J Mol Model (2007) 13:357–365
DOI 10.1007/s00894-006-0150-y

ORIGINAL PAPER

# A hydrophobic similarity analysis of solvation effects on nucleic acid bases

**Jordi Muñoz-Muriedas · Xavier Barril · José María López · Modesto Orozco · Francisco Javier Luque**

Received: 15 January 2006 / Accepted: 27 June 2006 / Published online: 21 September 2006


**Abstract** We investigate the changes in the solvation properties of the natural nucleic acid bases due to the formation of the canonical Watson–Crick hydrogen-bonded complexes. To this end, the changes in the free energy of solvation of the bases induced upon hydrogen-bonded dimerization are analyzed by means of the hydrophobic similarity index, which relies on the atomic contributions to the free energy of solvation determined by the partitioning method implemented in the framework of the MST continuum model. Such an index is also used to examine the hydrophobic similarity between the canonical nucleic acid bases and a series of highly apolar analogues, which have been designed as potential candidates to expand the genetic alphabet. The ability of these analogues to be incorporated into modified DNA duplexes can be related to the large reduction in the hydrophilicity of the natural bases upon formation of the canonical hydrogen-bonded dimers. The results illustrate the suitability of the hydrophobic similarity index to rationalize the role played by solvation in molecular recognition.

**Keywords** Solvation · Hydrophobicity · Molecular similarity · Hydrophobic index · Continuum models

Proceedings of "Modeling Interactions in Biomolecules II", Prague, September 5th–9th, 2005.

J. Muñoz-Muriedas · J. M. López · F. J. Luque (✉)
Departament de Fisicoquímica, Facultat de Farmàcia,
Avgda. Diagonal 643,
Barcelona 08028, Spain
e-mail: fjluque@ub.edu

X. Barril
ICREA and Departament de Fisicoquímica,
Facultat de Farmàcia,
Avgda. Diagonal 643,
Barcelona 08028, Spain

M. Orozco (✉)
Departament de Bioquímica i Biología Molecular,
Facultat de Química, Universitat de Barcelona,
Martí i Franqués 1,
Barcelona 08028, Spain
e-mail: modesto@mmb.pcb.ub.es

M. Orozco
Unitat de Modelització Molecular i Bioinformàtica,
Parc Científic de Barcelona,
Josep Samitier 1-6,
Barcelona 08028, Spain

## Introduction

Modern pharmaceutical research relies heavily on screening very large numbers of compounds. This can be achieved by means of high-throughput screening assays or by the combination of computational filtering of virtual libraries with lower-throughput assays [1]. It is, therefore, essential to maintain a library of good quality (drug-like, non-reactive, etc.), but also to be able to compare molecules with each other in order to select either diverse or target-focused subsets of the library for maximum chemical space coverage or improved hit rates, respectively [2]. In this context, the concept of molecular similarity/diversity is of crucial importance [3–5].

The assumption that the structural *similarity* between two molecules implies similar biological effects resides in the grounds of structure–activity relationships. Nevertheless, there are different ways to define similarity between molecules, leading to different methodological approaches.

In 1980 Carbó and coworkers developed the concept of *molecular quantum similarity measure*, [6, 7] based on the assumption that similar molecules must have similar electron densities (see Eq. 1)

$$C_{AB} = \frac{Z_{AB}}{(Z_{AA}Z_{BB})^{1/2}} \quad (1)$$

where $Z$ involves an integral over the electron densities of the two interacting molecules, A and B (Eq. 2).

$$Z_{AB}(\Omega) = \int\int \rho_A(r_1)\Omega(r_1, r_2)\rho_B(r_2)dr_1dr_2 \quad (2)$$

where $\rho_A(r_1)$ and $\rho_B(r_2)$ are the first-order density functions associated to molecules A and B, respectively, and $\Omega$ is a positive definite operator that defines the similarity measure (typically, overlap-like and Coulomb-like measures have been used).

The calculation of Carbó-like indices faces some problems, such as the expensiveness due to the use of *ab initio* wavefunctions of reasonable quality and the bias effect due to the core electron density on the similarity measure. These difficulties can be alleviated using several strategies, such as the use of valence electron densities, [8] the addition of nuclear charges to screen the core electronic charge, [9] or to restrict the similarity analysis to electron-density properties evaluated at bond critical points [10, 11].

Other similarity schemes, specially those working in the framework of classical mechanics, use different properties than density as molecular descriptors. One approach is to resort to topological descriptors based on connectivity indices between atoms, [12, 13] or electrostatic potentials and fields [14–23]. Finally, another strategy consists of the simultaneous comparison of electrostatic and steric properties of molecules [24–28]. Among this latter approach, the so-called *Comparative Molecular Field Analysis* [29–34] has become especially popular. Inclusion of hydrophobic properties, nevertheless, is probably the most challenging issue in molecular similarity measures.

Hydrophobic/hydrophilic regions in a molecule are often identified based on intuitive, but somewhat arbitrary, chemical concepts, like polar versus apolar accessible surfaces [35–38] or from atom-based parameters, like atomic partial charges [39]. Hydrophobic regions are also identified on the basis of energetic criteria derived from the interaction of the molecule with suitable probes [40, 41]. The hydrophobicity can also be characterized from the concept of "hydrophobic moment", which was introduced to examine the stability of protein structures in water [42, 43]. Recently, such a concept has been generalized to the definition of both "solvation dipole" and "solvent-transfer dipole". These terms, which have proved to be valuable for comparison of partition-related properties in series of structurally related compounds, [44, 45] reflect the distribution of atomic contributions to the solvation free energy and transfer free energy between water and an organic solvent (i.e., octanol) along the molecule. An alternative approach is the use of "lipophilic" potentials [46–50] computed by using expressions similar to that given in Eq. (3), which provides an easy tool to visualize the lipophilicity of molecules [51–53]. A heuristic lipophilicity potential that exploits the electrostatic potential created by the molecule has also been defined recently [54, 55]. By paralleling work on the calculation of MEP similarity indices, a similarity index can be defined by comparison of the lipophilic potential values computed in grids of points around the molecule [56].

$$LP(r) = \sum_i f_i g(r - r_i) \quad (3)$$

where $f_i$ is the hydrophobic contribution of fragment $i$ in the molecule to the partition coefficient, and $g(r-r_i)$ is a distance-dependent function.

We have recently proposed a hydrophobic similarity index [57] based on a rigorous partitioning scheme [58] of the solvation or solvent-transfer (referred simply as transfer from here onwards) free energy into fragment contributions within the framework of Miertus–Scrocco–Tomasi (MST) continuum model [59]. The partitioning scheme divides the solvation/transfer free energy into contributions assigned to the surface elements that define the solute/solvent interface, which can be subsequently integrated to derive atomic or group contributions. Here we use the hydrophobic similarity index to study the changes in solvation properties of nucleic acid bases induced by the formation of hydrogen-bonded complexes. This analysis provides a qualitative basis to explain the stability of duplexes where canonical bases have been replaced by apolar compounds.

## Methods

In this section we briefly describe the main features of the MST continuum model, [59, 60] which is based on the Polarizable Continuum Model, [61] and the partitioning scheme used to decompose the solvation free energy into atomic contributions. Finally, the hydrophobic similarity index defined from those atomic contributions is presented.

*The MST continuum method* In the MST method the solvation free energy, $\Delta G_{sol}$ , is expressed by the addition of three contributions: cavitation, van der Waals, and electrostatic. The cavitation free energy, $\Delta G_{cav}$, is determined following Pierotti's scaled particle theory [62] adapted to molecular shaped cavities by using the procedure proposed by Claverie [63]. Thus, the cavitation free energy of atom $i$, $\Delta G_{C-P,i}$, is determined weighting the

contribution of the isolated atom, $\Delta G_{P,i}$, by the ratio between the solvent-exposed surface of such an atom, $S_i$, and the total surface of the molecule, $S_T$ (Eq. 4).

$$\Delta G_{cav} = \sum_{i=1}^{N} \Delta G_{C-P,i} = \sum_{i=1}^{N} \frac{S_i}{S_T} \Delta G_{P,i} \tag{4}$$

where $N$ is the number of atoms.

The van der Waals term, $\Delta G_{vW}$, is computed using a linear relationship to the solvent-exposed surface of each atom, as noted in Eq. (5), where $\Delta G_{vW,i}$ is the van der Waals free energy of atom $i$, and $\xi_i$ is the atomic surface tension, which is determined by fitting to the experimental free energy of solvation.

$$\Delta G_{vW} = \sum_{i=1}^{N} \Delta G_{vW,i} = \sum_{i=1}^{N} \xi_i S_i \tag{5}$$

Finally, the electrostatic contribution, $\Delta G_{\text{ele}}$, is determined assuming that the solvent is a continuum polarizable medium, which reacts against the solute charge distribution. The solvent's reaction field is introduced into the Schrödinger equation by means of a perturbation operator, $\widehat{V}_R$, which consists of a set of imaginary charges located on the solute cavity (Eq. 6).

$$\hat{V}_R = \sum_{j=1}^{M} \frac{q_j}{|r_j - r|} \tag{6}$$

where M is the total number of surface elements, $j$, in which the solute cavity is divided and $\{q_j\}$ denotes the set of charges (located at $r_j$) that represents the solvent response.

*Partitioning of the free energy of solvation* Partitioning of the non-electrostatic terms into atomic contributions is straightforward, since they are related to the solvent-accessible surface of atoms (see Eqs. 4 and 5). With regard to the electrostatic term, such a partitioning is facilitated by using a perturbation treatment of the mutual polarization between solute and solvent, [64] which permits to write the electrostatic component of the solvation free energy as

$$\Delta G_{ele} = \langle \Psi^o \left| \frac{1}{2} V_R^{sol} \right| \Psi^o > \tag{7}$$

where the index "sol" means that the perturbation operator is adapted to the fully relaxed charge distribution of the solute in solution, and the index "$o$" stands for the gas phase environment.

Equation 7 permits us to decompose $\Delta G_{\text{ele}}$ into atomic contributions, as noted in Eq. (8). According to this partitioning scheme, which is denoted *surface-based* partitioning method, the fractional contribution to $\Delta G_{\text{ele}}$ of a given atom $i$ is determined from the interaction energy between the whole charge distribution of the molecule with the apparent solvent-induced charges located at the surface elements pertaining to the portion of the cavity generated from that atom.

$$\Delta G_{ele} = \sum_{i=1}^{N} \Delta G_{ele,i} = \sum_{i=1}^{N} \frac{1}{2} \sum_{\substack{j=1 \\ j \in i}}^{M} \langle \Psi^o | \frac{q_j}{|r_j - r|} | \Psi^o \rangle \tag{8}$$

The free energy of solvation in a given solvent can then be expressed in terms of atomic contributions, $\Delta G_{\text{sol},i}$, by adding the corresponding contributions to the electrostatic, cavitation and van der Waals terms (Eq. 9)

$$\begin{aligned} \Delta G_{sol} &= \sum_{i=1}^{N} \Delta G_{sol,i} \\ &= \sum_{i=1}^{N} \left( \Delta G_{ele,i} + \Delta G_{C-P,i} + \Delta G_{vW,i} \right) \end{aligned} \tag{9}$$

Finally, the fractional contribution to the transfer free energy between two solvents (typically, water and an organic solvent) can be obtained from the fractional contributions to the free energy of solvation of each atom in the two solvents (see Eq. 10), where X stands for each of the three contributions to the free energy of solvation in the MST model).

$$\begin{aligned} \Delta G_{w \to o} &= \sum_{i=1}^{N} \Delta G_{w \to o,i} \\ &= \sum_{i=1}^{N} \left( \Delta\Delta G_{ele,i} + \Delta\Delta G_{C-P,i} + \Delta\Delta G_{vW,i} \right) \end{aligned} \tag{10}$$

where $\Delta\Delta G_{X,i} = \Delta G_{X,i}(organic) - \Delta G_{X,i}(water)$

*Hydrophobic similarity index* The partitioning of the solvation/transfer free energy in atomic contributions permits us to define a simple index to compare the similarity in hydrophobic/hydrophilic properties of molecules [57]. Given a pair of molecules, A and B, the hydrophobic similarity measure can be determined from

$$\Lambda_{AB}(R_{AB}) = \sum_{i \in A} \sum_{j \in B} - \frac{\Delta G_i \Delta G_j}{\left( r_{ij}^{n} + d_o \right)} \tag{11}$$

where $\Delta G_i$ denotes the contribution of atom $i$ in molecule A to the solvation/transfer free energy, $r_{ij}$ is the interatomic distance, $n$ is a parameter used to control the shape of the similarity function, and $d_o$ is a reference distance that avoids the occurrence of singularities in $\Lambda_{AB}$ ($n$=1 and $d_o$=2 in present calculations).

Since the similarity measure depends on the relative orientation of the two molecules, a Monte-Carlo algorithm combined with a multiple-copy strategy is used to maximize the overlap between hydrophobic (and hydrophilic) regions in the two molecules that are being compared. The best fitted structures are finally refined using a minimization algorithm to optimize the hydrophobic similarity measure, $\Lambda_{AB}$.

Finally, the hydrophobic similarity index, $\gamma_{AB}$, can be defined by using a Carbó-like expression, as noted in Eq. (12).

$$\gamma_{AB} = \frac{\Lambda_{AB}}{(\Lambda_{AA}\Lambda_{BB})^{1/2}} \tag{12}$$

where $\Lambda_{AA}$ is the hydrophobic self-similarity of molecule A.

*Computational details* Calculations were performed for the separate and hydrogen-bonded bases in water and octanol using the standard parameters of the HF/6-31G(d) version of the MST model [59, 65]. According to the MST HF/6-31G(d) parameterization, the geometries were optimized in the gas phase and kept frozen for the calculations in solution. MST calculations were carried out using a local version of the MonsterGauss [66] computer program.

## Results and discussion

*Solvation properties of separated and hydrogen-bonded natural bases* The interactions between nucleic acids are fundamental for the maintenance of the genetic information, which is reflected in the specific pairings between adenine (A):thymine (T) and guanine(G):cytosine(C) in the DNA duplex. This specificity is determined by the arrangement of hydrogen-bond donors and acceptors in the natural bases, leading to the Watson–Crick hydrogen bonded A:T and G:C dimers. The canonical G:C and A:T pairings in B-DNA are intrinsically stable, with interaction energies close to −26 and −12 kcal mol$^{-1}$ as determined from different studies (see, for instance, References [67, 68]).

Table 1 reports selected topological properties of A:T and G:C hydrogen-bonded complexes (Fig. 1). Those properties include the electron density ($\rho$) at the bond critical point of the hydrogen bonds formed upon complex formation following Bader's theory of atoms in molecules, [69] and the core-valence bifurcation index ($\nu$) defined in the framework of the electron localization function. [70] Both parameters have been shown to provide useful information on the strength of the hydrogen bond [71–73]. The results in Table 1 indicate that those topological parameters are mostly unaffected upon solvation of the dimer, as expected from the fact that the polar groups involved in hydrogen-bonding are very little exposed to the solvent. Accordingly, no major changes in the intrinsic strength of the hydrogen-bond interaction are expected to occur upon solvation of the dimer.

**Table 1** Topological properties for the Watson–Crick hydrogen-bonded G:C and A:T pairs: electron density ($\rho$) at the bond and ring critical points and core-valence bifurcation index ($\nu$) for the Watson–Crick hydrogen-bonded base pairs determined both in the gas phase and in aqueous solution

| Dimer | $\rho$ (gas) | $\nu$ (gas) | $\rho$ (water) | $\nu$ (water) |
|---|---|---|---|---|
| Guanine:Cytosine | | | | |
| N2–H2···O2 | 0.0228 | −0.061 | 0.0233 | −0.059 |
| ring | 0.0040 | – | 0.0041 | – |
| N1–H1···N3 | 0.0317 | −0.011 | 0.0322 | −0.007 |
| ring | 0.0052 | – | 0.0052 | – |
| O6···H4–N4 | 0.0425 | −0.006 | 0.0427 | −0.005 |
| Adenine:Thymine | | | | |
| N1···H3–N3 | 0.0291 | −0.020 | 0.0293 | −0.018 |
| ring | 0.0045 | – | 0.0045 | – |
| N6–H6···O2 | 0.0327 | −0.041 | 0.0328 | −0.038 |

The electron density is in atomic units.

Although the dimer is mostly unchanged, its formation in aqueous solution is energetically largely unfavorable owing to the loss of favorable interactions of the polar groups that mediate hydrogen bonds in the base pair with water molecules [74]. This effect can be determined from the differential free energy of solvation, $\Delta\Delta G_{sol}$, of the nucleobases in the hydrogen-bonded dimer (X:Y) relative to the

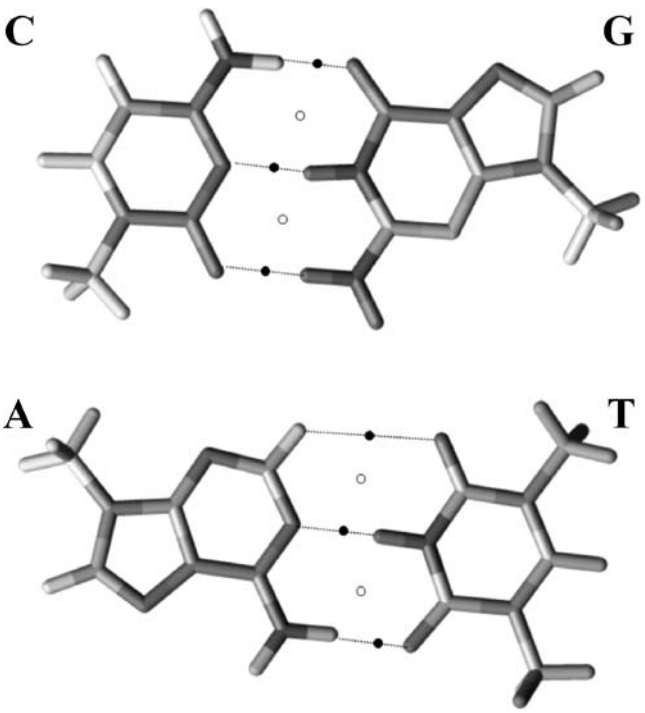



**Fig. 1** Bond (*black*) and ring (*white*) critical points in G:C and A:T Watson–Crick dimers

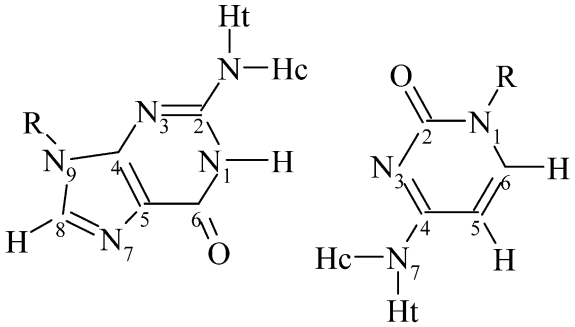
separate bases (X, Y) (see Eq. (13), where X:Y denotes A:T and G:C). The hydration free energies estimated from MST calculations amount to −23.0, −18.6, −14.0 and −11.2 kcal mol$^{-1}$ for methylated G, C, A and T, respectively, whereas the hydration free energies of Watson–Crick G:C and A:T dimers are −20.6 and −15.3 kcal mol$^{-1}$. Therefore, the differential hydration free energies are 21.0 and 9.9 kcal mol$^{-1}$ for G:C and A:T, indicating the strong destabilizing influence played by hydration on the hydrogen-bonded pairing of nucleobases.

$$\Delta\Delta G_{sol} = \Delta G_{sol}(X:Y) - \Delta G_{sol}(X) - \Delta G_{sol}(Y) \qquad (13)$$

The fractional analysis allows us to determine the magnitude of the water-induced destabilization for each base in the dimer, as shown in Tables 2 and 3. For G:C, hydration destabilizes G and C by 9.1 and 11.9 kcal mol$^{-1}$, respectively. Such destabilization amounts to 4.3 and 5.6 kcal mol$^{-1}$ for A and T in A:T, respectively. Thus, the solvent-induced destabilization of the hydrogen-bonded dimers is larger for the most polar bases. Inspection of the fractional contributions shown in Fig. 2 also shows that such destabilization is mostly due to the groups directly involved in hydrogen-bonding. This can be easily understood considering that the strongest binding in the gas phase is obtained for the dimer with the larger dipole annihilation, which in turn makes it to be less well solvated in water.

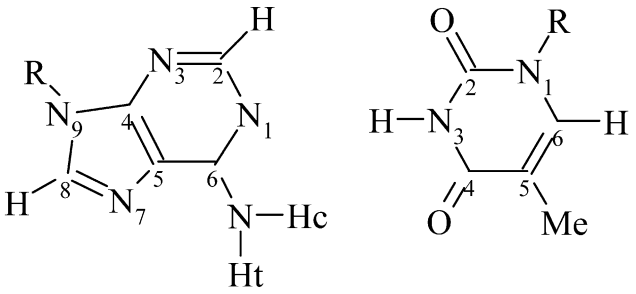
In order to examine the change in the hydrophobic properties, the free energy of solvation in octanol of the hydrogen-bonded dimers and the separate bases was also determined from MST calculations. The solvation free energies in octanol are −21.2, −15.7, −14.6 and −11.2 kcal mol$^{-1}$ for G, C, A and T, respectively, whereas the values for G:C and A:T dimers are −25.4 and −20.7 kcal mol$^{-1}$, respectively. Accordingly, the differential solvation free energies (see Eq. (13)) are 11.5 and 5.1 kcal mol$^{-1}$ for G:C and A:T, which are ~50% of the values obtained for the solvation in water. It is then clear that octanol disfavors hydrogen bonding of the nucleobases, but at much reduced extent than water. On the basis of the results obtained from the fractional analysis for the solvation in octanol (see Tables 2 and 3), the solvent-induced destabilization of the hydrogen-bonded bases

**Table 2** Fractional contributions to the solvation free energy in octanol and water (kcal mol$^{-1}$) and to the octanol/water partition coefficient (in log P units) of guanine–cytosine pair. The difference with regard to the contribution of the separated bases is given in parenthesis




| Guanine | | | | Cytosine | | | |
|---|---|---|---|---|---|---|---|
| **Atom** | **Octanol** | **Water** | **logP** | **Atom** | **Octanol** | **Water** | **logP** |
| **N1** | -1.0 (-0.7) | 0.0 (+0.7) | 0.8 (+0.1) | **N1** | -0.5 (0.0) | 0.0 (0.0) | 0.4 (0.0) |
| **C2** | -0.6 (+0.1) | -0.2 (+0.3) | 0.3 (+0.1) | **C2** | -0.8 (0.2) | -0.5 (-0.1) | 0.2 (-0.1) |
| **N3** | -2.0 (-0.6) | -3.0 (-1.0) | 0.8 (-0.7) | **N3** | -1.1 (+2.1) | -0.5 (+4.2) | 0.4 (+1.6) |
| **C4** | -0.7 (0.0) | -0.2 (+0.2) | 0.4 (+0.1) | **C4** | -0.7 (0.0) | -0.4 (-0.1) | 0.2 (-0.1) |
| **C5** | -0.9 (-0.1) | -0.4 (-0.1) | 0.4 (+0.1) | **C5** | -0.3 (0.0) | 0.0 (0.0) | 0.2 (0.0) |
| **C6** | -0.8 (0.0) | -0.1 (+0.1) | 0.5 (+0.1) | **C6** | -0.4 (0.0) | -0.4 (-0.1) | 0.0 (0.0) |
| **N7** | -4.0 (-0.3) | -5.8 (+0.6) | -1.3 (-0.2) | **N(C4)** | -2.0 (+0.6) | -0.9 (+0.6) | 0.9 (-0.1) |
| **C8** | -0.3 (0.0) | 0.1 (+0.1) | 0.3 (+0.1) | **O(C2)** | -0.8 (3.6) | -1.3 (+6.5) | -0.4 (+2.1) |
| **N9** | -0.5 (0.0) | -0.2 (+0.1) | 0.4 (0.0) | **$H_c(NH_2)$** | -0.1 (+0.5) | 0.1 (+1.0) | 0.1 (+0.2) |
| **N(C2)** | -2.0 (+1.5) | -0.8 (+1.1) | 0.9 (-0.3) | **$H_t(NH_2)$** | -0.5 (+0.1) | -1.0 (+0.3) | -0.4 (+0.1) |
| **O(C6)** | -1.9 (+1.6) | -3.3 (+3.1) | -0.9 (+1.0) | **H(C5)** | -0.6 (0.0) | -0.4 (0.0) | 0.1 (0.0) |
| **$H_t(NH_2)$** | -0.4 (+0.2) | -0.7 (+0.6) | -0.3 (+0.3) | **H(C6)** | -0.9 (-0.4) | -1.0 (-0.1) | -0.1 (0.0) |
| **$H_c(NH_2)$** | -0.1(+0.7) | 0.1 (+1.8) | 0.1 (+0.5) | | | | |
| **H(N1)** | -0.1 (+0.8) | 0.1 (+2.0) | 0.1 (0.7) | | | | |
| **H(C8)** | 0.2 (+0.6) | 0.0 (0.0) | 0.1 (0.0) | | | | |

**Table 3** Fractional contributions to the solvation free energy in octanol and water (kcal $mol^{-1}$) and to the octanol/water partition coefficient (in log P units) of adenine–thymine pair. The difference with regard to the contribution of the separated bases is given in parenthesis



| Adenine | | | | Thymine | | | |
|---|---|---|---|---|---|---|---|
| **Atom** | **Octanol** | **Water** | **logP** | **Atom** | **Octanol** | **Water** | **logP** |
| **N1** | -1.1 (+1.0) | -0.5 (+2.9) | 0.4 (+1.4) | **N1** | -0.5 (0.0) | 0.0 (+0.0) | 0.4 (0.0) |
| **C2** | -0.2 (0.0) | 0.2 (+0.1) | 0.3 (+0.1) | **C2** | -0.7 (0.0) | -0.5 (+0.3) | 0.3 (+0.1) |
| **N3** | -2.3 (-0.3) | -3.4 (-0.4) | -0.8 (-0.1) | **N3** | -1.1 (+0.3) | 0.0 (+0.5) | 0.8 (+0.1) |
| **C4** | -0.7 (0.0) | -0.3 (-0.3) | 0.3 (0.0) | **C4** | -0.8 (0.0) | -0.5 (+0.2) | 0.3 (+0.1) |
| **C5** | -0.8 (0.0) | -0.3 (0.0) | -0.4 (+0.1) | **C5** | -0.8 (0.0) | -0.4 (+0.0) | 0.3 (+0.1) |
| **C6** | -0.7 (+0.5) | -0.2 (+0.2) | 0.4 (+0.1) | **C6** | -0.3 (0.0) | -0.3 (+0.1) | 0.1 (0.0) |
| **N7** | -2.5 (-0.3) | -3.7 (-0.5) | -0.9 (-0.2) | **O(C4)** | -0.8 (+1.3) | -1.3 (+2.3) | -0.4 (+0.7) |
| **C8** | -0.3 (0.0) | 0.0 (+0.1) | 0.2 (0.1) | **O(C2)** | -1.2 (+0.4) | -2.3 (+0.4) | -0.8 (0.0) |
| **N9** | -0.5 (0.0) | 0.0 (+0.0) | 0.4 (0.0) | **H(N3)** | 0.0 (+0.9) | 0.1 (+1.6) | 0.1 (+0.6) |
| **N(C6)** | -2.0 (+0.4) | -1.1 (+0.4) | 0.7 (0.4) | **H(C6)** | -0.6 (0.0) | -0.6 (+0.1) | 0.0 (+0.1) |
| **$H_t(NH_2)$** | -0.3 (+0.1) | -0.5 (+0.3) | -0.2 (+0.1) | **Me** | -0.5 (+0.1) | 0.7 (0.0) | 0.6 (0.0) |
| **$H_c(NH_2)$** | -0.1 (+0.4) | 0.1 (+1.2) | 0.1 (0.7) | | | | |
| **H(C2)** | 0.0 (0.0) | 0.4 (+0.2) | 0.3 (+0.1) | | | | |
| **H(C8)** | -0.4 (0.0) | -0.3 (+0.1) | 0.1 (0.0) | | | | |

relative to the separate bases amounts to 5.5, 6.0, 2.1 and 3.0 for G, C, A and T, respectively.

The octanol/water partition coefficients (log P) of isolated G and C are negative, while those of the most hydrophobic nucleobases A and T are positive (Table 4). Interestingly, the octanol/water partition coefficients of the G:C and A:T pairs are 3.5 and 4.0, thus reflecting the increased hydrophobicity of the bases resulting upon formation of the dimer. Such an increase in the hydrophobicity can be decomposed in contributions from the nucleobases by using the atomic fractional analysis (i.e. adding up the atomic contribution to log P of each atom in a base), as it is shown in Table 4, which also reports the partition coefficients of the each base separated and in the hydrogen-bonded dimer. The log P value of G, C, A and T in the dimer is estimated to be +1.3, +2.2, +2.1 and +1.9. Therefore, relative to the separated bases, the hydrophobicity of G, C, A and T is enlarged by 3.6, 4.4, 2.5 and 1.9 log P units upon formation of the hydrogen-bonded complex.

The largest contributions to the partition coefficient of the G:C and A:T pairs (see Tables 2 and 3) correspond to the groups directly involved in hydrogen-bond interactions. In the G:C pair, the contributions of the $NH_2$, NH and carbonylic O atoms in G account for an increase of 2.4 log P units in the hydrophobicity of G, whereas the $NH_2$, N

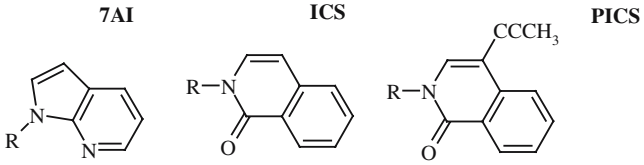



**Fig. 2** Schematic representation of the 7-azaindole (7AI), isocarbostyril (ICS) and its propynyl-substituted derivative (PICS) as apolar bases that can be incorporated in DNA duplexes

**Table 4** Octanol/water partition coefficients (in log P units) of the nucleic acid bases separated and in the Watson–Crick hydrogen-bonded dimer

| Complex | Separated | Dimer |
|---|---|---|
| G:C | | +3.5 |
| G | −2.3 | +1.3 |
| C | −2.2 | +2.2 |
| A:T | | +4.0 |
| A | +0.4[a] | +2.1 |
| T | +0.0 | +1.9 |

[a] The experimental octanol/water partition coefficient for methyladenine is 0.0 [77].

and carbonylic O atoms in C contribute to the enlarged lipophilicity by 4.1 log P units. In A:T, the contributions of the $NH_2$ and N atoms in A account for an increase in the lipophilicity of the base of 2.0 log P units, and those of NH and carbonylic O atoms in T are enlarged by 1.5 log P units.

*Hydrophobic similarity analysis of natural bases and apolar analogues* The change in the hydrophobicity of the nucleobases originated upon formation of the Watson–Crick dimers can be quantified by means of the hydrophobic similarity index defined in Eq. (12), which can be evaluated using the atomic fractional contributions to the octanol/water partition coefficient of the bases separated and in the dimers. Inspection of the results shown in Table 5 reveals the large change in the hydrophobic properties of the bases upon complex formation. The largest similarity indexes (between 0.31 and 0.51) are found for the comparison of A and T, which are the less polar nucleobases. In turn, comparison of G and C gives rise to similarity indexes lower than 0.2, thus revealing the marked effect that desolvation exerts in these bases upon formation of the complex.

Recently, several studies [75, 76] have reported that Watson–Crick dimers in DNA duplexes can be replaced by apolar bases without altering dramatically the stability of the modified DNA duplex. This is the case, for instance, of the apolar bases 7-azaindole (7AI), isocarbostyril (ICS) and its propynyl-substituted derivative (PICS) shown in Fig. 2, since the melting temperatures of the 5′-dGCGTACX CATGCG (3′-dCGCATGYGTACGC) duplex (where X, Y denote canonical and apolar bases) are fully comparable to the values of the unmodified DNAs, as can be stated from inspection of the data given in Table 6.

Table 7 reports the similarity indices between canonical (both separated and in the hydrogen-bonded complexes) and apolar bases determined from the atomic contributions to the octanol/water partition coefficient. To this end, the alignment between natural and unnatural bases started from a configuration where a common orientation for the linkage between the base and the ribose unit was maintained in order to take into account the orientational effects imposed by the DNA structure. Compared to the separated bases, the results reveal that the apolar bases are more similar to A and T (similarity indexes ranging from 0.26 to 0.38) than to G and C (from 0.14 to 0.31), as expected from the lower polarity of A and T. Nevertheless, when comparison is made with the canonical bases in the hydrogen-bonded dimers, there is a marked increase in the similarity index, which ranges from 0.53 to 0.75. Therefore, the similarity indices vary by around 0.4 when apolar bases are compared with regard to either separated or hydrogen-bonded bases, thus showing a notable sensitivity to capture the increase in hydrophobicity of the hydrogen-bonded pair relative to the separated bases (see above). This finding allows to realize the unexpected finding that apolar bases, though having a completely different chemical structure compared to the separate natural bases, can be incorporated into the DNA leading to stable duplexes.

**Table 5** Hydrophobic similarity index determined from atomic fractional contributions to the octanol/water partition coefficient for the nucleic acid bases separated and in the Watson–Crick hydrogen-bonded dimers

| | A | T | G | C |
|---|---|---|---|---|
| A* | **0.51** | 0.31 | 0.09 | 0.08 |
| T* | 0.48 | **0.33** | 0.12 | 0.04 |
| G* | 0.21 | 0.19 | **0.17** | 0.08 |
| C* | 0.43 | 0.27 | 0.04 | **0.04** |

The bases in the dimers are denoted by an asterisk.

**Table 6** Melting temperatures ($T_m$) for canonical and modified DNA duplexes

| 5′-dGCGTAC**X**CATGCG | | |
|---|---|---|
| 3′-dCGCATG**Y**GTACGC | | |
| X[a] | Y | $T_m$ |
| A | T | 59.2 |
| G | C | 61.8 |
| 7AI | 7AI | 55.5 |
| ICS | ICS | 59.3 |
| PICS | PICS | 62.6 |

[a] see Fig. 2 for the representation of the unnatural apolar bases.

## Conclusions

The analysis of the solvation properties of both separate and hydrogen-bonded bases reflects the enhanced hydrophobicity of the natural bases upon formation of the hydrogen-bonded dimer. This finding provides a basis to rationalize the experimental evidences that apolar analogues of the natural nucleic acid bases can be incorporated into the DNA duplex, suggesting that hydrophobicity is an important driving force for the stability and selective pairing of bases in the DNA duplex. The partitioning scheme adopted here

**Table 7** Hydrophobic similarity indexes determined from atomic fractional contributions to the octanol/water partition coefficient between apolar and canonical bases either separate or in the G:C and A:T dimers

| | A | T | G | C | A* | T* | G* | C* |
|---|---|---|---|---|---|---|---|---|
| 7AI | 0.46 | 0.38 | 0.17 | 0.15 | 0.81 | 0.83 | 0.45 | 0.77 |
| ICS | 0.38 | 0.38 | 0.31 | 0.27 | 0.74 | 0.75 | 0.53 | 0.69 |
| PICS | 0.30 | 0.26 | 0.21 | 0.14 | 0.65 | 0.68 | 0.60 | 0.61 |

The bases in the dimers are denoted by an asterisk.

allows us to quantify the magnitude of the changes in the solvation properties arising from the formation of the Watson–Crick dimers. Moreover, the hydrophobic similarity index provides a simple way to quantify the global similarity in the hydrophobicity/hydrophilicity of both natural and apolar bases. We think that these results illustrate the potential impact of this simple parameter as a tool to gain insight into the determinants that modulate the bioactivity of molecules and for the design of new compounds in drug discovery.

**Acknowledgments** We thank the Ministerio de Ciencia y Tecnología (grants CTQ2005-09365 and BIO2003-06848) for financial assistance and the Centre de Supercomputació de Catalunya for computational facilities.

## References

1. Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003) Nat Rev Drug Discov 52:369–378
2. Schuffenhauer A, Popov M, Schopfer U, Acklin P, Stanek J, Jacoby E (2004) Comb Chem High Throughput Screen 7:771–781
3. Dean PM (ed) (1995) Molecular similarity in drug design. Blackie Academic, London
4. Johnson MA, Maggiora GM (eds) (1990) Concepts and applications of molecular similarity. Wiley, New York
5. Carbó-Dorca R, Gironés X, Mezey PG (eds) (2001) Fundamentals of molecular similarity. Kluwer Academic/Plenum Publishers, New York
6. Carbó R, Leyda L, Arnau M (1980) Int J Quant Chem 17:1185–1189
7. Carbó-Dorca R, Robert D, Amat L, Girones X, Besalú E (2000) Lecture notes in chemistry, vol 73. Springer, Berlin, Hiedelberg, New York
8. Lee C, Smithline S (1994) J Phys Chem 98:1135–1138
9. Bowen-Jenkins PE, Richards WG (1986) J Chem Soc, Chem Commun 133–135
10. Popelier PLA (1999) J Phys Chem A 103:2883–2890
11. O'Brien SE, Popelier PLA (1999) Can J Chem 77:28–36
12. Kier LB, Hall LH (1986) Molecular connectivity letchworkin structure-activity analysis. Research Studies, Letchwork
13. Hall LH, Kier LB (1991) The molecular connectivity chi indexes and kappa shape indexes in structure-property relations. In: Lipkowitz KB, Boyd DB (eds) Reviews in Computational Chemistry, vol 2. VCH, New York pp 367–442
14. Luque FJ, Sanz F, Illas F, Pouplana R, Smeyers YG (1988) Eur J Med Chem 23:7–10
15. Richard AM (1991) J Comput Chem 12:959–969
16. Hernández B, Orozco M, Luque FJ (1996) J Comput Aided Mol Des 10:535–544
17. Náray-Szabó G, Ferenczy GG (1995) Chem Rev 95:829–847
18. Hodgkin EE, Richards WG (1987) Int J Quant Chem 14:105–110
19. Burt C, Richards WG, Huxley P (1990) J Comput Chem 11: 1139–1146
20. Manaut F, Sanz F, José J, Milesi M (1991) J Comput Aided Mol Des 5:371–380
21. Petke JD (1993) J Comput Chem 8:928–933
22. Rodríguez J, Manaut F, Sanz F (1993) J Comput Chem 14:922–927
23. Thorner DA, Willet P, Wright PM, Taylor R (1997) J Comput Aided Mol Des 11:163–174
24. Meyer AM, Richards WG (1991) J Comput Aided Mol Des 5:427–439
25. Mezey PG (1993) Shape in chemistry: an introduction to molecular shape and topology. VCH, New York
26. Tokarski JS, Hopfinger AJ (1994) J Med Chem 37:3639–3654
27. Jain N, Dietterich TG, Lathrop RH, Chapman D, Critchlow Jr RE, Bauer BE, Webster TA, Lozano-Pérez T (1994) J Comput Aided Mol Des 8:635–652
28. Cramer RD III, Paterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
29. Klebe G, Mietzner T, Weber F (1994) J Comput Aided Mol Des 8:751–778
30. Perkins TDJ, Mills JEJ, Dean PM (1995) J Comput Aided Mol Des 9:479–490
31. Mestres J, Rohrer DC, Maggiora GM (1997) J Comput Chem 18:934–954
32. Kubinyi H (ed) 3D QSAR in drug design: theory, methods and applications. (1993) ESCOM, Leiden
33. Lemmen C, Lengauer T, Klebe G (1998) J Med Chem 41:4502–4520
34. Miller MD, Sheridan RP, Kearsley K (1999) J Med Chem 42:1505–1514
35. Palm K, Luthman K, Ungell AL, Strandlund G, Artursson P (1996) J Pharm Sci 85:32–39
36. Palm K, Luthman K, Ungell AL, Strandlund G, Beigi F, Lundahl P, Artursson P (1998) J Med Chem 41:5382–5392
37. Clark DE (1999) J Pharm Sci 88:807–814
38. Clark DE (1999) J Pharm Sci 88:815–821
39. Kantola A, Villar HO, Loew GH (1991) J Comput Chem 12:681–689
40. Segarra V, López M, Ryder H, Palacios JM (1999) Quant Struct-Act Relatsh 18:474–481
41. Sulea T, Purísima EO (1999) Quant Struct- Act Relatsh 18:154–158
42. Eisenberg D, McLachlan AD (1986) Nature 319:199–203
43. Eisenberg D, Schwarz E, Komaromy M, Wall R (1984) J Mol Biol 179:125–142
44. Barril X, Muñoz J, Luque FJ, Orozco M (2000) Phys Chem Chem Phys 2:4897–4905
45. Muñoz J, Barril X, Luque FJ, Gelpí JL, Orozco M (2001) Partitioning of free energies of solvation into fragment contributions: applications in drug design. In: Carbó-Dorca R, Gironés X, Mezey PG (eds) Fundamentals of molecular similarity. Kluwer Academic Plenum Publishers, New York pp 143–168
46. Audry E, Dubost JP, Colleter JC, Dallet P (1986) Eur J Med Chem 21:71–72
47. Brasseur R (1991) J Biol Chem 266:16120–16127
48. Heiden FW, Moeckel G, Brickmann J (1993) J Comput Aided Mol Des 7:503–514
49. Kellogg E, Semus SF, Abraham DJ (1991) J Comput Aided Mol Des 5:545–552
50. Gaillard P, Carrupt PA, Testa B, Boudon A (1994) J Comput Aided Mol Des 8:83–96
51. Furet P, Sele A, Cohén, NCJ (1988) J Mol Graph 6:182–189
52. Croizet F, Langlois MH, Dubost JP, Braquet P, Audry E, Dallet P, Colleter JC (1990) J Mol Graph 8:153–155
53. Fauchére JL, Quarendon P, Kaetterer L (1988) J Mol Graph 6:203–206
54. Du Q, Arteca GA (1996) J Comput Aided Mol Des 10:133–144
55. Du Q, Arteca GA, Mezey PG (1997) J Comput Aided Mol Des 11:503–516
56. Bone RGA, Villar HO (1995) J Mol Graph 13:201–208
57. Muñoz J, Hernández B, Barril X, Orozco M, Luque FJ (2002) J Comput Chem 23:554–563
58. Luque FJ, Barril X, Orozco M (1999) J Comput Aided Mol Des 13:139–152
59. Curutchet C, Orozco M, Luque FJ (2001) J Comput Chem 22:1180–1193
60. Luque FJ, Curutchet C, Muñoz-Muriedas J, Bidon-Chanal A, Morreale A, Gelpi JL, Orozco M (2003) Phys Chem Chem Phys 5:3827–3836

61. Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55:117–129
62. Pierotti RA (1976) Chem Rev 76:717–726
63. Claverie P (1978) In: Pullman B (ed) Intermolecular interactions: from diatomics to biomolecules. Wiley, Chichester
64. Luque FJ, Bofill JM, Orozco M (1995) J Chem Phys 103:10183–10191
65. Orozco M, Bachs M, Luque FJ (1995) J Comput Chem 16:563–575
66. Peterson M, Poirier, R MonsterGauss. Department of Biochemistry, Univ Toronto, Canada. Version modified by Cammi R, Tomasi J (1987) and by Curutchet C, Orozco M, Luque FJ (2004)
67. Huertas O, Orozco M, Luque FJ (2006) J Phys Chem A 110:510–518
68. Hobza P, Sponer J (1999) Chem Rev 99:3247–3276
69. Bader RFW (1991) Chem Rev 91:893–928
70. Silvi B, Savin A (1994) Nature 31:683–686
71. Alkorta I, Rozas I, Elguero J (1998) Struct Chem 9:243–248
72. Muñoz J, Sponer J, Hobza P, Orozco M, Luque FJ (2001) J Phys Chem B 105:6051–6060
73. Fuster F, Silvi B (2000) Theor Chem Acc 104:13–21
74. Orozco M, Cubero E, Barril X, Colominas C, Luque FJ (1999) Nucleic acid bases in solution. In: Leszczynski J (ed) Computational molecular biology. Theoretical computational chemistry, vol 8 Elsevier, Amsterdam, pp 119–165
75. Wu Y, Ogawa AK, Berger M, McMinn DL, Schuitz PG, Romesberg FE (2000) J Am Chem Soc 122:7621–7632
76. Loakes D (2001) Nucleic Acids Res 29:2437–2447
77. Hansch C, Leo A (1995) Exploring QSAR: hydrophobic, electronic and steric constants. American Chemical Society, Washington