Stair Motifs at Protein−DNA Interfaces: Nonadditivity of H-Bond, Stacking, and Cation−π Interactions

Christophe Biot,*,† René Wintjens,§ and Marianne Rooman†

Unité de Bioinformatique génomique et structurale, Université Libre de Bruxelles, CP 165/61, 50 avenue Roosevelt, B-1050 Bruxelles, Belgium, and Institut de Pharmacie, Campus de la Plaine, Université Libre de Bruxelles, CP206/04, boulevard du Triomphe, 1050 Bruxelles, Belgium

Received January 22, 2004; E-mail: cbiot@ulb.ac.be

At protein−DNA interfaces, two successive nucleobases (Gua, Ade, Thy, or Cyt) along the B-DNA stack and a positively charged side chain (Arg, Lys) or a side chain carrying a partially charged group (Asn, Gln) are often observed to form simultaneously three different pairwise interactions: aromatic base stacking, hydrogen bonding, and cation/amino-π. They are called stair motifs as they have a stair-like shape, with the H-bond forming the horizontal part of the stair, and the cation−π interaction, the vertical part (Figure 1a). A better appreciation of the subtle balance between these noncovalent interactions, and of their nature and electron-transfer properties, plays an important role in understanding and predicting molecular recognition and macromolecular functioning.

In a data set of 52 high-resolution (≤2.5 Å) crystal structures of protein/double-stranded DNA complexes, 77 stair motifs have previously been identified,[1] presenting favorable vacuum interaction energies computed at the second-order Møller−Plesset perturbation theory (MP2).[2] Note that this type of calculation appears to be the most appropriate, as density functional theory (DFT) methods fail completely for base stacking[3] and cation−π representation, and as the size of the stair motifs exclude nonperturbative coupled cluster (e.g., CCSD(T)) calculations.

Here we focused on the stair motif Gua∴Arg∨Gua (∴denotes cation−π and ∨ H-bond) occurring in the DNA-binding domain of Tc3 transposase from *Caenorhabditis elegans* (1TC3) between Arg C236 and the two successive Gua A7 and A8. This complex constitutes an academic case for studying the relative importance of the three simultaneous interactions and the resulting nonadditive effects induced by the different energy and entropy contributions, both in a vacuum and in different solvents.

Hartree−Fock (HF) and MP2 gas-phase interaction energies of the Gua∴Arg∨Gua motif are given in Table 1. Strikingly, the stabilization of this stair motif is strongly influenced by dispersion attraction: the electronic correlation contribution ΔE_{Cor} is important for the Arg∨Gua H-bond (6 kcal/mol) and especially for the Gua||Gua stacking (10 kcal/mol), the latter being unfavorable at the HF level. The poor ΔE_{Cor} contribution of Gua∴Arg (∼1 kcal/mol) results from the location of the guanidinium group above the extracyclic atoms rather than above the cycle center, and thus from the weak overlapping of the p-orbitals of both partners, due to steric constraints imposed by the DNA. In the protein/ligand context, where there is less steric constraint, London dispersion and electrostatic interactions have been shown to contribute equally to the stabilization of Gua∴Arg.[5] Note *en passant* that non-BSSE corrected HF interaction energies are not equivalent to BSSE corrected MP2 interaction energies, as sometimes stated (see Supporting Information [SI]).

The three-body energy term $\Delta E^{(3)}$ is almost identical at HF and MP2 levels: +4 kcal/mol (Table 1). The estimation of dispersion

Table 1. BSSE Corrected[a] HF and MP2 Interaction Energies (in kcal/mol) of the 1TC3 C236-A7-A8 Stair Motif without Geometry Optimization[b] Using the 6-31G({0.8,0.2},p) Basis Set[4]

	G∴R	G||G	R∨G	G∴R∨G	$\Delta E^{(3)c}$
HF	−12.7	+6.8	−23.7	−25.5	+4.1
MP2	−13.6	−3.0	−29.5	−41.9	+4.2

[a] Correction for the basis set superposition error (BSSE) is performed using the standard counterpoise method.[5] [b] Only intramolecular geometries of the isolated partners are optimized. All energy values are computed with the Gaussian 98 package.[6] Details are in refs 7,8. [c] The three-body term $\Delta E^{(3)}$ is: $\Delta E^{(3)} = \Delta E(G∴R∨G) - \Delta E(R∨G) - \Delta E(G||G) - \Delta E(G∴R)$.

nonadditivities indeed requires a higher level of calculation than MP2.[9,10] Also, $\Delta E^{(3)}$ is nearly independent of the BSSE correction (see SI). To check these results, we computed the $\Delta E^{(3)}$ term for the 15 Gua∴Arg∨Gua motifs previously identified.[1] All $\Delta E^{(3)}$ values were found to be in the +2 to +4 kcal/mol range.

Positivity of the $\Delta E^{(3)}$ term can be suspected to be due to the nonoptimal geometries of the crystal structures. To analyze this possibility, optimizations were performed at HF and MP2 levels, using various split valence or correlation consistent basis sets (from 6-31G(d,p) to aug-cc-pVDZ). Two constraints mimicking the effect of the protein environment were applied on the complexes. The first imposes parallelism between the planes of both Gua bases and allows optimization of the interbase distance. The second forces the guanidinium group to stay in the plane of the Gua with which it is H-bonded and permits adjustment of the distance and orientation. This approach is justified by the experimental observations that cation−π geometries[11] and stacked conformations in DNA crystals[12] are rather variable, unlike H-bond conformations.

Interaction energies of the optimized complexes are given in Table 2. The $\Delta E^{(3)}$ term is always positive, between +3 and +6 kcal/mol. The use of more extended basis sets for the optimizations at the HF level only slightly reduces the positivity of the $\Delta E^{(3)}$ term. In contrast, optimization at MP2/6-31G(d,p) level causes the largest $\Delta E^{(3)}$ value. This is to be related to the drastic changes observed in the energies and geometries of the π-electron involving interactions. Indeed, interaction energy of the cation−π is much more favorable than in the HF optimized complexes, and that of the stacked bases is much less favorable ($\Delta E_{HF} = +15$ kcal/mol instead of +5 kcal/mol). This highly unfavorable electrostatic contribution, which provokes repulsion, is due to the smaller distance between the stacked bases (3.2 Å instead of 3.6 Å in the crystal structure) and is probably caused by the BSSE uncorrected optimizations.

Thus, we may conclude that the Gua∴Arg∨Gua stair motif exhibits an anticooperative behavior in a vacuum at HF level, originating from many-body polarization effects: the charged guanidinium moiety influences the electrostatic potentials of both Gua bases in a similar way, leading to less favorable charge distributions on their aromatic cycles[13] (see Figure 1b).

† Unit de Bioinformatique Gnomique et Structurale.
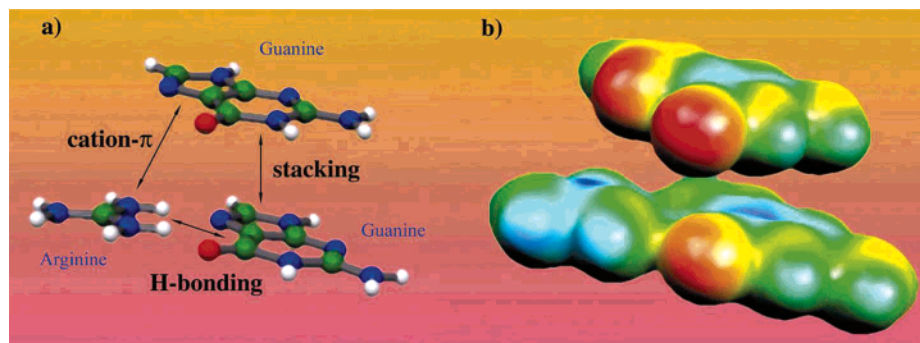§ Institut de Pharmacie, Campus de la Plaine.

Figure 1. (a) Structure and (b) molecular electrostatic potential (MEP) of the 1TC3 C236-A7-A8 stair motif. MEPs isoenergy contours were superimposed onto the total electron density surface (0.002 e/au^3). The most negative potentials are colored in red, and the most positive are in blue.

Table 2. BSSE Corrected MP2/6-31G(0.2,p) and HF/6-31G(0.2,p) (in parentheses) Interaction Energies (in kcal/mol) of the 1TC3 C236-A7-A8 Stair Motif, Whose Geometry Is Optimized at HF or MP2 Level Using Various Basis Sets

optimization	G∴R	G‖G	R∨G	G∴R∨G	$\Delta E^{(3)}$
HF	−12.8	−1.9	−36.1	−47.2	+3.6
6-31G(d,p)	(−13.0)	(+5.2)	(−34.6)	(−38.8)	(+3.6)
HF	−11.7	−1.3	−36.2	−46.0	+3.2
6-311++G(d,p)	(−12.0)	(+4.4)	(−34.7)	(−39.1)	(+3.2)
HF	−13.2	−2.1	−36.4	−48.0	+3.7
cc-pVDZ	(−13.4)	(+5.5)	(−34.9)	(−39.1)	(+3.8)
HF	−11.7	−1.4	−36.3	−46.2	+3.2
aug-cc-pVDZ	(−12.1)	(+4.5)	(−34.8)	(−39.1)	(+3.3)
MP2	−18.7	−0.7	−36.2	−49.7	+5.9
6-31G(d,p)	(−18.2)	(+15.0)	(−34.2)	(−31.6)	(+5.8)

Table 3. BSSE Corrected MP2 Interaction Energies, and HF Thermal and Solvation Free Energies (in kcal/mol), Using the 6-31G({0.8,0.2},p) Basis Set, for the 1TC3 C236-A7-A8 Stair Motif

	G∴R	G‖G	R∨G	G∴R∨G	$\Delta X^{(3) d}$
ΔE_{MP2}	−13.6	−3.0	−29.5	−41.9	+4.1
$\Delta G_{vib}{}^a$	−1.7	−6.4	−0.8	−5.0	+3.9
$\Delta G_{rot+trans}{}^b$	+15.6	+17.0	+15.6	+35.4	−12.8
$\Delta\Delta G_{water}{}^c$	+16.6	+2.9	+32.4	+48.3	−3.6
$\Delta\Delta G_{acetone}{}^c$	+15.4	+4.0	+24.4	+36.7	−7.1
$\Delta\Delta G_{THF}{}^c$	+15.7	+5.5	+22.2	+36.2	−7.2
$\Delta\Delta G_{CCl4}{}^c$	+10.7	+7.0	+14.0	+24.8	−6.9

^{a} $\Delta G_{vib} = \Delta E^{th}{}_{vib} − T\Delta S_{vib}$. ^{b} $\Delta G_{rot+trans} = \Delta E^{th}{}_{rot+trans} − T\Delta S_{rot+trans}$. ^{c} Calculated using the integral equation formalism polarizable continuum model.[6,14] ^{d} $\Delta X^{(3)}$ is given by: $\Delta X^{(3)} = \Delta X(G∴R∨G) − \Delta X(R∨G) − \Delta X(G‖G) − \Delta X(G∴R)$, where X stands for E, G, or ΔG.

Next, we analyze the influence of thermal interaction free energy contributions on the nonadditivity of the interactions. As noted in Table 3, the vibrational three-body term $\Delta G_{vib}{}^{(3)}$ is positive (+4 kcal/mol). The anticooperativity of this free energy contribution is supported by similar calculations on two other Arg-containing stair motifs, yielding $\Delta G_{vib}{}^{(3)}$ values of +0.4 kcal/mol (in 1A1G A180-B1-B2) and +7.2 kcal/mol (in 6MHT A240-D425-D426). There are indeed less vibrational degrees of freedom in the ternary complex than in the three binary complexes taken together.

In contrast, the rotational and translational three-body term $\Delta G_{rot+trans}{}^{(3)}$ term is highly negative (−13 kcal/mol). This result is not surprising, as the loss of rotational and translational degrees of freedom upon interaction of three molecular groups is smaller when two of the groups are already in contact.

Finally, we estimated the influence of the solvent on the nonadditivity of the interactions. In water, acetone, THF, and CCl$_4$, the three-body term $\Delta\Delta G_{solv}{}^{(3)}$ is systematically negative, between −4 and −7 kcal/mol (Table 3). The negativities of these values and of those calculated on six other stair motifs indicate that the solvent plays a crucial role in inducing cooperativity. Basically, the cavitation free energy shows a cooperative behavior ($\Delta G^{(3)}{}_{cavitation}$

= −31kcal/mol in CCl$_4$) which cancels the anti-cooperativity of other contributions (e.g. $\Delta G^{(3)}{}_{electrostatic}$ = 17kcal/mol).

The error on the thermal energy and entropy contributions of the complexes, due essentially to their estimation slightly away from the energy minimum, is limited but not negligible.[9] Thus, we prefer not to add these contributions to estimate the full free energy three-body term. Nevertheless, the cooperative or anticooperative tendencies of the individual contributions appear quite well defined. They can be summarized as follows: $\Delta E_{HF}{}^{(3)} > 0$; $\Delta G_{vib}{}^{(3)} > 0$; $\Delta G_{rot+trans}{}^{(3)} < 0$; and $\Delta\Delta G_{solv}{}^{(3)} < 0$. Hence, our calculations clearly indicate that, in absence of an estimation of the dispersion nonadditivity, the cooperativity for Gua∴Arg∨Gua stair motifs arises from the environment.

Supporting Information Available: HF and MP2 interaction energies of the 1TC3 C236-A7-A8 stair motif without BSSE correction; Cartesian coordinates of the relevant optimized structures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Rooman, M.; Liévin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67.
(2) Wintjens, R.; Biot, C.; Rooman, M.; Liévin, J. *J. Phys. Chem. A* **2003**, *107*, 6249.
(3) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.
(4) Two variants of the standard basis set 6-31G(d,p), with d-polarization function exponent $\alpha_d = 0.8$ on the heavy atoms C, N, and O, were used: 6-31G(0.2,p) with $\alpha_d = 0.2$, and 6-31G({0.8,0.2},p) with $\alpha_d = 0.8$ and $\alpha_d = 0.2$. These basis sets were shown to allow a more accurate description of the three interactions considered.[2]
(5) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(6) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7, Gaussian, Inc.: Pittsburgh, PA, 1998
(7) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816.
(8) Biot, C.; Buisine, E.; Rooman, M. *J. Am. Chem. Soc.* **2003**, *125*, 13988.
(9) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
(10) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76.
(11) Lamoureux, J. S.; Maynes, J. T.; Glover, J. N. M. *J. Mol. Biol.* **2004**, *335*, 399.
(12) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
(13) Sponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 5951.
(14) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211.

JA049620G