Basis Set and Electron Correlation Effects on ab Initio Calculations of Cation-π/H-Bond Stair Motifs

René Wintjens,*,†,‡ Christophe Biot,§ Marianne Rooman,§ and Jacky Liévin||

Institut de Biologie de Lille, Institut Pasteur de Lille, UMR 8525 CNRS-Université de Lille II, 1 rue du Professeur Calmette, BP 447, 59021 Lille cédex, France, Ingénierie Biomoléculaire CP 165/64, Université Libre de Bruxelles, 50 Avenue F. D. Roosevelt, B-1050 Bruxelles, Belgium, Chimie Quantique et Physique Atomique CP 160/09, Université Libre de Bruxelles, 50 Avenue F. D. Roosevelt, B-1050 Bruxelles, Belgium

Received: January 15, 2003; In Final Form: April 18, 2003

Cation−π/H-bond stair motifs are recurrently found at the binding interface between protein and DNA. They involve two nucleobases and an amino acid side chain, and encompass three different types of interactions: nucleobase stacking, nucleobase-amino acid H-bond and nucleobase−amino acid cation−π interaction. The interaction energies of the 77 stair motif geometries identified in a data set of 52 high-resolution protein−DNA complexes were investigated by means of ab initio quantum chemistry calculations. Using the standard 6-31G* basis set, we first establish the value of the Gaussian α_d-exponent of d-polarization functions on heavy atoms, which optimizes the MP2 interaction energies. We show that, although the default value of $\alpha_d = 0.8$ is appropriate to minimize the total MP2 energy of a system, the value of $\alpha_d = 0.2$ is optimal for the three types of pairwise interactions studied and yields MP2 interaction energies quite similar to those calculated with more extended basis sets. Indeed, the more diffuse nature of the $\alpha_d = 0.2$ basis functions allows a spatial overlap between the orbitals of the interacting partners. Such functions are also shown to improve the multipole electric moments in the interaction region, which results in a stabilizing polarization effect and a better description of the dispersive energy contributions. Using the MP2 computation level and the 6-31G* basis set with $\alpha_d = 0.2$ instead of $\alpha_d = 0.8$, we computed the interaction energies of the 77 observed stair motif geometries and found that, in a vacuum, the cation−π energy is much less favorable, about 3 times, than the H-bond energy and of the same order of magnitude as the $\pi-\pi$ stacking energy. Furthermore, the convergence of the MP perturbation theory expansions was analyzed by computing the MP3 and MP4 corrections on simplified complexes. These expansions exhibited an oscillatory behavior, where MP2 seems to provide a satisfactory approximation, albeit slightly overestimated, to the interaction energy.

Introduction

Several kinds of noncovalent interactions contribute concurrently to determine macromolecular structures such as proteins and DNA. Among these, hydrogen bonds, salt bridges, van der Waals, and $\pi-\pi$ stacking interactions are known to be quite important. Hydrogen bonds are found extensively in protein structures where they are at the basis of α-helices and β-sheets.[1,2] They are also responsible for the DNA base pairing scheme. van der Waals and $\pi-\pi$ stacking interactions aid the stabilization of the protein core, and $\pi-\pi$ stacking between nucleic bases sustains the unique structure of DNA. Salt bridges have been noticed to increase the thermostability of proteins.[3−5] These different interactions are moreover all involved in biomolecular association processes, in particular in protein/protein, protein/DNA, and protein/ligand binding.

Cation−π interactions between a π electron cloud and a positive charge, as well as amino−π interactions between a π electron cloud and the partial positive charge of an amino group, also revealed to be important in macromolecular structures.[6,7] These interactions, which we both refer to as cation−π interactions, are found to be common in protein structures.[8−12] Their role in molecular recognition is extensively studied,[13−17] in particular in acetylcholine receptor,[18,19] toxin/K^+ channels,[20] protein−DNA binding,[21] antigen−antibody interaction,[22,23] and enzyme−substrate binding.[24−26] Experimental studies measured cation−π contributions to protein or peptide stability in the range of 0.4−1.0 kcal/mol, depending on the experimental conditions, on the protein analyzed and on the amino acid considered.[27−32]

Recently, a survey of X-ray structures of protein/DNA complexes has exhibited the recurrence of a particular motif, named stair motif, which involves at the same time $\pi-\pi$ nucleobase stacking, nucleobase-amino acid H-bond and nucleobase-amino acid cation−π interactions.[33] These motifs have the shape of a stair, with the H-bond forming the horizontal part of the stair and the cation−π interaction the vertical part (see Figure 1). Because they simultaneously encompass three different interactions, these motifs constitute an excellent system for studying the importance of cation−π interactions relative to other noncovalent interactions, an issue that has not yet been totally settled. They are also particularly interesting for investigating the cooperativity of noncovalent interactions. Indeed, we still do not have a good idea of the energy gain stemming

* Corresponding author. Tel: +32 2 650 52 07. Fax: +32 2 650 54 39. E-mail: Rene.Wintjens@ulb.ac.be.
† UMR 8525 CNRS-Université de Lille II.
‡ Current address: Université Libre de Bruxelles, Institut de Pharmacie, Chimie Generale, CP 206/4, Campus de la Plaine, Boulevard du triomphe, B-1050 Bruxelles, Belgium.
§ Université Libre de Bruxelles.
|| Université Libre de Bruxelles.

10.1021/jp034103q CCC: $25.00
Published on Web 07/22/2003

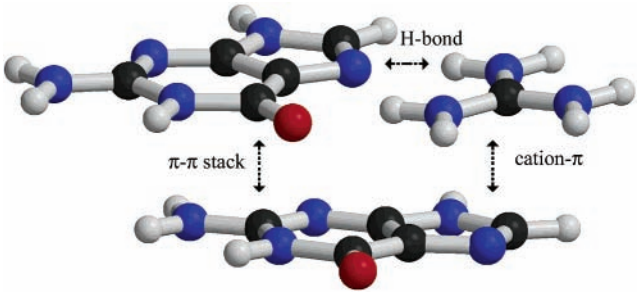

Figure 1. Cation−π/H-bond stair motif. The geometry is taken from the X-ray structure of tc3 transposase (protein code 1TC3) and the interacting residues are Arg-C236, Gua-A7, and Gua-A8. Atoms O, N, C, and H are colored in red, blue, black and gray, respectively.

from the cooperative association of several noncovalent interactions into a particular molecular scheme.

In the absence of experimental data, ab initio quantum chemistry calculations represent a useful tool to address these questions. But to accurately evaluate the bonding energies between biological moieties, ab initio calculations require taking electron correlation effects into account and the use of very large basis sets. Such levels of calculations are not always attainable considering the size of molecular systems studied. Therefore, methodological tests were first performed to determine a correct level of calculation, in particular to define a basis set adapted to our systems. We showed that an adjustment of the α_d exponents of the polarization d-functions of the medium size 6-31G* basis set allows a better estimate of the binding energy without use of massive computer time and memory. The optimal values of the α_d exponents are justified by means of an analysis of the multipole moments of the interacting partners and of the shape of the corresponding orbital. Furthermore, high-level ab initio calculations (MP3, MP4) were performed on simplified complexes to investigate the convergence of the perturbation theory expansions.

Methods

Geometric Definition of Cation-π/H-Bond Stair Motifs. A stair motif involves three interacting partners, two consecutive nucleic acid bases (Ade, Gua, Cyt, Thy) along the DNA duplex, and an amino acid side chain that carries a full positive charge (Arg, Lys) or that carries a partial positive charge $\delta(+)$ on its amino group (Asn, Gln). The two nucleic acid bases are stacked, and the amino acid side chain interacts through an H-bond with one of the bases and through a cation−π interaction with the other.[33] This motif has a stairlike shape: the H-bond constitutes the horizontal part of the stair and the cation−π the vertical part (Figure 1).

Cation−π interactions are defined geometrically by distance and angle criteria.[21] In brief, the atom carrying the net or $\delta(+)$ positive charge is required to be located inside a cylinder of 4.5 Å height, whose basis is a disk including the aromatic ring of the nucleic acid base and of radius twice that of the ring. H-bonds are assigned using the program HBPLUS.[34]

Minimal Representation of Stair Motifs. Each stair motif identified in the X-ray complexes was reduced to a minimal system, suitable for performing ab initio quantum chemistry calculations. Arg residues were represented by their guanidinium groups, DNA bases by their aromatic systems, Asn and Gln by their side chain formamide groups and Lys residues by their ammonium group. As the crystal structures sometimes display unrelaxed intramolecular geometries likely to yield distorted wave functions and wrong interaction energies, we considered, instead of the crystal structure coordinates of these minimal molecular groups, optimized coordinates obtained using the HF/6-31G** level of ab initio calculations (see below). The crystal structure coordinates of molecular groups involved in a stair motif were replaced by the optimized coordinates using the structure superposition algorithm U3BEST.[35]

This procedure allows us to unequivocally position the H-atoms, which are not observed in the crystal structures, except for Lys. In this case, one of the H atoms of the NH_4^+ ammonium group is positioned along the Nζ−Cϵ axis, but there is a an indeterminacy for the three others due to the rotational symmetry. Accordingly, we considered two different geometries. In the first, one of the three remaining H-atoms is positioned as close as possible to the center of the aromatic ring, considering the constraint induced by the positioning of the first H-atom. In the second geometry, one of the three remaining H-atoms is positioned as far as possible from the center of the aromatic ring. The latter two H-atoms are then unambiguously fixed.

Interaction Energies from ab Initio Quantum Chemistry Calculations. The pairwise interaction energy ΔE(A−B) between two molecules A and B is estimated as the difference between the energy of the complex E(A−B) and the energies of isolated partners: ΔE(A−B) = E(A−B) − E(A) − E(B). Similarly, the total interaction energy ΔE(A−B−C) of a stair motif with three interacting partners A, B, and C is defined by

$$\Delta E(\mathrm{A-B-C}) = E(\mathrm{A-B-C}) - E(\mathrm{A}) - E(\mathrm{B}) - E(\mathrm{C}) \quad (1)$$

It can also be defined in terms of pairwise interaction energies and the three-body correction ΔE_3, which reflects the possible nonadditivity of the pairwise interactions:

$$\Delta E(\mathrm{A-B-C}) = \Delta E(\mathrm{A-B}) + \Delta E(\mathrm{B-C}) + \Delta E(\mathrm{C-A}) + \Delta E_3 \quad (2)$$

All calculations were corrected for the basis set superposition error (BSSE) by using the standard counterpoise method.[36] More details on ab initio calculation methods can be found in ref 21.

Level of Quantum Chemistry Calculations. Calculations were performed with the second-order Møller−Plesset (MP2) perturbation theory,[37,38] which includes the electron correlation energy in addition to the Hartree−Fock (HF) energy. The use of such level of calculation is fully justified by the fact that the description of base stacking requires calculations with explicit inclusion of the electron correlation.[39] The interaction energy at a given order of the Møller−Plesset (MP) perturbation expansion is calculated as

$$\Delta E_{\mathrm{MP}n} = \Delta E_{\mathrm{HF}} + \sum_{i=2}^{n} \Delta E_{\mathrm{Corr}}(\mathrm{MP}i) \quad (3)$$

where ΔE_{HF} is the HF energy and ΔE_{Corr}(MPi) is the ith order perturbative correction to the correlation energy. Only the valence electrons were explicitly correlated in our computations, which corresponds to the usual frozen core approximation. We have also limited the perturbation expansion (3) to the second order, which is expected to take the major contributions to the van der Waals energies (electrostatic, polarization, dispersion, electron transfer and exchange contributions) into account. However, this expansion is not guaranteed to have converged at $n = 2$.[40] This convergence has been investigated by means of test calculations using higher-order correlation energy contributions (MP3 and MP4). The ΔE_{Corr}(MPn) contributions are evaluated from sum-over-states expressions involving the

following classes of excited states for $n \leq 4$: the singly (S), doubly (D), triply (T), and quadruply (Q) excited states with respect to the zeroth-order Hartree−Fock ground-state wave function. The MP2 and MP3 contributions arise from the D class only, whereas the MP4 correction implies S, D, T, and Q excitations. The full MP4 calculation is thus noted MP4(SDTQ), to be distinguished from calculations limited to given types of excitations, like for instance MP4(DQ) or MP4(SDQ). As such calculations require a large amount of computer power, they were performed on simplified complexes.

All MP2 calculations were performed with the GAUSSIAN 98 program,[41] and MP3 and MP4 calculations with the MOLPRO-2000 program.[42]

Results and Discussion

Stair Motifs in Protein−DNA Complexes. A set of 52 high-resolution (resolution ≤2.5 Å) crystal structures of protein/double-stranded DNA complexes were searched for stair motifs, involving two stacked nucleic acid bases and a positively or δ(+) charged amino acid side chain forming simultaneously an H-bond and a cation−π interaction with a base (Figure 1). A total of 77 stair motifs were identified, distributed among 36 complexes. They are given in Table 1. The list of proteins in the set and a structural description of these motifs can be found in ref 33.

Basis Sets for MP2 Calculations on Stair Motifs. Quantum chemistry calculations based on MP perturbation theory were performed up to the second-order. This level of calculations has been shown to be adequate for estimating cation−π binding energies[43−45] as well as base stacking energies.[46] However, the MP2 interaction energies strongly depend on the basis set used.[45,47,48] We therefore first investigate this dependency and determine the optimal basis set for describing the three types of interactions present in our system: nucleic acid base stacking, base-amino acid H-bond, and base-amino acid cation−π interaction.

Given the size of our systems, the computer needs (CPU, memory and disk space) for MP2 calculations rapidly increase with the size of the basis set. This is the reason the medium size 6-31G* or 6-31G** [49,50] basis sets are the most frequently used with biomolecules. They offer a valence double-ζ quality and are augmented by a single polarization function on each atomic center, except on H atoms with 6-31G*. The parameters of these basis sets (Gaussian exponents and coefficients) are optimized on bonded model systems, making them well suited for representing isolated molecules. It is, however, known that a proper description of nonbonding interactions is not ensured, in particular in the case of stacked aromatic species.[46,51] A basis set extension is thus required to improve the flexibility of the wave functions at short and medium range (valence region) and at long range (nonbonding interaction region). In principle both regions should be simultaneously improved to ensure a balanced description of all interactions.

One solution consists of adopting larger basis sets from the literature, like for instance the correlation consistent polarized Dunning's basis sets,[52,53] offering an increasing flexibility along the series cc-PVXZ (with X = T, Q, 5). The "augmented" versions of these basis sets, aug-cc-PVXZ,[54] contains additional diffuse functions that may be efficient in the context of nonbonding interactions. Such basis sets are, however, too prohibitive for dealing with the large systems considered in the present work. To get an idea, let us take the Gua∴Arg∨Gua motif as an example, where ∴ and ∨ denote cation−π and H-bond interactions, respectively. The basis set size N grows in the following way along the cc-PVXZ and aug-cc-PVXZ series:

cc-PVXZ:

$N = 444$, 1004, 1910, and 3246 for X = D, T, Q, 5, respectively

aug-cc-PVXZ:

$N = 742$, 1564, 2816, and 4582 for X = D, T, Q, 5, respectively

For comparison, a similar increase occurs within Pople's basis sets, when going for instance from 6-31G** ($N = 470$) to 6-311++G** [55] ($N = 684$). The dramatic increase of N is of course to be related with the corresponding computer costs, which in the best case scales formally with $O(N^4)$ at the MP2 level of calculation.[56]

An alternative to deal with large complexes is to keep the medium size 6-31G* basis set (422 basis functions) for its relative simplicity, but to adjust some of its parameters to improve its flexibility at long-range distances. Following the ideas developed in the literature[48,57−59] and applied with success in the framework of biomolecules,[46,60−62] we decided to optimize a single parameter, the Gaussian α_d-exponent of the d-polarization functions on the heavy atoms C, N, and O. The analytical form of such a d-basis function is

$$\chi_d(x,y,z;\alpha_d) = Nx^a y^b z^c e^{-\alpha_d (x^2+y^2+z^2)} \quad \text{with} \quad a+b+c = 2 \quad (4)$$

where N is a normalization factor, x, y, and z are the Cartesian coordinates (in a.u.) of the electron with respect to the nuclear center and a, b, and c are positive integers defining the Cartesian projections of the d atomic orbital. Also note that, as in the standard 6-31G* basis set, we decided to use the same value of α_d for describing the atoms of carbon, nitrogen and oxygen.[50]

In the following, the 6-31G* basis set with modified α_d values will be noted 6-31G(α_d). The basis set consisting of a standard 6-31G* basis set with $\alpha_d = 0.8$ plus one additional, more diffuse, d-function of exponent α_d, will be referred to as 6-31G*(α_d).

Optimal α_d Value for Intermolecular Binding Energies. To determine the appropriate α_d value for stair motifs, we computed MP2 interaction energies as a function of the α_d exponent, using the 6-31G(α_d) basis set, for nucleobase stacking, H-bond and cation−π interactions in a Gua∴Arg∨Gua stair motif. The geometry of the system is taken from the stair motif Gua(A7)∴Arg(C236)∨Gua(A8) in the crystal structure of tc3 transposase in complex with DNA (protein code 1TC3), which is depicted in Figure 1.

Figure 2 shows the interaction energies ΔE, computed at the MP2/6-31G(α_d) level, for the stacked Gua||Gua system (where || denotes stacking interaction), the Arg∨Gua H-bond, and the Arg∴Gua cation−π as a function of the α_d exponent. The MP2 energies appear to be strongly affected by the value of α_d. In particular, the energy of the Gua||Gua stacking changes from −0.2 kcal/mol for the default 0.8 α_d value to −2.8 kcal/mol for $\alpha_d = 0.2$. The H-bond and cation−π interactions show smaller ΔE changes of about 1 kcal/mol. The effects of α_d are more critical for interactions involving large dispersion contributions, and thus in relative terms, the energy changes follow the ordering: stacking ≫ cation−π > H-bond. Strikingly, the minimal interaction energy is invariably found, for all three types of interactions, at an α_d value close to 0.2. This value is similar to the one ($\alpha_d = 0.25$) used by other authors to estimate the stacking energies between aromatic systems.[46,47,60]

The adequacy of the value $\alpha_d = 0.2$ for computing interaction energies is further supported by the observation that MP2

TABLE 1: Ab Initio ΔE_{MP2}/6-31G(α_d=0.2) Interaction Energies of the Cation−π/H-bond Stair Motifs Observed in the 52-Protein Data Set[a]

Minor Groove

Ade ∴ Arg ∨ Thy	A ∴ Arg	A ‖ T	Arg ∨ T	A ∴ Arg ∨ T	Ade ∴ Arg ∨ Thy	A ∴ Arg	A ‖ T	Arg ∨ T	A ∴ Arg ∨ T
1MNM A19-E9-E8	−8.0	−2.4	−3.6	−11.8	1FJL A51-D4-D3	−5.9	−3.2	−16.5	−24.1
9ANT A51-D219-D218	−7.2	−4.1	−1.0	−8.7	2HDD A51-C12-C11	−5.2	−4.2	−11.3	−19.7
Ade ∴ Arg ∨ Cyt	**A ∴ Arg**	**A ‖ C**	**Arg ∨ C**	**A ∴ Arg ∨ C**					
1E3M A48-E12-E11	−0.9	−3.1	−27.4	−30.5					

Major Groove

Gua ∴ Arg ∨ Gua	G ∴ Arg	G ‖ G	Arg ∨ G	G ∴ Arg ∨ G	Gua ∴ Arg ∨ Gua	G ∴ Arg	G ‖ G	Arg ∨ G	G ∴ Arg ∨ G
1TC3 C236-A7-A8	−13.5	−2.7	−29.4	−41.6	1IGN A546-D23-D24	−10.3	−3.6	−34.4	−44.8
1A3Q A52-D606-D607	−12.8	−2.7	−31.5	−42.9	1A1G A124-B7-B8	−9.6	−3.7	−34.0	−43.9
1BC8 C61-A5-A6	−11.9	−3.4	−34.8	−46.1	1A3Q A54-D605-D606	−9.3	−3.5	−31.7	−41.5
1A1G A146-B6-B7	−11.1	−3.1	−34.9	−45.0	1PUE E232-A8-A9	−8.7	−3.8	−21.5	−31.3
2RAM A33-D6-D7	−10.9	−3.2	−29.3	−40.3	1PUE E235-A7-A8	−7.7	−2.1	−17.2	−25.0
1GD2 E82-B-6-B-5	−10.9	−4.5	−23.8	−36.5	1IGN A404-D30-D31	−6.9	−2.6	−15.1	−22.5
2NLL B328-C514-C515	−10.5	−3.5	−23.2	−34.4	1AM9 A336-G48-G49	−5.9	−3.6	−16.7	−24.3
1AWC A376-D8-D9	−10.3	−3.6	−27.3	−38.2					
Gua ∴ Arg ∨ Ade	**G ∴ Arg**	**G ‖ A**	**Arg ∨ A**	**G ∴ Arg ∨ A**					
1MNM D185-E1-E2	−6.5	−8.3	−16.0	−28.7					
Ade ∴ Arg ∨ Gua	**A ∴ Arg**	**A ‖ G**	**Arg ∨ G**	**A ∴ Arg ∨ G**	**Ade ∴ Arg ∨ Gua**	**A ∴ Arg**	**A ‖ G**	**Arg ∨ G**	**A ∴ Arg ∨ G**
1A1G A180-B1-B2	−9.5	−3.6	−34.7	−44.5	1MEY C72-A4-A5	−3.6	−6.4	−33.2	−41.7
1A73 A74-D17-D18	−6.5	−5.3	−33.9	−43.5	6MHT A240-D425-D426	−3.3	−5.6	−34.1	−41.8
Thy ∴ Arg ∨ Gua	**T ∴ Arg**	**T ‖ G**	**Arg ∨ G**	**T ∴ Arg ∨ G**	**Thy ∴ Arg ∨ Gua**	**T ∴ Arg**	**T ‖ G**	**Arg ∨ G**	**T ∴ Arg ∨ G**
1LAT A466-D11-D12	−3.6	−3.3	−33.3	−39.2	1MEY C78-A2-A3	−1.1	−4.2	−24.8	−29.3
1TRO A69-I1-I2	−2.8	−3.1	−33.7	−38.6	1BHM A155-C3-C4	−0.9	−4.4	−32.5	−37.3
1SKN P519-A7-A8	−2.7	−4.0	−25.6	−31.2	1B72 B290-D7-D8	−0.3	−4.4	−33.3	−37.1
1AU7 A49-D483-D484	−1.9	−3.5	−28.9	−33.6	1TSR B280-E12-E13	+0.1	−4.8	−33.5	−37.6
1AKH B185-C5-C6	−1.8	−4.2	−28.2	−32.8	1IGN A542-D22-D23	+0.4	−5.0	−29.0	−32.5
1UBD C342-B31-B32	−1.2	−4.5	−32.0	−36.5					
Cyt ∴ Arg ∨ Gua	**C ∴ Arg**	**C ‖ G**	**Arg ∨ G**	**C ∴ Arg ∨ G**	**Cyt ∴ Arg ∨ Gua**	**C ∴ Arg**	**C ‖ G**	**Arg ∨ G**	**C ∴ Arg ∨ G**
1A1G A174-B3-B4	−0.7	−4.0	−33.9	−37.9	1CRX A259-D11-D12	+0.4	−5.4	−34.6	−39.7
1GD2 E94-A-1-A1	+0.3	−4.2	−18.0	−21.8	1BC8 C64-A4-A5	+1.7	−5.8	−32.2	−36.6
Gua ∴ Lys ∨ Gua	**G ∴ Lys**	**G ‖ G**	**Lys ∨ G**	**G ∴ Lys ∨ G**	**Gua ∴ Lys ∨ Gua**	**G ∴ Lys**	**G ‖ G**	**Lys ∨ G**	**G ∴ Lys ∨ G**
2HDD A50-D28-D29	−32.0	−2.0	−35.4	−62.0	1A3Q B221-C507-C508	−13.6	−4.5	−32.9	−47.4
1MNM A38-F47-F48	−25.3	−3.2	−29.3	−52.9	1UBD C339-B32-B33	−11.5	−3.8	−35.5	−46.4
1TSR B120-F7-F8	−15.1	−3.6	−34.3	−47.1					
Gua ∴ Lys ∨ Thy	**G ∴ Lys**	**G ‖ T**	**Lys ∨ T**	**G ∴ Lys ∨ T**					
1HCQ A32-C5-C6	−32.4	−4.7	−16.8	−48.9					
Ade ∴ Lys ∨ Gua	**A ∴ Lys**	**A ‖ G**	**Lys ∨ G**	**A ∴ Lys ∨ G**	**Ade ∴ Lys ∨ Gua**	**A ∴ Lys**	**A ‖ G**	**Lys ∨ G**	**A ∴ Lys ∨ G**
1LAT A461-C5-C6	−13.2	−7.7	−36.5	−52.8	1HCQ A28-C3-C4	−4.9	−7.0	−30.4	−40.0
1MEY C22-A8-A9	−10.2	−5.6	−36.8	−48.9	2IRF L2075-C1026-C1027	−3.4	−6.5	−29.0	−37.6
1CRX A86-E14-E15	−8.8	−6.5	−37.1	−49.5					
Thy ∴ Lys ∨ Gua	**T ∴ Lys**	**T ‖ G**	**Lys ∨ G**	**T ∴ Lys ∨ G**					
1A73 A65-C2-C3	−2.7	−4.1	−36.4	−41.0					
Cyt ∴ Lys ∨ Gua	**C ∴ Lys**	**C ‖ G**	**Lys ∨ G**	**C ∴ Lys ∨ G**					
1LMB 4 3-2 29-2 30	−2.5	−3.0	−34.0	−39.3					
Gua ∴ Asn ∨ Gua	**Asn ∴ G**	**G ‖ G**	**G ∨ Asn**	**G ∴ Asn ∨ G**					
1LMB 4 55-1 13-1 14	+1.9	−3.6	−2.1	−4.1					
Gua ∴ Asn ∨ Ade	**Asn ∴ G**	**G ‖ A**	**G ∨ Asn**	**G ∴ Asn ∨ A**	**Gua ∴ Asn ∨ Ade**	**Asn ∴ G**	**G ‖ A**	**G ∨ Asn**	**G ∴ Asn ∨ A**
1B72 A286-D8-D9	+0.7	−8.3	−7.7	−15.4	1MEY C19-A9-A10	+1.4	−8.6	−7.7	−15.0
1AKH A120-C25-C26	+0.9	−8.4	−8.0	−15.7	1BGB A185-C804-C805	+2.2	−7.5	−6.9	−12.3
1B72 A253-D12-D13	+1.0	−8.0	−7.9	−15.2					
Gua ∴ Asn ∨ Cyt	**Asn ∴ G**	**G ‖ C**	**C ∨ Asn**	**G ∴ Asn ∨ C**					
1A1G A121-B8-B9	+0.8	−10.0	+1.7	−7.7					
Ade ∴ Asn ∨ Ade	**Asn ∴ A**	**A ‖ A**	**A ∨ Asn**	**A ∴ Asn ∨ A**	**Ade ∴ Asn ∨ Ade**	**Asn ∴ A**	**A ‖ A**	**A ∨ Asn**	**A ∴ Asn ∨ A**
1MNM D182-F50-F51	+0.0	−6.3	−8.8	−15.2	2HDD A51-C12-C13	+0.7	−7.3	−8.4	−15.1
1GD2 E86-A3-A4	−0.0	−6.4	−4.0	−10.3	1UBD C369-B29-B30	+1.4	−6.0	−6.4	−11.3
9ANT A51-D219-D220	+0.0	−6.4	−7.8	−14.3	1FJL A51-D4-D5	+3.0	−6.9	−9.4	−13.4
Ade ∴ Asn ∨ Cyt	**Asn ∴ A**	**A ‖ C**	**C ∨ Asn**	**A ∴ Asn ∨ C**					
1IGN A401-C7-C8	−0.2	−4.9	−3.9	−9.1					
Thy ∴ Asn ∨ Gua	**Asn ∴ T**	**T ‖ G**	**G ∨ Asn**	**T ∴ Asn ∨ G**					
3PVI A141-D10-D11	−2.7	−1.9	−6.7	−11.5					
Cyt ∴ Asn ∨ Ade	**Asn ∴ C**	**C ‖ A**	**A ∨ Asn**	**C ∴ Asn ∨ A**					
3PVI A140-C6-C7	−1.1	−2.6	−9.3	−12.9					
Gua ∴ Gln ∨ Ade	**Gln ∴ G**	**G ‖ A**	**A ∨ Gln**	**G ∴ Gln ∨ A**					
1A73 A63-D16-D17	+0.1	−6.8	−8.1	−15.0					
Ade ∴ Gln ∨ Ade	**Gln ∴ A**	**A ‖ A**	**A ∨ Gln**	**A ∴ Gln ∨ A**	**Ade ∴ Gln ∨ Ade**	**Gln ∴ A**	**A ‖ A**	**A ∨ Gln**	**A ∴ Gln ∨ A**
1RPE I28-A24-A25	−0.7	−5.7	−8.7	−15.1	1MEY C16-A10-A11	+0.5	−6.3	−3.1	−8.8
1UBD C396-B27-B28	−0.0	−5.1	−6.3	−11.4					
Thy ∴ Gln ∨ Ade	**Gln ∴ T**	**T ‖ A**	**A ∨ Gln**	**T ∴ Gln ∨ A**	**Thy ∴ Gln ∨ Ade**	**Gln ∴ T**	**T ‖ A**	**A ∨ Gln**	**T ∴ Gln ∨ A**
1AU7 A44-C459-C460	−0.9	−4.7	−7.7	−13.4	1LMB 3 44-1 3-1 4	−0.5	−5.0	−8.4	−13.8
3CRO I28-B3-B4	−0.6	−4.6	−7.0	−12.3					
Cyt ∴ Gln ∨ Ade	**Gln ∴ C**	**C ‖ A**	**A ∨ Gln**	**C ∴ Gln ∨ A**					
1MEY C44-A7-A8	−0.1	−3.5	−8.0	−11.5					

[a] The symbol ∴ means cation−π interaction, ∨ means H-bond, and ‖ means π−π stacking interaction.

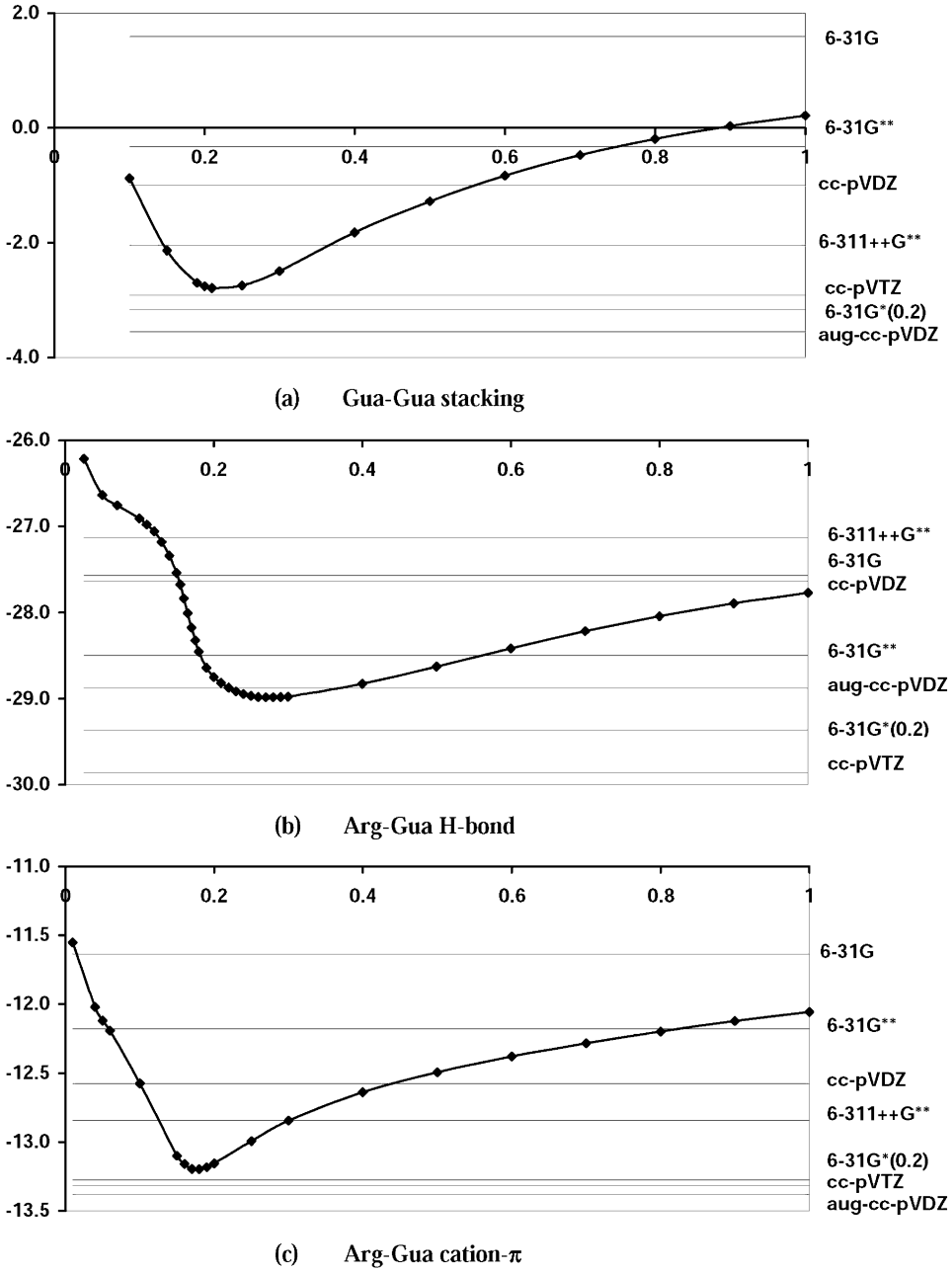

Figure 2. MP2 interaction energies ΔE with the 6-31G(α_d) basis set for (a) stacked Gua||Gua, (b) Arg∨Gua H-bond, and (c) Arg∴Gua cation−π, as a function of the α_d exponent of the d-polarization functions on C, N, and O atoms. The geometries of complex were taken from the 1TC3 (A7-C236-A8) stair motif, depicted in Figure 1. The energies obtained with the basis sets 6-31G*(α_d=0.2), 6-31G, 6-31G**, 6-311++G**, cc-pVDZ, cc-pVTZ, and aug-cc-pVDZ are shown with horizontal lines. The α_d(N), α_d(C), and α_d(O) coefficients are equal to 0.8, 0.8, and 0.8 in the basis 6-31G**, to 0.913, 0.626, and 1.292 in 6-311++G**, and to 0.817, 0.55, and 1.185 in cc-pVDZ. More extended bases contain two sets of α_d coefficients, namely 0.817, 0.55, and 1.185 and 0.23, 0.151, and 0.332 in aug-cc-pVDZ and 1.654, 1.097, and 2.314 and 0.469, 0.318, and 0.645 in cc-pVTZ. The 6-31G basis set has no d-polarization function. Moreover, p-polarization functions on H atoms are added in all extended basis sets and a set of f-polarization functions on heavy atoms in cc-pVTZ basis set.

interaction energies computed with the modified basis set 6-31G-(α_d=0.2) are close to those computed with aug-cc-pVDZ and cc-pVTZ, the most extended basis sets considered here, for all three types of interactions (Figure 2). The 6-31G(α_d=0.2) basis seems thus close to optimal; it yields −13.2 kcal/mol for the cation−π interaction, −28.7 kcal/mol for the H-bond, and −2.8 kcal/mol for the stacking interaction.

Note that for stacking and cation−π interactions, the interaction energies improve, as expected, when the basis set is extended from 6-31G to 6-31G**, cc-pVDZ, 6-311++G**, cc-pVTZ, and aug-cc-pVDZ. In contrast, for H-bond interactions, the cc-pVTZ basis sets gives the lowest values, roughly equivalent to the modified 6-31G*(α_d=0.2) basis, but the 6-311++G** basis gives less favorable interaction energies than the less extended bases 6-31G and 6-31G**. This is due to the fact that extending a basis set does not mean keeping the set unchanged and allowing additional flexibility. In contrast, all parameters, and in particular the α_d values, may be modified, as is clearly apparent in the legend to Figure 2. The global stabilization of a system with a given basis set comes from a competition between the different ways the system can acquire energy from the available basis functions. When an energy difference is calculated, like an interaction energy, both components of the difference should ideally acquire flexibility

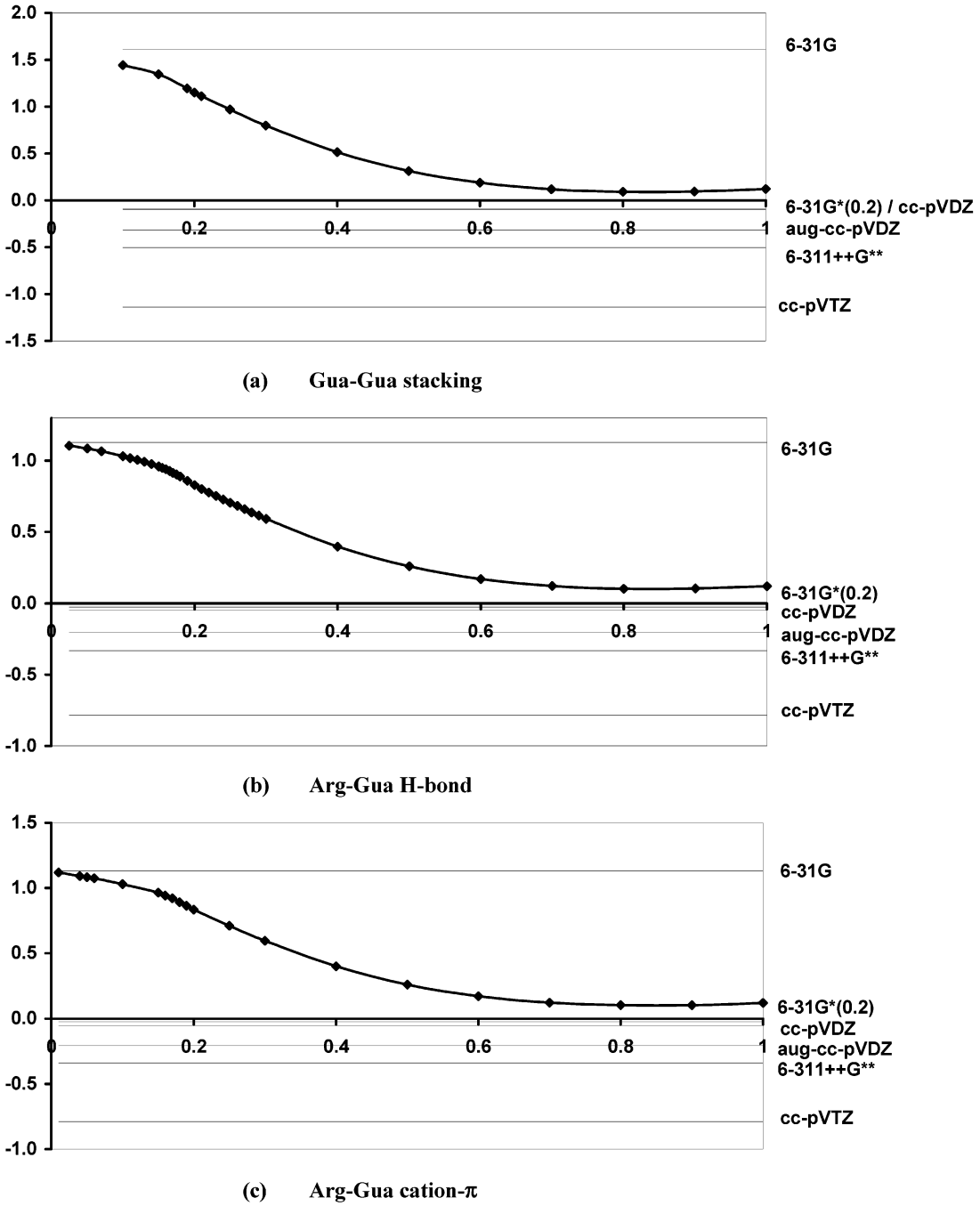

Figure 3. Difference between MP2 energies *E* calculated with the 6-31G(α_d) basis set and with the standard 6-31G** basis set, for (a) stacked Gua||Gua, (b) Arg∨Gua H-bond, and (c) Arg∴Gua cation−π, as a function of the Gaussian α_d exponent of the d-polarization functions on heavy atoms. The geometries of complex were taken from the 1TC3 (A7-C236-A8) stair motif, depicted in Figure 1. The corresponding energy differences computed with the basis sets 6-31G*(α_d=0.2), 6-31G, 6-311++G**, cc-pVDZ, cc-pVTZ, and aug-cc-pVDZ are indicated by horizontal lines.

in a balanced way. This is not always satisfied and may explain the a priori surprising results obtained with 6-311++G**.

Optimal α_d Value for Total System Energies. It must be stressed that a relevant basis set should in principle optimize the total energy of the system rather than the interaction energy. As seen in Figure 3, the value of α_d in the 6-31G(α_d) basis yielding the most favorable total energies is close to 0.80, for the three types of interactions considered; this generalizes to stair motifs previous results on smaller molecular systems.[50] The divergent trends of interaction and total energies need thus to be reconciled. This is achieved by defining a basis set containing simultaneously the values of α_d that optimize the interaction and total energies, i.e., $\alpha_d = 0.2$ and $\alpha_d = 0.8$; this basis set will be denoted 6-31G*(α_d=0.2). Figures 2 and 3 show that this basis set indeed allows to optimize both energies.

Moreover, the interaction energy computed with the 6-31G*-(α_d=0.2) basis set is quite close to that calculated with the most extended basis sets considered, i.e., aug-cc-pVDZ and cc-pVTZ. Indeed, 6-31G*(α_d=0.2) yields binding energies of −3.2, −29.4, and −13.3 kcal/mol for Gua||Gua stacking, Gua∨Arg H-bond, and Arg∴Gua cation−π, respectively, whereas the best interaction energies are −3.6 kcal/mol for base stacking (aug-cc-pVDZ basis), −29.9 kcal/mol for H-bond (cc-pVTZ basis), and −13.4 kcal/mol for cation−π (aug-cc-pVDZ basis).

The basis set 6-31G*(α_d=0.2) seems thus the most adequate for our system. However, calculations using a basis with two

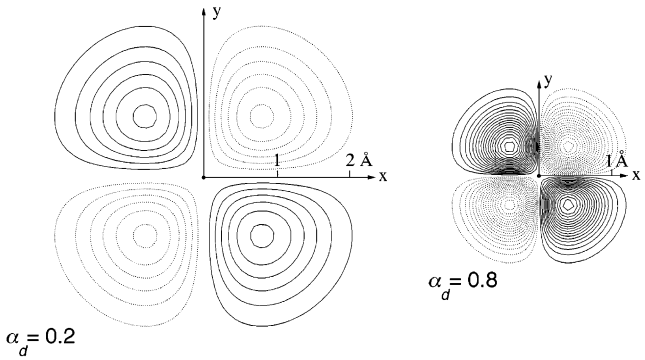

Figure 4. Isocontour plots in the *xy* plane of d_{xy} polarization basis functions (see eq 4) with an α_d exponent equal to 0.2 and 0.8. The pictures were generated with the MOLDEN program.[80] The scales on the *x* axes give an idea of the relative spatial extent of both functions. Full and dotted lines correspond to positive and negative values of the wave functions, respectively. The contour lines are drawn by steps of 0.0125 (absolute value of the wave function), starting from an absolute value of 0.0125 for the external line of each *d* lobe.

different α_d values introduces 5 extra functions per C, N, or O atom in the complex, which results in a significant increase of the computational costs. Therefore, as we focus here on interaction energies for which 6-31G*(α_d=0.2) and 6-31G-(α_d=0.2) yield roughly the same values, we decided to use the basis set 6-31G(α_d=0.2) in all subsequent MP2 calculations.

Interpretation of the Influence of the α_d-Exponent on Interaction Energies. A first interpretation of the observed trend that $\alpha_d = 0.8$ and $\alpha_d = 0.2$ are better suited for computing total and interaction energies, respectively, comes from the spatial extension of the d-polarization orbitals, which changes significantly as a function of α_d, as can be visualized in Figure 4. This picture, which presents isocontour maps of a $\chi = d_{xy}$ (eq 4 with $a = b = 1$, $c = 0$) in the *xy* plane, shows how the spatial extension of the wave function evolves with the value of the Gaussian exponent. The electron distribution in the $\alpha_d = 0.8$ orbital is indeed more contracted on the nucleus than in the $\alpha_d = 0.2$ orbital. The more diffuse $\alpha_d = 0.2$ orbital spreads over 2.4 Å from the atomic center and presents a maximum of electron density at about 0.8 Å. The $\alpha_d = 0.8$ orbital asymptotically ends at about 1.2 Å, with a maximum at 0.6 Å. Considering that the distance between two stacked bases or between the two cation−π partners of a stair motif is about 4 Å, there is a possible overlap of about 1 Å between two $\alpha_d = 0.2$ orbitals. No overlap is possible with $\alpha_d = 0.8$.

Another explanation, directly related to the above geometrical arguments, is that the diffuse nature of the $\alpha_d = 0.2$ orbitals allows a polarization of the wave function in the interaction region between the partners of the complex. As shown before,[57] this polarization effect improves significantly the multipole electric moments of the interacting partners, with, as a consequence, a better description of the dispersive contributions to the interaction energy.[40,63] To check this hypothesis, we calculated the dipole, quadrupole, and octopole moments of an isolated Gua at the HF/6-31G(α_d) level as a function of the α_d exponent value and compared them with those obtained using the HF/aug-ccVQZ level, taken as a reference. As seen in Figure 5, the difference between the moments computed with the 6-31G(α_d) and aug-ccVQZ basis sets tends to vanish for $\alpha_d = 0.2$, which demonstrates the correlation between optimal values of multipoles and of interaction energies. The same test calculations were not performed at the MP2 level to save computer time, but as already pointed out before,[57] the addition of the correlation will change the absolute values of the multipoles but not their qualitative behavior.

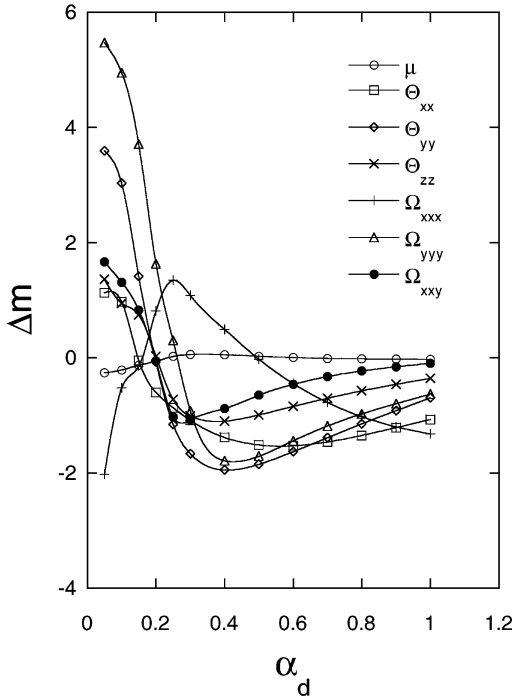

Figure 5. Variation of the HF values of the multipole electric moments of an isolated Gua molecule as a function of the α_d value. Δm is the difference between a given multipole component computed with the reference aug-ccpVQZ and the 6-31G(α_d) basis sets. Only the largest components of the multipole moments have been represented, corresponding to an orientation where the aromatic plane lies in the *xy* plane. Dipole (μ), quadrupole (Θ), and octopole (Ω) Cartesian components are given in D, DÅ, and DÅ^2, respectively. The absolute values of the moments calculated at the HF/aug-ccVQZ level are $\mu = 7.0343$ D, $\Theta_{xx} = -41.0077$ DÅ, $\Theta_{yy} = -70.9961$ DÅ and $\Theta_{zz} = -64.6958$ DÅ, $\Omega_{xxx} = 55.9881$ DÅ^2, $\Omega_{yyy} = -54.9324$ DÅ^2, and $\Omega_{xxy} = -29.6588$ DÅ^2.

MP2 Interaction Energies for Stair Motifs. Quantum chemistry calculations were performed on the 77 stair motifs identified in the set of protein/DNA complexes (Table 1), in view of determining the relative strength of the three pairwise interactions contained in the stair motif. The pairwise interaction energies computed at the MP2/6-31G(α_d=0.2) level show that the most favorable energies, −37 kcal/mol, are reached by the Lys∨Gua H-bonds. The most favorable stacking energy observed is equal to −10 kcal/mol, for the Gua||Cyt pair, and the most favorable cation−π energies occur for Arg∴Gua and reach −13 kcal/mol; actually, some of the Lys∴Gua have energies up to −32 kcal/mol but these are due to the simultaneous formation of an H-bond.

The first conclusion is thus that the cation−π energy is much less favorable, about 3 times, than the H-bond energy, and roughly as favorable as the π−π stacking energy. Note, however, that the energy computations were performed in a vacuum, and may not be directly transposed to more realistic environments consisting of water and/or protein residues. Indeed, ab initio energies of H-bonds are largely overestimated in a vacuum compared to water,[64,65] whereas stacking interactions are less overestimated;[65] cation−π interactions could be expected to display an intermediate behavior.

A more careful analysis of the results shows important fluctuations in the H-bond energies and frequencies. In particular, the overwhelming majority of H-bonds in the major groove involve Arg∨Gua and Asn/Gln∨Ade pairs. This can be

TABLE 2: ΔE_{MPn}/6-31G(α_d=0.2) Interaction Energies (kcal/mol) for a Stacking Interaction, a Cation−π Interaction, and a Stair Motif, as a Function of the Order n of the Pertubation Theory Contributions ($n = 2-4$) (See Eq 2)

stair motif		HF	MP2	MP3(D)	MP4(DQ)	MP4(SDQ)	MP4(SDTQ)
1TC3 (C236-A7-A8)	G || G	6.57	−2.76	0.38	−0.12	−0.68	
1LAT (A466-D11-D12)	T || G	1.53	−3.15	−1.91	−1.96	−2.31	
	T ∴ Arg	−1.84	−4.70	−4.02	−3.90	−4.16	
	Arg ∨ G	−31.89	−33.80	−33.70	−32.92	−33.26	
	T ∴ Arg ∨ G	−31.12	−40.29	−38.24	−37.44	−38.37	
1AKH (B185−C5−C6)	T ∴ Arg	0.82	−2.02	−1.34	−1.26	−1.46	−1.97

attributed to the fact that Arg can make a double H-bond with Gua in the major groove, and Asn and Gln with Ade. Though double H-bonds are expected to be more favorable energetically than single H-bonds, which is moreover supported by the observation that the substitution of an Arg into a Lys residue in an ets domain inhibits protein/DNA recognition,[66] this does not always appear to be true in our ab initio calculations. In particular, Lys∨Gua H-bonds are computed to be as favorable as, or even slightly more favorable than, Arg∨Gua, yielding energies of more than 30 kcal/mol. A possible explanation to this issue is that our calculations are performed in a vacuum and not in water or in a protein environment, and that when transposing vacuum energies to energies in a solvent, the energy values of systems with a localized net charge are more reduced than those of systems with a delocalized charge or with partial charges.[64,65,67,68]

In the case of cation−π interactions involving an Arg side chain, the frequency of occurrences and energetic values are quite well correlated. Indeed, Arg∴Gua, Arg∴Ade, Arg∴Thy, and Arg∴Cyt have minimal energies of −13.5, −9.5, −3.6, and −0.7 kcal/mol, respectively, and occur 16, 9, 11, and 4 times in the data set. The same is true for cation−π interactions involving a Lys side chain: Lys∴Gua, Lys∴Ade, Lys∴Thy, and Lys∴Cyt have minimal energies of −15.1 (if the matches where an H-bond is simultaneously formed are overlooked), −13.2, −2.7, and −2.5 kcal/mol and occur 6, 5, 1, and 1 times. In contrast, the higher frequency of Arg compared to Lys cannot be explained on an energetic basis. Here also, a possible explanation is related to the fact that the MP2 energy contribution is much more important in cation−π interactions involving Arg than in those involving Lys, due to the delocalization of the charge; as the electrostatic contributions are more overestimated than the electron correlation contributions in a vacuum compared to water,[64,65] the vacuum energies of Lys-containing cation−π interactions could be expected to be more reduced when transposed to water or protein environments than those of Arg-containing cation−π interactions,[26] thereby reconciling the calculated and observed trends.

In the case of cation−π interactions involving the partial charge located on the amino group of Asn or Gln side chains, also termed amino−π interactions, the computed energy values are in general slightly unfavorable for Ade and Gua, and slightly favorable for Thy and Cyt. For these interactions, the HF energy is generally unfavorable, whereas the MP2 energy is favorable. Hence the inclusion of the solvent effect may here also modify the conclusions and render all the interactions, including those involving Gua and Ade, favorable. Note that the energy values computed here are, in general, less favorable than in protein/ligand complexes.[26] This is due to the fact that in the protein/ligand context the amino acid side chains can be positioned straight above the aromatic cycles, whereas in protein/DNA complexes they are sterically hindered.

We also estimated the ΔE_3 term in eq 2, which measures the cooperativity of the interactions, for all stair motif geometries in Table 1, and found values in the range −0.4 to +7.4 kcal/mol. This result reflects the expected nonadditivity of the pairwise interactions, but not their cooperativity, which has however been suggested to be important in cation−π[69,70] and H-bond interactions.[71−73] The reason we do not observe the cooperative behavior of the interactions may be due to the fact that we deal with nonoptimized intermolecular geometries. This issue will be addressed in future work.

MP3/MP4 Corrections to the Interaction Energies. All above calculations were performed at the MP2 level. However, higher correlation energy corrections have been shown to be important,[45−48,63] and their contribution to the different components of the interaction energy (electrostatic, dispersion, induction, ...) has been quantified for small van der Waals systems.[40]

In the case of stair motifs, the introduction of MP3 and MP4 corrections or the use of the high level CCSD(T) coupled cluster approach[74−76] are too demanding in computer power to be systematically adopted. Rather, we performed test calculations on three of the complexes listed in Table 1, a Gua||Gua stacking, a Thy∴Arg∨Gua stair, and a Thy∴Arg cation−π interaction, so as to evaluate the order of magnitude of the MP3 and MP4 corrections. The calculations have been limited to MP4(SDQ) for the first two examples and were pushed to full MP4(SDTQ) in the last one. Note that to limit the computer costs the Thy∴Arg∨Gua complex has been simplified by replacing the amino group of Gua, the methyl group of Thy, and one amino group of the guanidinium moiety of Arg with an H atom. Despite these simplifications the computer needs remain however important, with, for instance, in the smaller case (Thy∴Arg), the following factors in the CPU time: 1 (HF), 1.5 (MP2), 15 (MP3), 26 (MP4(SDQ)) and 390 (MP4(T)). One sees that the addition of triples to MP4 is very expensive but gives a non-negligible energy increment,[46,63] as shown in the results which are summarized in Table 2. Figure 6 shows how the correlation energy corrections evolve along the MP perturbation theory expansion (5) for all considered complexes.

We remark that overall, the MP2 correction stabilizes the complexes, whereas the MP3 contribution produces an opposite effect of smaller intensity (between 5 and 34%). The full MP4 correction, only obtained for the smallest complex, a Thy∴Arg cation−π interaction, is globally stabilizing if one takes all the different classes of excited states (S, D, T, and Q) in the fourth-order perturbative correction into account. More specifically, as illustrated in Figure 6, the double and quadruple excited states yield a small positive increment to the interaction energy, whereas the single and triple excitations have a more important negative effect. Note that similar trends have been observed in the calculation of the interaction energies of the dimers of benzene and naphthalene[63] and of the CH_4/NH_3 complex.[48] It is worth noting that in the Thy∴Arg cation−π interaction, for which the full MP4 calculation has been performed, the MP4-(SDTQ) value of ΔE is close to the MP2 value, which is a nice example of oscillatory behavior of the MP expansions[48,63]and a demonstration of the reliability of the MP2 level of theory.

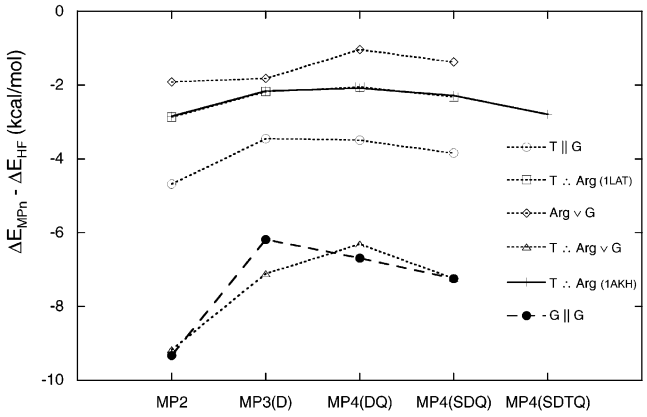

Figure 6. Evolution of the total correlation energy contributions $\Delta E_{MPn} - \Delta E_{HF}$ (eq 3) to the interaction energies as a function of the order n of the perturbation expansion, for a G||G stacking in 1TC3 (C236-A7-A8) (full line), a cation−π interaction in 1AKH (B185−C5−C6) (dashed line), and a Thy∴Arg∨Gua stair motif in 1LAT (A466-D11-D12) (dotted line). Detailed energies are given in Table 2.

The optimistic conclusion drawn in this particular case is, however, to be considered with care and is certainly not to be generalized.

Conclusions

A systematic survey of X-ray protein/DNA complexes allowed us to identify 77 stair motifs, exhibiting simultaneously a cation−π, an H-bond and a π−π stacking interaction. The recurrence of such cation−π/H-bond stair motifs at the protein−DNA interface suggests that they must play an important role, which, as yet, is not fully understood. Of course they play a stabilizing role. The conjunction of H-bond and cation−π interactions is indeed favorable energetically, though we were not able to demonstrate their cooperativity. The increased delocalization of an extraneous electron should also improve stability. Cation−π/H-bond stair motifs must moreover have a structural role, as their presence requires and induces very specific conformations of the DNA double helix. They must equally play a role in the specificity of protein−DNA recognition, not only through structural requirements but also because the strength of the cation−π interactions depends on the type of base and amino acid. It can moreover be argued[21,33] that they should play a role in the charge migration known to occur in double-stranded DNA upon oxidation of a Gua base.[77−79]

Furthermore, we showed that more diffuse polarization functions need to be introduced to correctly estimate the interaction energy of van der Waals type systems than for evaluating their total energy. In particular, setting the Gaussian α_d exponent of d-polarization functions in the medium size 6-31G* basis equal to 0.2 provides the best MP2 interactions energies of stair motifs, close to those obtained with the largest basis sets practicable with our molecular systems (aug-cc-pVDZ and cc-pVTZ), whereas $\alpha_d = 0.8$ yields the optimal total energies. This confirms earlier conclusions drawn from calculations on smaller interacting systems.[46,57−59,61] More diffuse polarization functions were found to improve the interactions energies, especially in the case of π−π stacking but also for cation−π and H-bond interactions, and more generally for interactions with important electron correlation contributions. This can be attributed to the properties of these polarization functions to allow a spatial overlap between the wave functions and a polarization in the interaction region, which improves the multipole electric moments of the interacting partners, and hence the description of the dispersive energy contributions.

The interaction energy of the 77 stair motifs observed in the data set was computed at the MP2 level using the 6-31G(0.2) basis set, with $\alpha_d = 0.2$. We found the stacking interactions to be the least favorable, with a minimum energy value of −10 kcal/mol, followed by the cation−π interactions reaching −13 kcal/mol, and the H-bond interactions up to −32 kcal/mol. Of course, these are vacuum energy values, and their transposition to values in water or a protein/DNA environment may entail significant modifications. The cation−π interactions involving a partial positive charge instead of a net charge, also termed amino−π interactions, are usually also computed to be favorable, but less. It must be stressed that it is the electron correlation contribution that renders these interactions attractive. Note that we chose not to optimize the stair motif geometries provided by X-ray structures. Indeed, the optimization of such motifs outside the protein/DNA context sometimes leads to distortions that are incompatible with the structural constraints provided by the environment in the native structures. This implies that the energy values computed in the present paper appear less favorable than those computed after intermolecular geometry optimization.

Finally, we also performed tests on the higher correlation MP energy corrections, for a few simplified complexes. We recovered the oscillatory behavior of the MP expansions, already observed before:[48,63] MP3 appears less favorable than MP2, and MP4 seems more favorable than MP3 but slightly less than MP2. Though these results need of course to be confirmed on many more examples, the similar values of MP2 and MP4 interaction energies can be taken to suggest that MP2 is a reliable approximation, at least for the stair systems considered.

Acknowledgment. We thank E. Buisine and J.-M. Kwasigroch for useful discussions. C.B. acknowledges support from a BioVal research program of the Walloon Region. M.R. is Research Director at the Belgian Fund for Scientific Research. J.L. and M.R. thank this fund for financial support through FRFC contracts. R.W., C.B., and M.R. also acknowledge support from the Communauté Française de Belgique through an Action de Recherche Concertée.

References and Notes

(1) Bränden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing Co.: New York, 1991.
(2) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman Company: New York, 1993.
(3) Yip, K. S.; Stillman, T. J.; Britton, K. L.; Artymiuk, P. J.; Baker, P. J.; Sedelnikova, S. E.; Engel, P. C.; Pasquo, A.; Chiaraluce, R.; Consalvi, V. *Structure* **1995**, *3*, 1147.
(4) Elcock, A. H. *J. Mol. Biol.* **1998**, *284*, 489.
(5) Gilis, D.; Wintjens, R.; Rooman, M. *Recent Res. Devel. Protein Eng.* **2001**, 1, 277.
(6) Dougherty, D. A. *Science* **1996**, *271*, 163.
(7) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303.
(8) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139.
(9) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709.
(10) Mitchell, J. B.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315.
(11) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459.
(12) Minoux, H.; Chipot, C. *J. Am. Chem. Soc.* **1999**, *121*, 10366.
(13) Lentz, T. L. *Biochemistry* **1991**, *30*, 10949.
(14) Luhmer, M.; Bartik, K.; Dejaegere, A.; Bovy, P.; Reisse, J. *Bull. Soc. Chim. Fr.* **1994**, *131*, 603.
(15) Zeng, H.; Moise, L.; Grant, M. A.; Hawrot, E. *J. Biol. Chem.* **2001**, *276*, 22930.
(16) Scrutton, N. S.; Raine, A. R. C. *Biochem. J.* **1996**, *319*, 1.
(17) Zacharias, N.; Dougherty, D. A. *Trends Pharm. Sc.* **2002**, *23*, 281.
(18) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, *253*, 872.
(19) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. *Proc. Natl. Acad. Sci U.S.A.* **1998**, *95*, 12088.

(20) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708.
(21) Wintjens, R.; Liévin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395.
(22) Novotný, J.; Bruccoleri, R. E.; Saul, F. A. *Biochemistry* **1989**, 28, 4735.
(23) Hsieh-Wilson, L. C.; Schultz, P. G.; Stevens, R. C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5363.
(24) Ortiz, A. R.; Pisabarro, M. T.; Gallego, J.; Gago, F. *Biochemistry* **1992**, *31*, 2887.
(25) Ortiz, A. R.; Pisabarro, M. T.; Gago, F. *J. Med. Chem.* **1993**, *36*, 1866.
(26) Biot, C.; Buisine, E.; Kwasigroch, J.-M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816.
(27) Loewenthal, R.; Sancho, J.; Fersht A. R. *J. Mol. Biol.* **1992**, *224*, 759.
(28) Fernández-Recio, J.; Vázquez, A.; Civera, C.; Sevilla, P.; Sancho, J. *J. Mol. Biol.* **1997**, *267*, 184.
(29) Fernández-Recio, J.; Romero, A.; Sancho, J. *J. Mol. Biol.* **1999**, *290*, 319.
(30) Ting, A. Y.; Shin, I.; Lucero, C.; Schultz, P. G. *J. Am. Chem. Soc.* **1998**, *120*, 7135.
(31) Burghardt, T. P.; Juranic, N.; Macura, S.; Ajtai, K. *Biopolymers* **2002**, *63*, 261.
(32) Shi, Z.; Olson, C. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 3284.
(33) Rooman, M.; Liévin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67.
(34) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777.
(35) Kabsch, W. *Acta Crystallog. Sect. A* **1978**, *34*, 827.
(36) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(37) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(38) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.
(39) Šponer, J.; Berger, I.; Špačková; N.; Leszczyński, J.; Hobza, P. *J. Biolmol. Struct. Dyn.* **2000**, *11*, 383.
(40) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
(41) 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.; Baboul, A. G.; 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.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.7; Gaussian Inc.: Pittsburgh, PA, 1998.
(42) MOLPRO (version 2000.1) is a package of ab initio programs written by H.-J. Werner and P. J. Knowles, with contributions from R. D. Amos, A. Bernhardsson, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, H. Stoll, A. J. Stone, R. Tarroni, and T. Thorsteinsson.
(43) Hoyau, S.; Norrman, K.; McMahon, T. B.; Ohanessian, G. *J. Am. Chem. Soc.* **1999**, *121*, 8864.
(44) Feller, D. *Chem. Phys. Lett.* **2000**, *322*, 543.
(45) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906.
(46) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
(47) Hobza, P.; Šponer, J.; Leszczyński, J. *J. Phys. Chem. B* **1997**, *101*, 8038.
(48) Rappé, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
(49) Ditchfield, R.; Hehre, W.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(50) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(51) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.
(52) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *98*, 4572.
(53) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
(54) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(55) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.
(56) Schütz, M.; Hetzer, G. ; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691.
(57) van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Mol. Struct. THEOCHEM* **1982**, *89*, 185.
(58) Kroon-Batenburg, L. M. J.; van Duijneveldt, F. B. *J. Mol. Struct. THEOCHEM* **1985**, *121*, 185.
(59) Tsuzuki, S.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **1998**, *102*, 2091.
(60) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
(61) Šponer, J.; Hobza P. *Chem. Phys. Lett.* **1997**, *267*, 263.
(62) Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871.
(63) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **2000**, *319*, 547.
(64) Floriàn, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884.
(65) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870.
(66) Pio, F.; Kodandapani, R.; Ni, C.; Shepard, W.; Klemz, M.; McKercher, S. R.; Maki, R. A.; Ely, K. R. *J. Biol. Chem.* **1996**, *271*, 23329.
(67) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.
(68) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622.
(69) Cadwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177.
(70) Eriksson, M. A. L.; Morgantini, P.-Y.; Kollman, P. A. *J. Phys. Chem. B* **1999**, *103*, 4474.
(71) Turki, N.; Milet, A.; Rahmouni, A. L.; Ouamerali, O.; Moszynski, R.; Kochanski, E.; Wormer, P. E. S. *J. Chem. Phys.* **1998**, *109*, 7157.
(72) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670.
(73) Šponer, J.; Burda, J. V.; Sabat, M., Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 5951.
(74) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.
(75) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.
(76) Scuseria, G. E.; Schaefer, H. F. III *J. Chem. Phys.* **1989**, *90*, 3700.
(77) Núñez, M. E.; Holmquist, G. P.; Barton J. K. *Biochemistry* **2001**, *40*, 12465.
(78) Meggers, E.; Michel-Beyerle, M. E.; Biese, B. *J. Am. Chem. Soc.* **1998**, *120*, 12950.
(79) Rajski, S. R.; Jackson, B. A.; Barton, J. K. *Mutat. Res.* **2000**, *447*, 49.
(80) Shaftenaar, G.; Noordik, J. H. *J. Comput. -Aided Mol. Design.* **2000**, *14*, 123.