*/M06-2X/6-31G(d,p) level.

The solvation free energies of individual compounds that include solvent-induced geometry relaxation (see Table 1 for nucleobase monomers and Table S2 in the ESI† for nucleobase dimers) were calculated as a difference between the free energy in solution at the corresponding liquid-phase geometry and the free energy in the gas phase at the gas-phase geometry calculated with the same basis set used for the liquid-phase computation.

Tables 8, 10–12 show the final SM*x*/M06-2X/*B* (*x* = 8, 8AD, D; *B* = 6-31G(d,p), 6-31 + G**) free energies of association of the tested dimers in chloroform (Tables 8, 10 and 11) and water (Table 12). The SC columns in Table 8 and all of Tables 10–12 were calculated by eqn (4) and (5), whereas the MC columns of Table 8 were calculated using eqn (4), (5) and (9).

Table 10 shows that in almost all cases relaxation of the structures of the unsubstituted natural nucleobase dimers and monomers in solution does not change the free energies of association appreciably when compared to the results obtained with gas-phase structures in Table 7. In addition, the results provided in Table 10 confirm the already observed experimental trend (Table 7) that, in chloroform, hydrogen bonding is more favorable than stacking for the systems here studied.

In Table 11 we compare the relative stabilities of each of the conformers of the alkylated nucleobase dimers here studied. We see that the Hoogsteen base pair is the dominant one in a chloroform solution of 9-ethyladenine and 1-cyclohexyluracil, whereas it is the third most stable conformation of the 9-hexyladenine–1-hexylthymine dimer in the same solvent. According to the methods employed, the reverse WC conformation is the most stable 9-hexyladenine–1-hexylthymine hydrogen-bonded pair. SM8 and SM8AD disagree with SMD with respect to the dominant dimer in a 9-hexyladenine–9-hexyladenine chloroform solution, since the 9-hexA–9-hexA HB3 dimer has a lower free energy of association according to the latter, while the HB1 dimer is the most stable according to the formers. On the other hand, all three methods agree in the prediction that the 1-hexylthymine–1-hexylthymine HB1 is the most stable dimer in a chloroform solution compared to the other two tested. The difference between the free energies of association given by the models employed in this work for the 1-cyclohexyluracil–1-cyclohexyluracil HB1, HB2 and HB4 dimers is between 0.01 and 0.5 kcal mol^{-1} and all of them give the HB4 dimer as the most stable in chloroform. We note that, except for the alkylated thymine homodimer, more than one conformation is important for the description of each dimer in solution, and therefore the approach that considers multiple conformations is required to compute accurate free energies.

Table 8 contains the free energies of association provided for the alkylated nucleobase dimers for which the geometries were optimized in chloroform with the SM*x* models (SC or MC//liq).

Table 11 SM*x*/M06-2X/6-31G(d,p) free energies of association of nucleobase dimers in chloroform calculated with nuclear relaxation[a]

	SM*x*/M06-2X/6-31G(d,p)		
Base pair	SM8	SM8AD	SMD
9-HexA–9-HexA HB1	4.21	3.80	4.44
9-HexA–9-HexA HB2	6.58	6.31	6.03
9-HexA–9-HexA HB3	5.11	4.64	3.82
1-HexT–1-HexT HB1	1.23	1.36	0.76
1-HexT–1-HexT HB2	10.28	10.14	9.31
1-HexT–1-HexT HB3	6.03	6.24	5.49
9-EtA–1-cHexU WC	0.65	0.22	−0.29
9-EtA–1-cHexU RWC	0.66	0.25	−0.34
9-EtA–1-cHexU H	−0.63	−1.14	−1.61
9-EtA–1-cHexU RH	−0.01	−0.48	−0.91
1-cHexU–1-cHexU HB1	1.26	1.41	0.99
1-cHexU–1-cHexU HB2	1.25	1.26	0.54
1-cHexU–1-cHexU HB3	4.25	4.18	3.60
1-cHexU–1-cHexU HB4	0.99	1.11	0.46
9-HexA–1-HexT WC	1.50	1.10	0.69
9-HexA–1-HexT RWC	0.66	0.37	−0.14
9-HexA–1-HexT H	1.68	1.18	0.71
9-HexA–1-HexT RH	1.87	1.43	1.01
AHOQM[b]			
9-HexA–9-HexA HB1	3.62	3.21	3.85
9-HexA–9-HexA HB2	5.99	5.72	5.44
9-HexA–9-HexA HB3	6.95	4.05	3.23

[a] Free energies of solvation computed by SM*x*/M06-2X/6-31G(d,p) with liquid-phase geometries; gas-phase association free energies obtained by M06-2X/6-31G(d,p), except for the AHOQM section. [b] Free energies of solvation computed by SM*x*/M06-2X/6-31G(d,p), and gas-phase free energies of association given by the AHOQM values provided by Table 6.

Table 12 Standard-state (1 M) free energies of association (kcal mol^{-1}) of natural nucleobases in water calculated with nuclear relaxation[a]

	SM*x*/M06-2X/6-31G(d,p)			SM*x*/M06-2X/6-31+G**		
Base pair	SM8	SM8AD	SMD	SM8	SM8AD	SMD
Hydrogen bonding						
A–T WC	2.21	1.92	1.36	2.89	2.26	5.22
A–U WC	1.64	1.25	0.98	3.22	2.59	4.77
C–G WC	−1.44	−1.58	−2.84	−0.68	−1.02	2.14
π–π stacking						
A–T	6.53	3.09	0.74	6.41	2.59	3.09
A–U	6.82	3.42	1.26	6.82	3.21	3.07
C–G	7.68	4.93	2.03	9.05	5.48	3.49

[a] Calculated at the SM*x*/M06-2X/6-31G(d,p) and SM*x*/M06-2X/6-31+G** level of theory (*x* = 8, 8AD, and D) with account for nuclear relaxation in solution as described in the text; gas-phase association free energies obtained by M06-2X/6-31G(d,p) and M06-2X/6-31+G(d,p), respectively.

It is shown that such an approach does not generally improve the agreement between theory and experiment. When only a single conformation is used, the SM*x*/6-31G(d,p) mean unsigned errors change from 1.4–1.9 to 2.1–2.8 kcal mol^{-1} after inclusion of nuclear relaxation, and conformational averaging also increases the mean errors by 0.6–0.9 kcal mol^{-1} compared to the analog calculations done with gas-phase geometries. Using gas-phase AHOQM association free energies along with the free energies of solvation obtained for optimized structures in solution to get the liquid-phase association energies does not improve on the accuracy of the results based on gas-phase M06-2X/6-31G(d,p) free energies. The mean unsigned errors increase by 0.7–1.7 kcal mol^{-1} when the AHOQM gas-phase association free energies are used to generate the liquid-phase free energies of hydrogen bonding.

Some caution should be exercised in comparing theory and experiment, as the experimental data have large uncertainties (see footnotes in Table 8), partially due to neglecting self-association of individual monomers when deriving the equilibrium constant values from experimental (NMR titration) data in the case of binary dimers and due to neglecting higher aggregates in all other cases.[52]

In addition, a large number of structures can be found in solution and the energetic order of the structures can be different than in the gas-phase, *e.g.*, structural rearrangement of the solvent might stabilize structures that otherwise would be highly unstable, and would not be minima on the free energy surface of the solute.[96–98] Although this means that one should be cautious about accepting conclusions based on only a few structures, Sartorius and Schneider found Nuclear Overhauser Effect (NOE) signals that correspond to the four alkylated A–T hydrogen-bonded structures (Watson–Crick, reverse Watson–Crick, Hoogsteen and reverse Hoogsteen base pairs) that we included in our description, which validates the choice of structures we made in this case.[52]

Table 12 (as compared with Table 9) indicates that accounting for nuclear relaxation in the case of π–π stacked dimers in water does not alter our previous conclusion that the SM8 and SM8AD models predict that the WC hydrogen-bonded nucleobase dimers in water are more stable than the corresponding π–π stacked dimers. The SMD/6-31+G** model still predicts that the π–π stacked nucleobase dimers in water are more stable at least in two out of three cases (WC A–T and WC A–U) by an average of 2 kcal mol^{-1}. In all these cases, the C–G dimer in water favors the hydrogen bonding interaction but the stability of the π–π stacked C–G structure relative to the corresponding WC structure increases by 3.7 kcal mol^{-1} if the SMD/6-31+G** model includes account for nuclear relaxation in solution.

A treatment that includes nuclear relaxation of the solute is supposed to improve on a method that uses gas-phase geometries, but as seen in Table 12 that is not enough for the prediction of all qualitative features of the dimerization of nucleobases in aqueous solutions by the SM*x* models. Former studies shed light on this, as very specific arrangements of water molecules are necessary to make stacked dimers more favorable than hydrogen-bonded ones in water and other polar solvents.[30,32,36] In addition, there might be stacked geometries in water that are responsible for this preference and might not be accessible to continuum solvent models as stationary points on the free energy surface. The fact that SMD is able to predict the favoring of stacked A–T and A–U compared to their base pairs is already encouraging since solvation phenomena depending on specific hydrogen bonding patterns are hard to describe in models that do not include explicit solvent molecules.

We can also consider the effect of nucleobase dimer energetics on nucleic acid duplexes. It is known[86,87] that RNA duplexes are more stable than DNA duplexes under laboratory and physiological conditions but it has not been established that hydrogen bonding in the WC adenine–uracil dimer is energetically

Table 13 Standard-state (1 M) free energies of association (kcal mol^{-1}) of natural nucleobases in chloroform and water using the multiprotocol method[a]

	Chloroform		Water	
Base pair	H-bonding (WC)	π–π stacking	H-bonding (WC)	π–π stacking
A–T	1.52	4.10	4.41	−0.23
A–U	1.08	4.02	4.00	−0.26
C–G	−7.29	2.49	0.33	−1.10

[a] These energies were calculated using several protocols: the M06-2X/6-31 + G** method for gas-phase association free energies, SMD/M06-2X/6-31 + G** with nuclear relaxation for the solvation free energies of π–π stacked nucleobase dimers, and SMD/M06-2X/6-31G(d,p) with nuclear relaxation for the solvation free energies of hydrogen-bonded nucleobase dimers and monomeric nucleobases.

more favorable than in the WC adenine–thymine case.[86,87] Except for the SM8 and SM8AD computations using diffuse functions, all methods tested here predict that WC A–U hydrogen bonding is 0.5–0.7 kcal mol^{-1} stronger than WC A–T hydrogen bonding in water.[88–92] Our work supports the findings[90] of Vakonakis and LiWang on the hydrogen-bond stabilization of A–U over A–T obtained in a study of the correlation of deuterium isotope effects and ^{13}C shielding constants as well as the findings of Acharya *et al.*[91] obtained in a study of the correlation between the hydrogen bond strength and experimentally measured pK_a values. Pérez *et al.*[93] attribute the greater thermal stability of RNA over DNA duplexes to other factors such as sugar puckering, intra and intermolecular stacking, and other subtle contributions to the global conformations of DNA and RNA duplexes as well as to bulk solvation. They also found that the free energy of association of A–T is lower compared to A–U when the geometry of the WC A–T pair is changed in order to resemble the structure of A–U in RNA. In the present study we find that solvation effects do not play a significant role in any extra stabilization of the hydrogen-bonded A–U nucleobase dimer compared to the corresponding A–T dimer: the difference between the free energy of association of A–T and the free energy of association of A–U calculated in the gas phase changes only slightly (less than 1 kcal mol^{-1}) when the same quantities are computed for the dimers in an aqueous solution.

3.2.3 A dual-basis methodology for the treatment of nucleotide bases interactions. In this section we explore the use of a dual-basis protocol that employs a selection of two basis sets to obtain the most reliable results for the nucleobase dimers in water. As shown above, the SMD/M06-2X/6-31 + G** model including nuclear relaxation predicts the most favorable π–π stacking interactions in water. Therefore, we select this method to compute the solvation free energies of π–π stacked nucleobase dimers. To compute the solvation free energies of hydrogen-bonded WC dimers, we again account for nuclear relaxation, but we select the SMD/M06-2X/6-31G(d,p) model. The latter model is also selected to compute the solvation free energies of monomeric nucleobases. The gas-phase association free energies are computed using the M06-2X/6-31 + G** method. Results of calculations using this dual-basis methodology are presented in Table 13.

Table 13 indicates that the dual-basis protocol predicts the correct experimental trend showing the preference of π–π stacked structures over hydrogen-bonded structures for all the tested WC nucleobase pairs (A–T, A–U, C–G) in water.

4. Conclusions

Many simulation methods in the literature are very approximate, and a careful comparison of the sensitivity of predictions to the method may provide some general guidance. We have presented here the chloroform–water partition coefficients of selected nucleotide bases (namely, cytosine, thymine, uracil, adenine, and guanine) and the free energies of association of these nucleobases in chloroform and water using the Minnesota density functional M06-2X and the Minnesota implicit solvation models SM8, SM8AD, and SMD combined with the 6-31G(d,p) and 6-31 + G** basis sets. These calculations are based on a single solute conformation in each case. For the free energies of association of alkylated nucleotide bases in chloroform, which we compare with previously published experimental data, we applied a more complete methodology, with the difference being that we included several solute conformations. The predicted chloroform–water partition coefficients values are in qualitative agreement with available experimental values for all the tested methods. But the results also demonstrate the sensitivity of the conclusions to geometry and to the choice of a basis set used in the treatment of solvation effects. Most of the tested methods predict that the hydrogen bonding in the adenine–uracil dimer in water is stronger than the hydrogen bonding in the corresponding adenine–thymine dimer in agreement with the observation that the RNA duplexes are more stable than the corresponding DNA duplexes under laboratory and physiological conditions. All the tested protocols correctly predict the preference for hydrogen bonding over π–π stacking in the gas-phase nucleobase dimers and in the corresponding dimers solvated in the non-polar solvent (chloroform). However, there is a qualitative discrepancy between theory and conventional wisdom for the description of nucleobase dimers solvated in water. Interpretation of existing experiment suggests π–π stacked aggregates of undetermined structure and molecularity to be more favorable in water than the hydrogen-bonded structures. However, in our calculations on the heterodimers, only SMD provides such a description for two out of the three tested nucleobase pairs (namely, for A–T and A–U), whereas no solvation model predicts the preference for π–π stacking over WC hydrogen bonding in the aqueous C–G dimer. Nevertheless, when a dual-basis protocol is used, where one selects each basis set where it is most successful, theory predicts that for the nucleobase pairs in water the stacking interaction is more favorable than the chosen hydrogen bonding structure for all of the three nucleobase pairs. It may be that implicit solvation treatment of the π–π

stacked nucleobase dimers in aqueous solution can be improved with inclusion of one or two water molecules explicitly, although that introduces new uncertainties in the placement of the explicit solvent molecules.[94,95]

Acknowledgements

This work was supported by the Army Research Office under grant US ARMY RES LAB/W911NF09-1-0377 and by the National Science Foundation under grants CHE06-10183, CHE07-04974, CHE09-52054, and CHE09-56776.

References

1 C. A. Hunter, *Chem. Soc. Rev.*, 1994, **23**, 101–109.
2 P. Hobza and J. Sponer, *Chem. Rev.*, 1999, **99**, 3247–3276.
3 K. Müller-Dethlefs and P. Hobza, *Chem. Rev.*, 2000, **100**, 143–168.
4 M. W. Mosesson, K. R. Siebenlist and D. A. Meh, *Ann. N. Y. Acad. Sci.*, 2001, **936**, 11–30.
5 J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3–31.
6 A. E Meyer, R. K. Castellano and F. Diederich, *Angew. Chem., Int. Ed.*, 2003, **42**, 1210–1250.
7 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
8 L. Stryer, *Biochemistry*, W. H. Freeman and Company, New York, 1999.
9 K. I. Yanson, A. B. Teplitsky and L. F. Sukhodub, *Biopolymers*, 1979, **18**, 1149–1170.
10 M. Kabelác, E. C. Sherer, C. J. Cramer and P. Hobza, *Chem.–Eur. J.*, 2007, **13**, 2067–2077.
11 R. R. Shoup, H. T. Miles and E. D Becker, *Biochem. Biophys. Res. Commun.*, 1966, **23**, 194–201.
12 L. Katz and S. Penman, *J. Mol. Biol.*, 1966, **15**, 220–231.
13 Y. Kyogoku, R. C. Lord and A. Rich, *J. Am. Chem. Soc.*, 1967, **89**, 496–504.
14 J. H. Miller and H. M. Sobell, *J. Mol. Biol.*, 1967, **24**, 345–350.
15 R. A. Newmark and C. R. Cantor, *J. Am. Chem. Soc.*, 1968, **90**, 5010–5017.
16 M. P. Schweizer, A. D. Broom, P. O. P. Ts'o and D. P. Hollis, *J. Am. Chem. Soc.*, 1968, **90**, 1042–1055.
17 T. Imoto, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1997, **475**, 409.
18 P. O. P. Ts'o, I. S. Melvin and A. C. Olson, *J. Am. Chem. Soc.*, 1963, **85**, 1289–1296.
19 N. I. Nakano and S. J. Igarashi, *Biochemistry*, 1970, **9**, 577–583.
20 T. N. Solie and J. A. Schellman, *J. Mol. Biol.*, 1968, **33**, 61–77.
21 A. Pohorille, L. R. Pratt, S. K. Burt and R. D. MacElroy, *J. Biomol. Struct. Dyn.*, 1984, **1**, 1257–1280.
22 A. Pohorille, S. K. Burt and R. D. MacElroy, *J. Am. Chem. Soc.*, 1984, **106**, 402–409.
23 P. Cieplak and P. A. Kollman, *J. Am. Chem. Soc.*, 1988, **110**, 3734–3739.
24 L. X. Dang and P. A. Kollman, *J. Am. Chem. Soc.*, 1990, **112**, 503–507.
25 J. Pranata and W. L. Jorgensen, *Tetrahedron*, 1991, **47**, 2491–2501.
26 J. Pranata, S. G. Wierschke and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810–2819.
27 V. I. Danilov, N. V. Zheltovsky, O. N. Slyusarchuk, V. I. Poltev and J. L. Alderfer, *J. Biomol. Struct. Dyn.*, 1997, **15**, 69–80.
28 M. Kabelác, F. Ryjacek and P. Hobza, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4906–4909.
29 M. Kabelác and P. Hobza, *Chem.–Eur. J.*, 2001, **7**, 2067–2074.
30 M. Kabelác, L. Zendlová, D. Reha and P. Hobza, *J. Phys. Chem. B*, 2005, **109**, 12206–12013.
31 L. Zendlova, P. Hobza and M. Kabelac, *ChemPhysChem*, 2006, **7**, 439–447.
32 L. Zendlova, P. Hobza and M. Kabeläc, *J. Phys. Chem. B*, 2007, **111**, 2591–2609.
33 J. Florian and A. Warshel, *J. Phys. Chem. B*, 1997, **101**, 5583–5595.
34 J. Florian, J. Sponer and A. Warshel, *J. Phys. Chem. B*, 1999, **103**, 884–892.
35 J. Langlet, C. Giessner-Prettre, B. Pullman, P. Claverie and D. Piazzola, *Int. J. Quantum Chem.*, 1980, **18**, 421–437.
36 D. Sivanesan, I. Sumathi and W. J. Welsh, *Chem. Phys. Lett.*, 2003, **367**, 351–360.
37 I. Komura, Y. Ishikawa, T. Tsukamoto, T. Natsume and N. Kurita, *THEOCHEM*, 2008, **862**, 122–129.
38 D. Sivanesan, K. Babu, S. R. Gadre, V. Subramanian and T. Ramasami, *J. Phys. Chem. A*, 2000, **104**, 10887–10894.
39 T. Yamaguchi and C. Nagata, *J. Theor. Biol.*, 1977, **69**, 693–707.
40 C. J. Cramer and D. G. Truhlar, *Chem. Phys. Lett.*, 1992, **198**, 74–80.
41 J. Li, C. J. Cramer and D. G. Truhlar, *Biophys. Chem.*, 1999, **78**, 147–155.
42 C. Aleman, *Chem. Phys. Lett.*, 1999, **302**, 461–470.
43 G. Chung, H. Oh and D. Lee, *THEOCHEM*, 2005, **730**, 241–249.
44 Y. Zhao and D. G Truhlar, *J. Chem. Theory Comput.*, 2005, **1**, 415–432.
45 Y. Zhao and D. G Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2701–2705.
46 Y. Zhao and D. G Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 289–300.
47 Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 364–382.
48 Y. Zhao and D. G Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
49 Y. Zhao and D. G Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.
50 Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2008, **4**, 1849–1868.
51 Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1969, **179**, 10–17.
52 J. Sartorius and H.-J. Schneider, *Chem.–Eur. J.*, 1996, **2**, 1446–1452.
53 A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2011–2033.
54 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2009, **5**, 2447–2464.
55 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
56 C. J. Cramer and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 760–768.
57 C. J. Cramer and D. G. Truhlar, *Acc. Chem. Res.*, 2009, **42**, 493–497.
58 O. Tapia, in *Quantum Theory of Chemical Reactions*, ed. R. Daudel, A. Pullman, A. L. Salem and A. Viellard, Wiley, London, 1981, vol. 2, p. 25.
59 S. C. Tucker and D. G. Truhlar, *Chem. Phys. Lett.*, 1989, **157**, 164–170.
60 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
61 G. J. Hoijtink, E. de Boer, P. H. van der Meij and W. P. Weijland, *Recl. Trav. Chim. Pays-Bas. Belg.*, 1956, **75**, 487.
62 J. W. Storer, D. J. Giesen, C. J. Cramer and D. G. Truhlar, *J. Comput.-Aided Mol. Des.*, 1995, **9**, 87–110.
63 R. M. Olson, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2046–2054.
64 C. C. Chamber, G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, **100**, 16385–16398.
65 T. Grycuk, *J. Chem. Phys.*, 2003, **119**, 4817–4826.
66 E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032–3041.
67 B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **106**, 5151–5158.
68 B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B*, 1997, **101**, 10506–10517.
69 J. Tomasi, B. Mennucci and E. Cancès, *THEOCHEM*, 1999, **464**, 211–226.
70 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2008, **4**, 877–887.
71 *Minnesota Gaussian Functional Module, version 4.1*, Y. Zhao and, D. G. Truhlar, University of Minnesota, Minneapolis, MN, 2008.
72 *Minnesota Gaussian Solvation Module, version 2010*, R. M. Olson, A. V. Marenich, A. C. Chamberlin, C. P. Kelly, J. D. Thompson, J. D. Xidos, J. Li, G. D. Hawkins, P. D. Winget, T. Zhu, D. Rinaldi, D. A. Liotard, C. J. Cramer, D. G. Truhlar, and M. J. Frisch, University of Minnesota, Minneapolis, MN, 2010.

73 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revisions D.02 and E.01*, Gaussian, Inc., Wallingford, CT, 2004.

74 M. M. Francl, W. J. Petro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.

75 J. Zheng, I. M. Alecu, B. J. Lynch, Y. Zhao and D. G. Truhlar, *Database of Frequency Scaling Factors for Electronic Structure Method*, University of Minnesota, Minneapolis, MN, 2009, http://comp.chem.umn.edu/truhlar/freqscale (accessed Nov. 12, 2010).

76 I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 2872–2887.

77 P. Winget, D. M. Dolney, D. J. Giesen, C. J. Cramer and D. G. Truhlar, *Minnesota Solvent Descriptor Database, version 1999*, University of Minnesota, Minneapolis, MN, 1999, http://comp.chem.umn.edu/solvation/mnsddb.pdf (accessed July 28, 2010).

78 P. Jow and C. Hansch, Pomona College, unpublished analysis; R. Vasanwala and C. Hansch, Pomona College, unpublished results; in *MedChem Software Masterfile*, ed. A. J. Leo, BioByte Corp., Claremont, CA, 1994.

79 K. Tsukada, S. Ueda and R. Okada, *Chem. Pharm. Bull.*, 1984, **32**, 1929–1935.

80 P. Shih, L. G. Pedersen, P. R. Gibbs and R. Wolfenden, *J. Mol. Biol.*, 1998, **280**, 421–430.

81 P. M. Cullis and R. Wolfenden, *Biochemistry*, 1981, **20**, 3024–3028.

82 D. J. Giesen, C. C. Chambers, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 1997, **101**, 5084–5088.

83 J. E. Eksterowicz, J. L. Miller and P. A. Kollman, *J. Phys. Chem. B*, 1997, **101**, 10971–10975.

84 P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

85 T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.

86 M. Riley, B. Maling and M. J. Chanberlin, *J. Mol. Biol.*, 1966, **20**, 359; V. A. Bloomfield, D. M. Crothers and I. Tinoco, *Nucleic Acids: Structure, Properties and Functions*, University Science Books, Sausalito, 2000.

87 S. Nakano, T. Kanzaki and N. Sugimoto, *J. Am. Chem. Soc.*, 2004, **126**, 1088–1095.

88 M. Swart, C. Fonseca Guerra and F. M. Bickelhaupt, *J. Am. Chem. Soc.*, 2004, **126**, 16718–16719.

89 S. Wang and E. T. Kool, *Biochemistry*, 1995, **34**, 4125–4132.

90 I. Vakonakis and A. C. LiWang, *J. Am. Chem. Soc.*, 2004, **126**, 5688–5689.

91 P. Acharya, P. Cheruku, S. Chatterjee, S. Acharya and J. Chattopadhyaya, *J. Am. Chem. Soc.*, 2004, **126**, 2862.

92 Y.-I. Kim, M. N. Manalo, L. M. Peréz, A. LiWang and A. J. Biomol, *NMR*, 2006, **34**, 229–236.

93 A. Perez, J. Sponer, P. Jurecka, P. Hobza, F. J. Luque and M. Orozco, *Chem.–Eur. J.*, 2005, **11**, 5062–5066.

94 C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161–2200.

95 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 2829–2844.

96 V. I. Danilov and T. van Mourik, *Mol. Phys.*, 2008, **106**, 1487–1494.

97 C. Colominas, F. J. Luque and M. Orozco, *J. Phys. Chem. A*, 1999, **103**, 6200–6208.

98 C. Colominas, J. Teixidó, J. Cemeli, F. J. Luque and M. Orozco, *J. Phys. Chem. B*, 1998, **102**, 2269–2276.

