MOLECULAR INTERACTIONS OF NUCLEIC ACID BASES. A REVIEW OF QUANTUM-CHEMICAL STUDIES

Jiří Šponer[a,b,*] and Pavel Hobza[b,*]

[a] *Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic; e-mail: sponer@ncbr.chemi.muni.cz*

[b] *J. Heyrovský Institute of Physical Chemistry and Centre for Complex Molecular Systems and Biomolecules, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic; e-mail: hobza@indy.jh-inst.cas.cz*

Received May 26, 2003
Accepted September 24, 2003

Dedicated to Professor Rudolf Zahradník on the occasion of his 75th birthday.

Ab initio quantum-chemical calculations with inclusion of electron correlation significantly contributed to our understanding of molecular interactions of DNA and RNA bases. Some of the most important findings are introduced in the present overview: structures and energies of hydrogen bonded base pairs, nature of base stacking, interactions between metal cations and nucleobases, nonplanarity of isolated nucleobases and other monomer properties, tautomeric equilibria of nucleobases, out-of-plane hydrogen bonds and amino acceptor interactions. The role of selected molecular interactions in nucleic acids is discussed and representative examples where these interactions occur are given. Also, accuracy of density functional theory, semiempirical methods, distributed multipole analysis and empirical potentials is commented on. Special attention is given to our very recent reference calculations on base stacking and H-bonding. Finally, we briefly comment on the relationship between

doi:10.1135/cccc20032231

advanced *ab initio* quantum-chemical methods and large-scale explicit solvent molecular dynamics simulations of nucleic acids.

Keywords: Molecular interactions; Nucleobases; Purines; Pyrimidines; Base pairs; DNA; RNA; Nucleic acids; *Ab initio* calculations; Quantum chemistry; Electron correlation.

1. INTRODUCTION

The structure and dynamics of nucleic acid (NA) molecules are influenced by a variety of contributions. Among those, the interactions present between the nucleic acid bases are of particular importance (Fig. 1). In DNA, the bases are involved in two qualitatively different mutual interaction types: hydrogen bonding and stacking. The H-bonded base-pair geometries observed at high resolution in crystal structures of DNA fragments correspond to the minima on potential energy surfaces of isolated DNA base pairs. In contrast, stacked configurations present in crystals of DNA fragments are rather variable and in many cases do not correspond to energetically optimal stacking arrangements[1]. Astonishing variability of molecular interactions of nucleic acid bases, far exceeding what is known from DNA, has been revealed by crystallographic studies of large RNA molecules, such as ribozymes, pseudoknots and recently even whole ribosomal subunits[2]. In RNA literally any kind of interactions may occur and be involved in formation of key tertiary interactions. Many of these geometries do not corre-

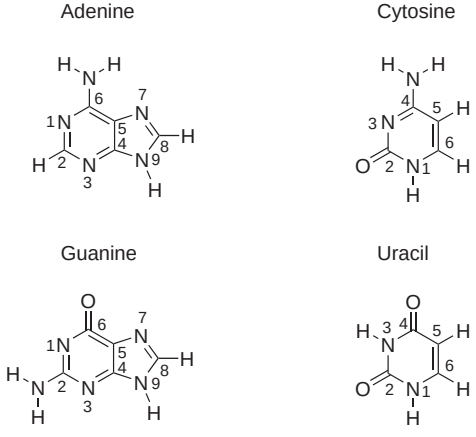

FIG. 1
Structures and atom numbering of RNA bases. Thymine base is the same as the uracil but it has a methyl group attached in position five instead of the hydrogen

spond even to the local minima on the intrinsic potential energy surfaces of the interacting subsystems[3].

Further, DNA bases interact with the water environment. High-resolution X-ray studies clearly reveal ordered hydration sites around nucleic acid bases[3] and ordered spine of hydration in DNA grooves[4]. DNA bases can further interact with metal cations, either directly (inner-shell binding), or indirectly, through a hydration shell around the cation[5]. The available experimental atomic resolution information regarding the metal binding in nucleic acids is very limited. Metal cations are involved not only in nonspecific charge screening (mostly through monovalent cations, *i.e.* K^+ and Na^+) but they also critically contribute to certain tertiary contacts, three-dimensional folds and even facilitate the RNA-catalyzed chemical reactions.

The electrostatic origin of stabilization of the H-bonded base pairs has long been widely accepted[6]. However, considerable uncertainty existed regarding the physical origin and magnitude of the base stacking phenomenon. Base stacking is characterized by an extensive overlap of the aromatic ring systems of adjacent base heterocycles, with delocalized π-electrons. It has been incorrectly postulated that this interaction type should result in very specific physical contributions rendering aromatic stacking qualitatively distinguishable from non-aromatic interactions[7]. These speculations became widespread in biological literature and are responsible for serious confusions. Until the mid-nineties, means did not exist to either verify or rule out the corresponding models of base stacking. It is one of the most important results of the recent state of the art quantum-mechanical studies to convincingly clarify that base stacking does not show any unusual properties that would set it apart from other molecular interactions[1,8]. Indeed, base stacking can be well described by utilizing any theoretical procedure properly describing electrostatic and dispersion terms, including the simplest form of all-atom empirical force fields[1,8]. In fact, base stacking is described by modern force fields more successfully than for example the sugar puckering (or amino acid conformations in proteins) as the bases are rigid and their electrostatic interaction may be well captured by constant atom-centered point charges derived quantum chemically from molecular wave functions.

To study the intrinsic interactions of DNA bases experimentally, one needs to carry out accurate gas phase experiments. Gas phase experiments provide large amounts of data giving insight into the physico-chemical origin of H-bonding. However, experiments on DNA base pairs are very difficult to perform. At this moment, regarding the association energetics, we still have to rely on the mass field spectroscopy data provided by Yanson

et al.[9a,9b] There is still no other reliable gas phase experiment reporting energetics of base pairing, even though any such experimental data would be highly vital. There exist valuable gas phase experimental studies on other aspects of nucleic acids and their constituents[9c–9h] and, in particular, the gas phase studies of de Vries and Kleinermanns on DNA base pairs should be mentioned. Thermal or laser desorption bring bases to vacuum and base pairs are formed in a beam of inert gas. The technique is combined with the mass spectrometry and IR vibration spectroscopy, which allows determination of the mass of a base pair as well as of its IR vibration characteristics. The experimental setup is limited to measurement of frequencies in the 3000–4000 cm^{-1} window but the O–H, N–H and C–H stretch vibration frequencies lie there. Combination of the method described with high level *ab initio* quantum-chemical calculations allows to determine the structure of a base pair *in vacuo*; valuable and also surprising data were collected. Unambiguous data about stacking of bases do not exist. The close to insurmountable obstacle for such studies is found in the fact that the stacked configurations might not be present in the course of the experiment, or the experiment shows a vast number of isomers populated simultaneously[10]. In fact, it has been suggested that the data by Yanson *et al.*[9a] likely show a mixture of structures, at least for some base pairs[10]. Some of their experiments with multiply methylated bases are likely to show predominantly stacking arrangements, though their structures are not known and the methylation certainly affects the energetics of stacking. Nevertheless, in recent years we have evidenced enormous efforts in the field of gas phase experiments of nucleobase complexes and we will certainly see a number of exciting new experimental results in the upcoming decade. As one of the major recent experimental results we consider the first experimental confirmation, carried out in helium droplets, that nucleobases are substantially nonplanar in their amino groups due to a partial sp^3 hybridization of their nitrogen atoms[1b,9i]. This has been convincingly predicted by electron correlation quantum-mechanical (QM) methods almost a decade ago, and nonplanar amino groups are nowadays assumed to affect important neighboring and tertiary interactions in nucleic acids, especially in RNA [1b]. Nevertheless, the absence of a convincing experimental counterpart to the QM predictions often significantly hampered our ability to convince structural biologists that the amino group nonplanarity is a real effect that is influencing the biomolecular structure and dynamics.

The other possibility open to study molecular complexes is based on the *ab initio* quantum-chemical theory with inclusion of electron correlation effects, a technique that has become feasible in the last decade due to the

rapid development of hardware and software. This computational approach should by no means be confused with older quantum-mechanical techniques of semi-empirical nature. Modern calculations with inclusion of electron correlation also brought about a substantial improvement compared to earlier *ab initio* computations that mostly neglected the electron correlation and relied on very small basis sets of atomic orbitals. Inclusion of correlation energy is substantial since it covers the London dispersion energy. The role of dispersion energy in biological environment is significant and this energy contribution plays a very decisive role. The unique role of the dispersion term is that it always stabilizes, contrary to other energy contributions (electrostatic) which might be attractive or repulsive. Dispersion energy thus plays a key role in the interaction of biomacromolecules. It is true that planar interaction of DNA bases leading to formation of H-bonded structures is satisfactorily described already at the Hartree–Fock (HF) level and dispersion only increases the stabilization by 20–50%. The other structural motif, stacked one, requires the full inclusion of the dispersion energy and if this energy is missing the theoretical treatment is meaningless. An important advantage of the quantum-chemical approach over the experiment is that by using quantum chemistry we can study any configuration of the complexes between interacting species, even those, which are far from the optimal geometry of the assemblies. For any single configuration of a particular complex we can unambiguously assign the corresponding energy. Thus, QM data provide a direct and unambiguous interrelation between molecular structures and energies. This is especially important since structures of biomolecular interactions can be obtained using X-ray crystallography but no information about energy is provided, leading sometimes misleading interpretations of the observed interactions. With contemporary computational tools, it would be very useful to rather routinely complement the structural data of novel and unusual interactions by advanced computations, and some examples can be found in the literature (see below).

A majority of QM data about energetics of molecular interactions has been in the last decade obtained using the electron-correlation second-order Møller–Plesset (MP2) perturbational method with medium-sized basis sets of atomic orbitals[1]. To obtain meaningful results, the basis set superposition error has to be corrected and, for stacking, diffuse polarization functions are needed. Good results for H-bonding can be also obtained using density functional methods (DFT) but the DFT approaches must strictly be avoided for base stacking due to their notoriously known failure for van der Waals interactions[1b]. First medium-level MP2 calculations appeared in the

mid-nineties providing qualitatively correct results for all types of molecular complexes of nucleic acid bases[1b]. It is now evident that accurate theoretical data requires the use of extended A-O basis sets in the MP2 treatment. Because the convergence of the MP2 energy is very slow, it is necessary to consider calculations with the infinite A-O basis set, which is realized by extrapolation of MP2 (and also HF) energies to the complete basis set (CBS) limit. Such calculations, when complemented by coupled cluster evaluations of higher-order contributions to electron correlation with consideration of triple contributions (CCSD(T) method), are assumed to be of quantitative accuracy. The inclusion of CCSD(T) terms is inevitable since differences between the MP2 and CCSD(T) stabilization energies are negligible in some cases (*e.g.* base pairing) but significant in others (*e.g.* base stacking). It is to be noted, nevertheless, that the CCSD(T) data are still prohibitively costly and only a limited set of geometries could be investigated.

Evidently, we cannot neglect experimental data from studies on base stacking and H-bonding in condensed phase, as these provide unique information about the association phenomena of nucleic acid bases in polar solvents and in nucleic acids[11]. Bulk water plays a key role and determines the structure of a pair. While H-bonded pairs are mostly more stable in the gas phase, stacked structures exist predominately in polar solvents. Nonpolar solvents behave similarly to the gas phase and the same structural motif is found in these solvents. Information on base-pair properties in polar solvents is difficult to obtain by computational methods, due to serious methodological problems and approximations concerning inclusion of solvent effects[12]. Not surprisingly, the number of computational studies on base–base interactions including solvent effects is quite limited[8b,13–15]. Obviously, the effects of environment are of primary importance, as nucleic acids do not exist in the gas phase. However, the condensed phase experiments are unable to unambiguously identify the physical origin of the direct (net, intrinsic) base–base interactions, as evidenced by mutually contradictory conclusions of such studies. In fact, different experimental methods may capture different aspects of the stacking phenomena and thus it is not so surprising that particular experimental designs may lead to conflicting conclusions, as we recently discussed[16]. A nice computational study on base stacking in a solvent has been recently presented by Luo and co-workers[14f]. These authors have designed calculations to closely parallel experimental conditions used by Newcomb and Gellman in their studies[11f]. These experimental data were originally interpreted as proving electrostatic origin of base stacking. The computational study shows that the experiments should be re-interpreted in favor of the dispersion origin of base stacking stabiliza-

tion, ending thus a long literature dispute[14f]. It is to be noted that the outcome of condensed phase experiments is affected not only by the common "continuum" solvent screening effects, but in some cases also by highly specific interactions between the studied system and its environment which may lead to entirely unexpected measured trends. Further, the experiments cannot be easily used to parametrize the empirical force fields which are needed for biomolecular modelling and simulations. This means that the gas phase approaches remain of primary relevance in studies of molecular interactions in nucleic acids and complement the condensed phase experimental approaches.

In the absence of data from gas phase experiments on energetics of nucleobase interactions, the nonempirical quantum-chemical calculations are the only tool currently available to evaluate the intrinsic interaction energies between bases. In the present short overview, we provide a very brief summary of the major outcomes of *ab initio* quantum-chemical calculations carried out in our laboratory in the past decade and their relevance to nucleic acid structure. Special attention is paid to the very recent results and the ongoing research. We would like to underline that the purpose of this contribution is not to provide an overall review of the field and the reader can find more information in our other review papers, including literature surveys, discussion of reliability and accuracy of the methodologies, and others[1b,16–20].

2. RESULTS

2.1. Tautomeric Equilibria of DNA Bases in the Gas Phase, in a Microhydrated Environment and in Aqueous Solution

DNA bases can undergo proton shifts while their neutrality is not changed, thus forming rare tautomers. Canonical tautomers are present in DNA and RNA but this does not mean they always make the global minimum. Rare tautomers may be involved in proton transfer processes, stabilize mispairs, promote point mutations and also play other so for unknown roles. It is, however, necessary to say that direct and unambiguous evidence of their presence in biomolecules is rare. The computational literature devoted to gas phase tautomerism of nucleobases is extensive but considerably less attention was devoted to tautomeric equilibria in a bulk solvent and also in a microhydrated environment. Here we summarize tautomeric equilibria of cytosine, guanine and adenine with a special attention paid to the role of solvent.

Tautomeric properties of cytosine. Canonical, enol and imino tautomers of cytosine (Fig. 2) were studied in the gas phase, in a microhydrated environment (1 and 2 molecules of water) and in bulk water[21a]. The relative energies of isolated tautomers presented in Table I were determined for RI-MP2/TZVPP geometries using the MP2 and CCSD(T) energies and Hartree–Fock thermodynamical characteristics. One of the enol tautomers forms the global minimum at all theoretical levels while the canonical form represents the first local minimum. However, already two water molecules reverse the relative stability of these two tautomers making the canonical form the global minimum. The effect of bulk water on relative stabilities of various tautomers was examined using the self-consistent reaction field, Monte Carlo and molecular dynamics free energy calculations. Bulk solvent unambiguously favored the canonical tautomer over the enol forms (*cf.* Table I) and the microsolvation data are in-between the gas phase and solvent predictions. These predictions basically agree with the values of dipole moments of various tautomers where the highest dipole moment possesses just the canonical form. The high dipole moment indicates a significant stabilization in a polar solvent and also high stabilization energy with one or two water molecules.

Tautomeric properties of guanine. A very different situation was found for tautomeric equilibria of guanine. Altogether eight keto and enol tautomers of guanine (Fig. 3) were studied theoretically in the gas phase, in a microhydrated environment (1 and 2 molecules o water) and in bulk water[21b]. Various imino tautomers, being considerably higher in energy,

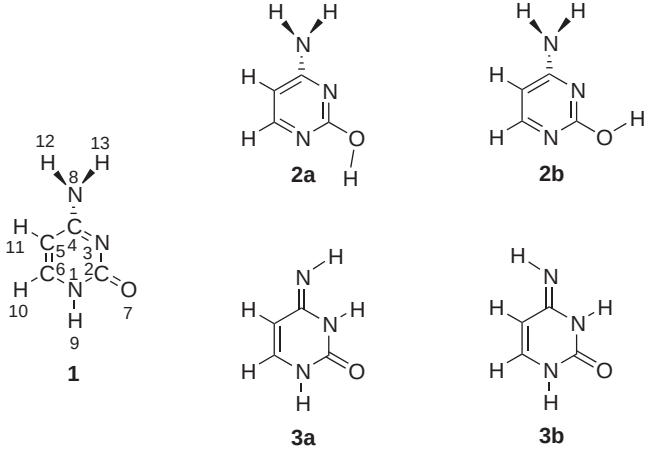

FIG. 2
Most stable cytosine tautomers

were not considered. The structures of isolated, mono- and dihydrated tautomers were determined at the RI-MP2 level using the TZVPP basis set. The relative energies of isolated tautomers included the correction to higher correlation energy terms evaluated at the CCSD(T)/aug-cc-pVDZ

TABLE I
Relative (compared to the canonical tautomer) gas phase energies (*E*) and relative hydration free energies (*G*) of various cytosine tautomers (in kcal/mol) shown in Fig. 2

Structure	E^a	G(EMST)^b	G(MC-FEP)^c	G(MD-TI)^d	$G_{av}{}^e$	$G_{tot}{}^f$
1	1.4	0.0	0.0	0.0	0.0	0.0
2a	0.0	7.1	7.5	6.5	7.0	6.4
2b	0.7	6.5	5.5	5.0	5.7	5.8
3a	3.7	4.1	3.5	3.7	3.8	7.3
3b	2.1	4.6	4.0	3.2	3.9	5.9

[a] CCSD(T)/complete basis set//RI-MP2/TZVPP. [b] Self-consistent reaction field. [c] Monte Carlo free energy perturbation. [d] Molecular dynamics-thermodynamic integration. [e] Average of previous three values. [f] G_{tot}, relative free energy in aqueous solution.

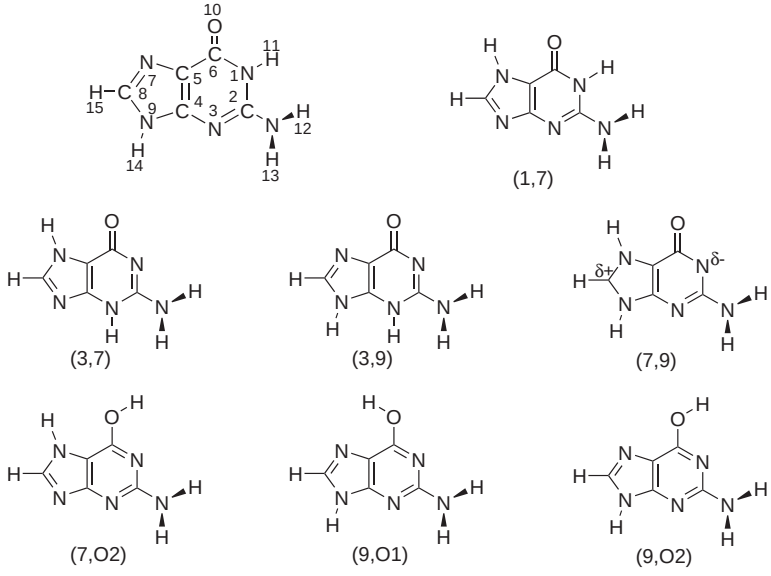

FIG. 3
Most stable guanine tautomers

level. The relative enthalpies at 0 K and relative free energies at 298 K were based on relative energies mentioned and zero-point vibration energies, temperature-dependent enthalpy terms and entropies evaluated at the MP2/6-31G** level. Final values of relative energies and free energies are presented in Table II. The keto form having hydrogen at N7 makes the global minimum at all theoretical levels in the gas phase while the canonical form having hydrogen at N9 represents the first local minimum. Both rare tautomers having hydrogens at N9 and N7, and at N3 and N9 are systematically considerably less stable and only four low-energy tautomers coexist in the gas phase. The theoretical predictions fully agree with existing theoretical as well as experimental results. Let us notice that the two tautomers having hydrogens at N9 and N7, and at N3 and N9 have extremely large dipole moments, considerably larger than that of canonical guanine which is the largest among all canonical DNA bases. As mentioned above, a large dipole moment indicates significant stabilization by a bulk water. The effect of bulk solvent on the relative stability of guanine tautomers was studied by self-consistent reaction field and molecular dynamics calculations of free energy using the thermodynamic integration method. Bulk solvent surprisingly strongly favored both unusual tautomers over all remaining low-energy tautomers and only these two forms are predicted to form in aqueous solution (*cf.* Table II). Addition of one or two water molecules does not change the relative stability order of isolated gua-

TABLE II
Relative gas phase energies (E) and relative hydration free energies (G) of various guanine tautomers (in kcal/mol) shown in Fig. 3

Structure	E [a]	G(MD-TI) [b]	G_{tot} [c]
(1,7)	–0.7	–0.9	–1.3
(9,O2)	0.2	5.6	5.7
(9,O1)	0.6	–	–
(7,O2)	3.0	3.1	6.0
(1,9)	0.0	0.0	0.0
(3,7)	5.8	–18.5	–13.0
(3,9)	19.0	–24.8	–7.1
(7,9)	20.0	–30.7	–10.9

[a] CCSD(T)/aug-cc-pVDZ//RI-MP2/TZVPP. [b] Molecular dynamics-thermodynamic integration. [c] G_{tot}, relative free energy in aqueous solution.

nine tautomers but the respective trends clearly support the surprising stabilization of both unusual forms. Addition of a higher number of water molecules (6–8) will probably change the tautomeric equilibrium even in the gas phase resulting in the rare tautomers as the most stable forms (to be published).

Tautomeric properties of adenine. Situation with adenine is again different[21c]. In this case we studied for the first time all 14 adenine tautomers. In addition to the standard tautomers for which the mesomeric structure can be assigned, we considered also two unusual tautomers having hydrogens at N7 and N9 (Fig. 4). As seen from the figure, the mesomeric structure can-

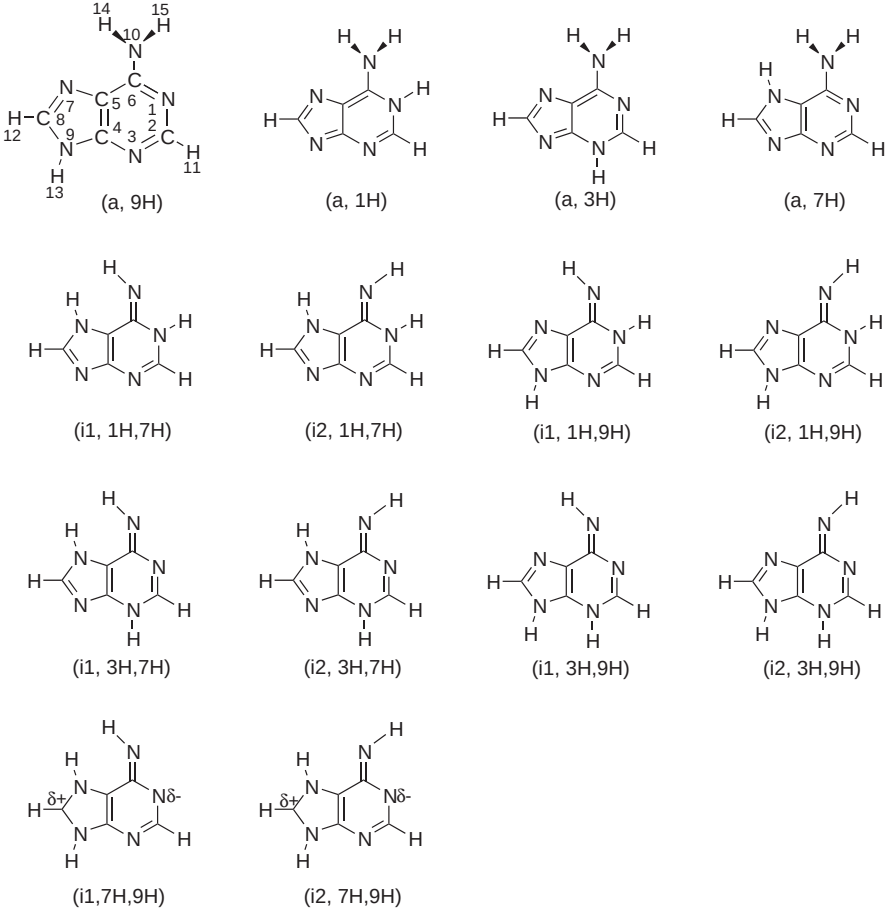

FIG. 4
Adenine tautomers

not be ascribed and N1 and C8 atoms bear formally a partial charge. Such structure is assumed to be associated with large dipole moment. Similarly to previous cases, we studied adenine tautomeric equilibria in the gas phase, in a microhydrated environment and in bulk water. All theoretical treatments were performed at the same level as in the case of guanine[21c]. Final values of relative energies and free energies in the gas phase are presented in Table III. Contrary to cytosine and guanine, the canonical form of adenine makes the global minimum and the first and second minima are well energetically separated (by about 7 kcal/mol). Remaining tautomers are energetically considerably less stable. Water phase plays an important role but, contrary to guanine, it does not change the stability order. The canonical structure remains the global minimum even in the water phase but the free energy differences between the global minimum and first two local minima become considerably smaller. This gives an evidence that the first and second local minima could co-exist with the dominant canonical form.

TABLE III
Relative gas phase energies (E) and relative hydration free energies (G) of various adenine tautomers (in kcal/mol) shown in Fig. 4

Structure	E [a]	G(MD-TI) [b]	G_{tot} [c]
(a,9H)	0.0	0.0	0.0
(a,1H)	17.7	–11.5	5.9
(a,3H)	8.0	–5.0	2.5
(a,7H)	7.6	–4.7	2.8
(i1,1H,7H)	16.1	–7.8	36.0
(i2,1H,7H)	16.6	–7.3	27.8
(i1,1H,9H)	18.5	–4.0	8.0
(i2,1H,9H)	12.1	–	–
(i1,3H,7H)	17.5	–	–
(i2,3H,7H)	24.3	–5.1	10.6
(i1,3H,9H)	31.9	–12.2	10.9
(i2,3H,9H)	31.6	–12.1	4.8
(i1,7H,9H)	35.5	–24.8	5.1
(i2,7H,9H)	45.0	–21.3	8.9

[a] CCSD(T)/aug-cc-pVDZ//RI-MP2/TZVPP. [b] Molecular dynamics-thermodynamic integration. [c] G_{tot}, relative free energy in aqueous solution.

Similarly to cytosine and guanine, the microhydration data support the results found in the bulk solvent.

Thus, recent calculations indicate that some nucleobase tautomers could be considerably more stable in polar environment than previously thought. Their formation in nucleic acids, however, remains to be proven.

2.2. Nonplanarity of Amino Groups of DNA Bases, Out-of-Plane H-Bonds and Amino Acceptor Interactions

One of the major outcomes of quantum-chemical studies of interactions of DNA bases was the discovery of the intrinsic nonplanarity and high flexibility of amino groups of DNA bases and their involvement in specific interactions in nucleic acids. The *ab initio* calculations unambiguously predict that the amino groups of bases are intrinsically nonplanar, with a partial sp^3 hybridization of the amino group nitrogen atoms (Table IV, Fig. 5)[22]. Direct experimental evidence of this nonplanarity was missing for a long time and only very recently Miller *et al.*[22e] fully confirmed our almost ten years old theoretical predictions. Pyramidalization means that the amino group hydrogens deviate from the nucleobase plane in one direction, while the amino nitrogen is slightly shifted in the opposite direction. In addition, a negatively charged lone-pair region arises above the nitrogen. The non-

TABLE IV

Nonplanar geometries of isolated DNA bases. (Reference MP2/6-311G(2df,p) data. Inversion barrier is the energy difference between nonplanar and planar optimized structures. Sum of amino group valence angles is 360° for a planar amino group.)

Base	Dihedral angle, °		Sum of amino group valence angles, °	Inversion barrier kcal mol^{-1}
Cytosine	C5–C4–N4–H41, –21.4	N3–C4–N4–H42, +12.6	351.9	–0.15
Adenine	C5–C6–N6–H41, –15.3	N1–C6–N6–H62, +16.5	352.9	–0.13
Guanine	N3–C2–N2–H21, –13.3	N1–C2–N2–N22, +39.2	339.6	–1.12
6-Thioguanine	N3–C2–N2–H21, –13.5	N1–C2–N2–H22, +38.0	340.6	–0.98
2-Aminoadenine	C5–C6–N6–H61, +18.1	N1–C6–N6–H62, –17.0	351.0	–
2-Aminoadenine	N3–C2–N2–H21, +22.2	N1–C2–N2–H22, –22.1	345.0	–0.79[a]

[a] Calculated for both amino groups.

planarity of guanine is larger than that of cytosine and adenine, and the amino group of guanine is also substantially rotated due to repulsion between the H1 hydrogen and the amino hydrogen[22b]. The amino group hydrogen atoms are very flexible and can be involved in two kinds of novel interactions. They can form very efficient out-of-plane hydrogen bonds and, further, the amino group nitrogen atom can serve as a weak H-bond acceptor. Both kinds of interactions are relevant for nucleic acid structure and have been studied in several recent crystallographic studies[23–26]. From the point of view of structural biology, the flexibility of the hydrogens is more important than the optimal nonplanar geometry itself. The available versions of AMBER and CHARMM force fields do not include these effects and out-of-plane H-bonds and amino acceptor interactions are neglected by current molecular dynamics simulations. Inclusion of the amino group nonplanarity into the molecular mechanics force fields, however, is far from being trivial. In reality, the amino group may be anywhere between the fully planar sp^2 arrangement and partial sp^3 pyramidalization, depending on its environment. The force field, in fact, can be tuned to be planar or nonplanar; however, it is not straightforward to achieve a switch between sp^2 and sp^3 arrangements and to model the lone pair with an analytical force field lacking molecular orbitals. It is assumed that in Watson–Crick base pairs and other planar H-bonds, the amino groups are mostly planar (as clearly evidenced for example by a systematic Cambridge Database search, and also the QM calculations). However, when the amino group is not fully saturated by primary in-plane H-bond, or in the presence of strong interactions promoting the nonplanarity, the sp^3 pyramidalization inherent to isolated bases may be expressed. The lack of reliable force field considering the amino group nonplanarity led us to reparametrize the empirical force field of Cornell *et al.*[26] Let us remind that this force field is by far the most suitable for simulation of DNA, providing surprisingly accurate data for interaction of DNA bases. It is important to mention that it

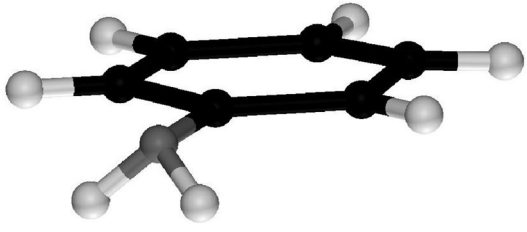

FIG. 5
Nonplanar amino group attached to an aromatic ring (sketch)

well describes both H-bonding and stacking of nucleobases. The amino group reparametrization was based on accurate geometries and energies of nonplanar and planar guanine, adenine and cytosine and was fully compatible with original parametrization in the other aspects. The new version of potential accurately described the nonplanarity of isolated bases and also (and this is more important) the changes of nonplanarity upon formation of H-bonded and stacked base pairs in the gas phase.

Nonplanar amino groups in nucleic acids. One of the interactions apparently influenced by amino group nonplanarity is interstrand mutual amino group contact in ApT B-DNA base-pair steps, invariably occurring in almost all high-resolution X-ray B-DNA structures. This ApT step has a highly conserved geometry and the average N6-N6 distance is 3.15 Å (refs[23,27]), making it the closest contact between adjacent base pairs in DNA crystals. Quantum-chemical calculations predict that close amino group contacts are inherently nonsymmetrical interactions and, indeed, they are systematically absent in those base-pair steps where a twofold symmetry is imposed by the crystal packing[23]. In the case of true two-fold symmetry of the ApT step, the N6–N6 distance increases to *ca* 3.4 Å.

The substantially nonplanar amino group of guanine was recently suggested to facilitate a novel minor groove binding mode of 4′,6-diamidino-2-phenylindole (DAPI) to a B-DNA duplex at 1.9 A resolution[25b]. In this particular case the guanine amino group involved in a standard Watson–Crick base pair with cytosine bends its outer (unpaired) hydrogen to accommodate the proximal amidine group of DAPI. Two proximal water molecules seen in the crystal appear to stabilize the interaction, one serving as donor and the other as acceptor with respect to the nonplanar amino group. Interestingly, as the amino group is pyramidal-rotated, the Watson–Crick base pairing is not affected. This local geometry is not reproduced in the course of empirical potential molecular dynamics[25c].

The out-of-plane bond between the nonplanar guanine amino group in the highly propeller twisted cis Watson–Crick G/A base pairs (Fig. 6) and the adjacent thymine in the $d(CCAAGATTGG)_2$ crystal structure is another convincing example (Fig. 6)[24,25d]. Here, the amino group of mismatched guanine is not involved in the primary base pairing and, in addition, faces a C2–H2 group on adenine. This makes the base pair substantially propeller twisted and the amino group profoundly pyramidal. The amino group then establishes an out of plane H-bond with O2 carbonyl oxygen on adjacent thymine. We have recently identified similar interactions in RNA including small and large ribosomal subunits[25d] where *ca* 40% of cis Watson–Crick G/A base pair form this out-of-plane bond. Further and more importantly

the phylogenetic analysis shows that the out-of-plane H-bond dramatically affects the conservation patterns. In the presence of the out-of-plane H-bond in the primary crystal structure the G/A base pair is highly conserved with no G/A to A/G covariation. In contrast, in the absence of the out-of-plane H-bond and tertiary contacts the G/A base pairs are not conserved. Thus, the amino group pyramidalization markedly contributes to the unique structural properties and conservation patterns of cis Watson–Crick G/A base pairs, this base pair is so far the most striking example of a biological role of the amino group pyramidalization effects[25d].

Nonplanar amino groups could also facilitate binding of divalent metal cations to the N7 position of adenine[28a]. A very similar geometry to what has been discussed above for the DNA–DAPI complex is predicted to occur in the case of N7 binding of hydrated metal cations to adenine. The N6 amino group serves as a H-bond acceptor with respect to one of the polarized water molecules from the cation hydration shell. Again, the H-bonding

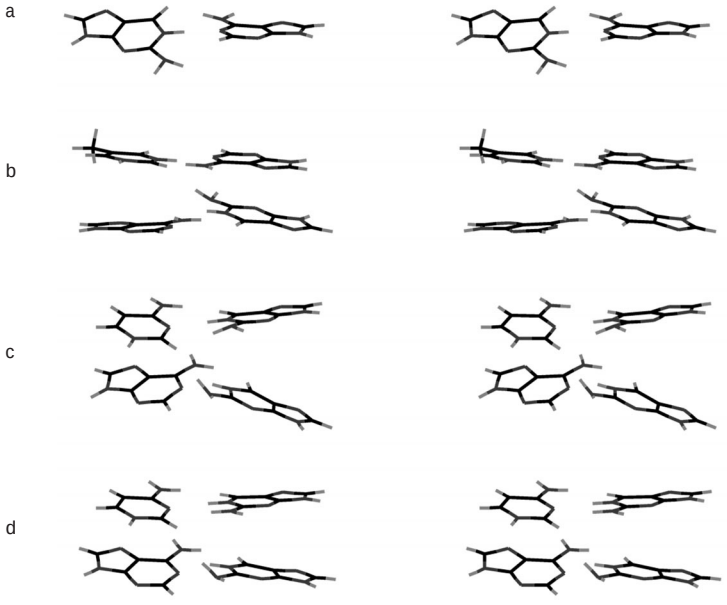

FIG. 6

Stereo views of cis Watson–Crick G/A base pair in the gas phase (a) and three base-pair steps (b–d) where the nonplanar amino group of G/A guanine is involved in out-of-plane H-bond with the adjacent base pair. These interactions were discovered based on gas phase QM calculations and have a major influence on the conservation patterns of G/A base pairs in ribosomal RNA [25d]

with thymine or uracil is unaffected as the amino group is pyramidal-rotated. Another potential role of amino group nonplanarity could be the intramolecular proton transfer between protonated and neutral cytosines in d(G.G.C) triplexes[28b]. It is interesting to note that current MD simulations do not reproduce the close amino group contacts in ApT B-DNA steps, which might be due to the lack of amino group pyramidalization effects in the force fields[28c].

Current reference values of amino group pyramidalization are based on gradient optimizations carried out at the MP2 level with 6-311G(2df,p) basis set[17] and the RI-MP2 level with TZVPP basis set (Table IV)[26], supplemented by CCSD(T) calculations on other molecules such as aniline and aminopyridines[22c]. For aniline, we have carried out a five-dimensional anharmonic vibrational analysis, which is in excellent agreement with experimental microwave spectra[22c]. We would like to underline that, as we have found out long time ago, DFT calculations show a rather wide variability of results concerning the nucleobase amino group pyramidalization and DFT results are thus method-dependent[22d].

Interestingly, when considering the *ab initio* studies aimed at explanation of bifurcated H-bonds and close amino group contacts involving paired amino groups of Watson–Crick base pairs, it seems, that the nonplanarities predicted by the model *ab initio* calculations are often rather subtle[23,25]. This indicates that the calculations perhaps underestimate the actual pyramidalization effects, even though we would intuitively assume that the pyramidalization might be exaggerated for calculations on small model complexes. This indicates that we might be overlooking some contribution. One of the possible sources of this discrepancy is neglect of the sugar-phosphate backbone in the calculations. Preliminary model calculations have shown that there might be a certain degree of charge transfer when including the negatively charged backbone segment. At the HF/6-31G* level, for isolated guanine, the N3–C2–N2–H and N1–C2–N2–H amino group dihedral angles are –31.4 and 11.9°, respectively. When extending the system to $dGMP^{-}$, the two dihedrals change to –38.9 and 10.1°, and when considering guanosine monophosphate dianion (a case not relevant to DNA already) even to –46.7 and 8.9°, respectively. The corresponding sums of the amino group valence angles change for the three structures in the following way: neutral G 346.3°, $dGMP^{-}$ 341.8° and $dGMP(-H)^{2-}$ 335.6°. Nevertheless, as the condensed phase situation can have a substantial effect on these charge-related effects, more calculations are necessary to prove whether the backbone has any substantial effect on the amino group pyramidalization effects in extended and solvated systems.

Protonation energies and dipole moments of nucleic acid bases. The quantum-chemical calculations can also be utilized to study other important monomer properties of nucleobases. The charge distribution, dipole moments and polarizability (Table V) and protonation energies (Table VI) are espe-

TABLE V

Electrostatic potential fitted atom-centered point charges (in a.u.) of nucleobases obtained by the MP2 method and extended aug-cc-pVDZ basis set for planar nucleobases. Dipole moments (μ, D; MP2/aug-cc-pVDZ method), polarizabilities (α, a.u.; values in parentheses show the vertical component of polarizability (α_{zz}); Becke3LYP/aug-cc-pVDZ method)[a]

Cytosine	C2 +0.967, N1 –0.634, C6 +0.238, C5 –0.740, C4 +1.055, N3 –0.817, O2 –0.619, N4 –1.100, H1 0.359, H6 0.138, H5 0.244, H41 0.451, H42 0.458, μ 6.39, α (α_{zz}) 103 (45)
Adenine	N3 –0.715, C2 0.465, N1 –0.694, C6 0.603, C5 0.128, C4 0.591, N9 –0.568, C8 0.243, N7 –0.578, N6 –0.897, H2 0.071, H8 0.129, H9 0.402, H61 0.421, H62 0.399, μ 2.56, α (α_{zz}) 122 (53)
Guanine	O6 –0.500, C6 0.524, N1 –0.710, C2 0.856, N3 –0.713, C4 0.480, C5 0.194, N7 –0.593, C8 0.302, N9 –0.602, N2 –1.053, H1 0.393, H8 0.105, H9 0.415, H21 0.458, H22 0.444, μ 6.55, α (α_{zz}) 119 (55)
Uracil	O4 –0.555, C4 0.817, N3 –0.609, C2 0.749, N1 –0.505, C6 0.133, C5 –0.554, O2 –0.554, H5 0.222, H3 0.349, H1 0.351, H6 0.156, μ 4.37, α (α_{zz}) 94 (41)
Thymine	O4 –0.531, C4 0.683, N3 –0.618, C2 0.744, N1 –0.520, C6 –0.043, C5 –0.041, O2 –0.559, C5M –0.474, H3 0.361, H1 0.364, H6 0.186, HM1 0.1575, HM2 0.133, HM3 0.1575, μ 4.31 D, α (α_{zz}) 112 (50)
6-Oxopurine	O6 –0.514, C6 0.601, N1 –0.644, C2 0.334, N3 –0.639, C4 0.550, C5 0.126, N7 –0.548, C8 0.251, N9 –0.549, H2 0.123, H1 0.383, H8 0.126, H9 0.400, μ 5.16, α (α_{zz}) 119 (50)
2-Aminoadenine	N6 –0.877, C6 0.455, N1 –0.663, C2 0.851, N3 –0.731, C4 0.487, C5 0.232, N7 –0.619, C8 0.279, N9 –0.601, N2 –1.011, H8 0.110, H9 0.412, H61 0.423, H62 0.399, H21 0.433, H22 0.421, μ 0.91, α (α_{zz}) 131 (60)
6-Thioguanine	S6 –0.272, C6 –0.129, N1 –0.295, C2 0.766, N3 –0.675, C4 0.480, C5 0.411, N7 –0.585, C8 0.241, N9 –0.587, N2 –1.081, H1 0.251, H8 0.130, H9 0.410, H21 0.465, H22 0.470, μ 7.28, α (α_{zz}) 169 (69)
Purine	N3 –0.731, C2 0.518, N1 –0.658, C6 0.231, C5 0.106, C4 0.770, N9 –0.709, C8 0.361, N7 –0.605, H6 0.114, H2 0.061, H8 0.112, H9 0.430, μ 3.75 α (α_{zz}) 115 (48)
2-Thiouracil	O4 –0.517, C4 0.543, N3 0.016, C2 –0.145, N1 0.033, C6 –0.125, C5 –0.313, S2 –0.276, H5 0.184, H3 0.194, H1 0.218, H6 0.188, μ 4.58, α (α_{zz}) 95 (54)
N3-Protonated cytosine	C2 0.699, N1 –0.455, C6 0.213, C5 –0.551, C4 0.816, N3 –0.532, O2 –0.430, N4 –0.933, H1 0.385, H6 0.188, H5 0.260, H41 0.482, H42 0.478, H3 0.380

[a] The amino hydrogens: cytosine H41 is cis to C5, adenine H61 is cis to N1, guanine and thioguanine H21 are cis to N3, 2-aminoadenine H21 is cis to N3, and H61 is cis to N1. Thymine HM3 hydrogen is in the plane of the base and points to O4.

cially relevant to assess the molecular interactions. Regarding the protonation energies, it is important to know that ring nitrogens have always higher proton affinities compared to the nucleobase amino nitrogens, although the amino groups themselves would be very good H acceptors (Table VI). This explains the lack of protonation of the nucleobase amino groups in solution, in contrast to aniline where there are no competing ring nitrogens suitable for protonation. The basicity argument (lack of amino group protonation) used in the past in bio-inorganic literature to rule out the amino group pyramidalization is thus completely unjustified. Interestingly, after any ring nitrogen is protonated, the nucleobase becomes planar due to electronic structure changes.

2.3. Structures and Energies of H-Bonded DNA Base Pairs

Large attention has been paid in recent years to obtain reliable data on structures and energies of H-bonded base pairs. Our initial data were based on HF optimizations of dozens of H-bonded base pairs at the HF/6-31G** level, with a subsequent evaluation of interaction energies at the MP2 level with similar basis sets[29] (Table VII). Besides standard and mismatched base pairs, we studied protonated base pairs[30], base pairs containing modified thio bases[31a] (Table VIII), trimers of DNA bases[31b] (Table IX), and others. In

TABLE VI
Protonation energies P (in kcal/mol) of selected nucleic acid bases evaluated at the MP2/6-31G** level

Cytosine	P_{N3} –241.4	P_{O2} –241.9[a]	P_{O2} –232.8[b]	P_{N4} –212.7	
Adenine	P_{N1} –234.8	P_{N3} –237.1	P_{N7} –228.6	P_{N6} –218.6	
Guanine	P_{N3} –223.5	P_{N7} –239.8	P_{O6} –233.5[c]	P_{O6} –224.3[d]	P_{N2} –205.3
Thymine	P_{O4} –217.5	P_{O4} –214.9[a]	P_{O2} –211.1	P_{O2} –209.8[a]	
Uracil	P_{O4} –216.4[b]	P_{O4} –213.4[a]	P_{O2} –208.0[b]	P_{O2} –206.6[a]	
6-Oxopurine	P_{N3} –217.7	P_{N7} –232.8	P_{O6} –228.3[c]	P_{O6} –220.1	
2-Aminoadenine	P_{N1} –238.8	P_{N3} –239.2	P_{N7} –235.2	P_{N6} –223.3	P_{N2} –230.8
6-Thioguanine	P_{N3} –221.9	P_{N7} –240.1	P_{S6} –235.1[a]	P_{S6} 231.3[b]	P_{N2} –203.8
2-Thiouracil	P_{O4} –215.5[b]	P_{O4} –212.5[a]	P_{S2} 210.9[b]	P_{S2} –210.3[a]	
Purine	P_{N1} –230.9	P_{N3} –219.4	P_{N7} –225.2		

[a] cis with respect to N3. [b] trans with respect the N3. [c] trans with respect to N1. [d] cis with respect to N1.

order to verify the reliability of calculations, we have re-optimized several base pairs at the MP2 level[32] and we have carried out a few optimizations of base pairs with an explicit inclusion of the basis set superposition error (BSSE) correction in the course of optimization[1b]. These reference calculations show that the HF/6-31G** optimizations provide a reasonable accuracy for the geometries, especially for a subsequent evaluation of interaction energies *via* single point calculations. Concerning the BSSE problem, our experience shows that, for optimized base-pair structures, BSSE correction is of the order of 3 kcal/mol at the HF and DFT levels with medium-sized basis sets of atomic orbitals. On the other hand, BSSE increases by additional *ca* 3 kcal/mol when the MP2 method is applied. This means that MP2 optimizations of base pairs are substantially influenced by BSSE. Such

TABLE VII

The interaction energies (in kcal/mol) of planar DNA base pairs (ΔE^{HF}, HF interaction energy; ΔE^{MP2}, interaction energy after adding the electron correlation contribution; ΔE^{T}, total interaction energy after adding the monomer deformation energies; $\Delta E^{T,est}$, estimated total interaction energy[a])

Pair[b]	ΔE^{HF}	ΔE^{MP2}	ΔE^{T}	$\Delta E^{T,est}$	Pair[b]	ΔE^{HF}	ΔE^{MP2}	ΔE^{T}	$\Delta E^{T,est}$
G-C WC	–24.6	–25.8	–23.8	–27.3	A-T WC	–9.7	–12.4	–11.8	–15.3
G-G1	–25.1	–24.7	–22.2	–25.7	G-G3	–16.0	–17.8	–17.0	–20.5
G-G4	–6.5	–10.0	–9.3	–12.8	G-A1	–12.2	–15.2	–14.1	–17.6
G-A2	–6.8	–10.3	–9.6	–13.1	G-A3	–10.8	–13.8	–13.1	–16.6
G-A4	–7.9	–11.4	–10.7	–14.2	A-T RH	–10.3	–13.2	–12.6	–16.1
A-T RWC	–9.6	–12.4	–11.7	–15.2	A-T H	–10.4	–13.3	–12.7	–16.2
C-C	–16.1	–18.8	–17.5	–21.0	G-C1	–11.6	–14.3	–13.4	–16.9
A-A1	–7.8	–11.5	–11.0	–14.5	A-A2	–7.2	–11.0	–10.3	–13.8
A-A3	–6.2	–9.8	–9.2	–12.7	G-T1	–14.2	–15.1	–13.9	–17.4
G-T2	–13.8	–14.7	–13.5	–17.0	T-C1	–8.7	–11.4	–10.7	–14.2
T-C2	–8.9	–11.6	–10.7	–14.2	T-T1	–9.3	–10.6	–10.0	–13.5
T-T2	–9.3	–10.6	–10.0	–13.5	T-T3	–9.3	–10.6	–9.9	–13.4
A-C1	–10.8	–14.3	–13.5	–17.0	A-C2	–10.4	–14.1	–13.2	–16.7

[a] Based on our continuing reference calculations on a smaller sample of base pairs and model systems. The $\Delta E^{T,est}$ value differs from the ΔE^{T} ones by –3.5 kcal/mol and is presently the most accurate estimate of base-pairing energetics to be used in force field verifications. [b] WC, Watson–Crick; H, Hoogsteen; R, reverse. Numbering of noncanonical base pairs according to ref.[29]

optimizations do not necessarily bring more accurate results compared with the HF optimizations. MP2 optimizations likely provide too short inter-base distances. The proper treatment requires the use of the BSSE in the gradient optimization and only now both geometry and stabilization energy are evaluated at a comparable level. These calculations are very tedious not

TABLE VIII

The interaction energies of planar modified DNA base pairs (ΔE^{HF}, HF interaction energy; ΔE^{MP2}, interaction energy after adding the electron correlation contribution; ΔE^{T}, total interaction energy after adding the monomer deformation energies; $\Delta E^{T,est}$, estimated total interaction energy) (*cf.* Table VII)[65]

Pair[a]	ΔE^{HF}	ΔE^{MP2}	ΔE^{T}	$\Delta E^{T,est}$	Pair[a]	ΔE^{HF}	ΔE^{MP2}	ΔE^{T}	$\Delta E^{T,est}$
^{6S}G-C WC	–23.1	–25.0	–22.5	–26.0	D-T WC	–11.9	–15.1	–14.1	–17.6
I.C-WC	–18.5	–19.4	–18.0	–21.5	A-^{4S}U WC	–8.4	–11.8	–11.1	–14.6
A-^{2S}U WC	–9.6	–12.8	–12.1	–15.6	^{2S}U-^{2S}U2	–8.7	–10.2	–9.6	–13.1
^{2S}U-^{2S}U1	–6.9	–9.3	–8.8	–12.3	^{6S}G-^{6S}G1	–19.3	–22.3	–19.9	–23.4
C-CH^{+}	–43.2	–44.8	–41.7	–45.2	A-F WC	–2.7	–3.9	–3.8	

[a] WC, Watson–Crick; H, Hoogsteen; R, reverse; D, 2-aminoadenine; I, 6-oxopurine (inosine); ^{6S}G, 6-thioguanine; ^{2S}U and ^{4S}U, 2- and 4-thiouracils; F, difluorotoluene; C-CH^{+}, triply-bonded protonated base pair (as seen in i-DNA)[30,31a].

TABLE IX

DNA base triplets: Total stabilization energies, individual pairwise base–base interaction energies, many-body contribution and deformation energy of bases[a] (all energies are in kcal/mol)

A.B-C[b] triplet	Total stabilization energy	Interaction between A and B	Interaction between B and C	Interaction between A and C	Many-body term (cooperativity)	Deformation energy of bases
G.G-C RH	–40.0	–18.2	–26.3	+1.0	–0.7	+4.2
G.G-C H	–44.6	–14.2	–26.2	–5.9	–3.8	+5.5
A.A-T RH	–21.7	–10.9	–12.3	+0.2	+0.1	+1.2
T.A-T H	–23.8	–13.3	–12.3	+0.6	0.0	+1.2
CH+.G-C	–65.5	–43.6	–26.1	+0.2	–3.5	+5.7

[a] The geometries were obtained *via* gradient optimization within the Hartree–Fock approximation using the 6-31G* basis set. The interaction energies were evaluated by the MP2 method[31b]. [b] Watson–Crick base pairing is between B and C.

only because five gradients instead of one (in the standard optimization) are required but mainly due to very slow convergence. Another very important advantage of counterpoise-corrected gradient optimizations is the fact that stabilization energy is defined as a mere difference between the energy of a complex and the energies of both subsystems, *i.e.* the concept of deformation energy is not considered.

Nonplanar base pairs. One of the major outcomes of the studies is the observation that many noncanonical base pairs, mainly all G/A mismatches and many other purine/purine base pairs, are intrinsically nonplanar. We suggest that in the case of G/A mismatches their large degree of flexibility is essential for the molecular recognition processes[24]. The studies show that gradient-optimized geometries and energies are superior to those obtained with rigid-monomer searches, in a sharp contrast to a recent misleading claim by Cybulski and co-workers[33]. This is because the monomers, upon formation of the H-bonded complexes, are substantially deformed and especially in the case of strong base pairs there is a substantial elongation of the X–H bonds in the X–H···Y interactions. With rigid monomers (suggested by Cybulski), this important feature inherent to the H-bonding is lost. This does not necessarily disqualify the use of rigid monomer approaches; however, they do not represent any step forward even though they allow easy correction for the basis set superposition error. Rigid monomer approaches of course fail for intrinsically nonplanar base pairs.

C–H···O hydrogen bonds. We have also analyzed the nature of C–H···O contacts occurring in DNA and RNA base pairs[32b] including MP2 vibrational analyses of base pairs accompanied by NBO and electron topology evaluations. These studies unambiguously show that the C–H···O contact in Watson–Crick A-T and A-U base pairs is entirely inactive and it is misleading to speculate that it is a C–H···O H-bond. On the other hand, a weak C–H···O bond, including a clear red shift, has been identified in certain uracil–uracil base pairs[32b].

Planarization of the amino group upon base pairing. An interesting problem is associated with planarization of amino group upon dimerization. It was observed that guanine amino N–H stretching frequency is blue-shifted upon formation of various guanine-containing DNA base pairs. Usually, the red shift of an X–H stretch frequency is observed and gives evidence about the formation of the XH···Y H-bonding. The blue shift in the guanine dimer was explained by planarization of both amino groups leading to contraction of N–H bonds and, consequently, to the blue shift of the respective stretching vibration frequencies[32d]. Planarization of both amino groups requires the energy but this is obtained from better conjugation when amino

groups are planar. The natural bond orbital (NBO) analysis reveals that electron density at the amino group nitrogen decreases upon dimerization (due to better conjugation), which leads to change of hybridization (from the sp^3 to sp^2).

Reliability of different QM methods and force fields. Accurate *ab initio* calculations of base pairing are still quite expensive. Therefore, there is a large interest in cheaper techniques and we tested a number of them. Hydrogen-bonded base pairs can be well studied by nonlocal DFT approaches with Becke3LYP energetics closely following the MP2 values with medium-size basis sets of atomic orbitals[29]. Regarding DFT optimizations of base pairs, the initial experience indicated that Becke3LYP method leads to too short interbase H-bonds (even shorter than those obtained from BSSE-uncorrected MP2 optimizations) and the intramolecular deformation of monomers is overestimated[29]. However, in view of the very recent RI-MP2 re-evaluations with large basis sets (see below), the DFT data appear to be quite good. More serious problem is that DFT-based methods do not properly include the intermolecular electron correlation effects. We have recently tested the PW91PW91 DFT method (unpublished data) and while it somewhat improves the H-bonding energies it remains to be entirely deficient for base stacking. DFT methods do not allow a systematic improvement of the interaction energy with increasing the size of the basis set[34]; thus, these techniques cannot replace conventional *ab initio* methods in reference calculations. Good performance of local MP2 has been suggested[35] but a systematic test of this technique is absent and our unpublished data are so far unsatisfactory. Comparison with other techniques shows a good performance of several recent empirical force fields including AMBER [36a]. However, semiempirical techniques do not provide sufficient accuracy for H-bonded DNA base pairs[36a] and should be avoided.

New reference calculations of base pairing. We have later verified[36b–36d] the above calculations using a more accurate technique (Table X). From the Table it is evident that extending the basis set over the DZ + P level is essential and brings an enlargement of stabilization. Further extension of the basis set by inclusion the of second polarization functions (*f*-functions on heavy atoms and *d*-functions on hydrogens) is still important and clearly indicates that passing to a complete basis set limit yields non-negligible changes of stabilization energies. Stabilization energies determined with extended aug-cc-pVQZ basis set are finally close to the basis set limit[36b,36d]. From the Table it is further evident that extrapolation of aug-cc-pVDZ and aug-cc-pVTZ data yields similar results to those obtained by much more expensive extrapolation of aug-cc-pVTZ and aug-cc-pVQZ data. This finding is

important for routine calculations. Investigating the role of higher correlation energy contributions (CCSD(T)) we found they are small and can be (in agreement with the original opinion) neglected for this class of DNA base pairs. The Table also gives support to our original theoretical level used for the evaluation of stabilization energies of DNA base pairs, namely the MP2/6-31G*(0.25)//HF-6-31G**. Evidently, this, from the present view, 6-31G*(0.25) basis set yields surprisingly good stabilization energies. However it must be added that this basis set is not suitable for geometry optimization since the basis set is unbalanced and resulting geometries are incorrect. This is clearly documented by very large deformation energies, being more than five times larger than those obtained with a balanced A-O basis set.

Evaluation of structures of H-bonded clusters. Let us shortly comment on the evaluation of geometry of molecular clusters. In the case of H-bonded DNA base pairs, we have originally used the HF/6-31G** geometries but then we demonstrated that inclusion of correlation energy affects the cluster geometry. There are two reasons for this. First, stabilization energy becomes larger due to inclusion of dispersion energy, and, second, the electrostatic energy is also affected since the correlated dipoles are by 15–20% smaller than the HF ones. Provided that contribution of correlation energy is modest, the

TABLE X

Interaction energies (in kcal/mol) of H-bonded guanine···cytosine and adenine···thymine base pairs in the Watson–Crick structures calculated at various theoretical levels for RI-MP2/TZVPP (identical to MP2/cc-pVTZ level) geometry (CBS, complete basis set extrapolations)

Method/Basis set	G-C WC	A-T WC
MP2/cc-pVDZ(0.25,0.15)	–26.2	
MP2/aug-cc-pVDZ	–28.7	–14.8
MP2/aug-cc-pVTZ	–30.4	–16.0
MP2/aug-cc-pVQZ	–31.1	
MP2/CBS limit (D-T)[a]	–31.2	–16.6
MP2/CBS limit (T-Q)[b]	–31.2	
CCSD(T)/cc-pVDZ(0.25,0.15)	–26.6	
E_{def}	3.4	1.5
Total	–28.3	–15.0

[a] Extrapolation to the MP2/complete basis set limit from MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ calculations. [b] Extrapolation to the MP2/complete basis set limit from MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ calculations.

MP2/6-31G** geometries are sufficiently accurate. When, however, the role of dispersion energy becomes larger, then extension of A-O basis set is inevitable. We can document this on the geometry of phenol dimer[36c] for which the rotational constant is available (it is very rare to have experimental geometries for more extended complexes). The absolute average difference of *A*, *B*, and *C* constants (with respect to experiment) was 4.5%, when the MP/6-31G* calculations were performed, and dropped to 1.5% when the RI-MP2/TZVPP [5s3p2d1f/3s2p1d] calculations were performed. The level used (equivalent to the MP2/cc-pVTZ method) is the first reliable level for obtaining accurate geometries and also stabilization energies of extended complexes including the DNA base pairs. Let us add that the level is acceptable for such extended systems only because the RI-MP2 method is at least one order of magnitude faster than the exact MP2 one.

2.4. Nature of Aromatic Base Stacking

One of the major results furnished by quantum-chemical studies on nucleic acids was clarification of the nature of base stacking. Proper evaluation of base stacking is more difficult compared with H-bonding of nucleobases. The calculations must be carried out with inclusion of electron correlation effects. The Hartree–Fock method, which neglects all electron correlation effects, cannot be used for the purpose. The main reason is the occurrence of dispersion energy as a result of correlation of electron motions. Second, it is imperative to use a basis set of atomic orbitals which contains at least one set of diffuse d-atomic orbitals for second-row elements. The diffuse d-atomic orbitals reach sufficiently far from the atomic nuclei and hence they can fill the empty space between two stacked bases where a substantial portion of the dispersion energy originates. Diffuse sp shells (like in the 6-31+G basis set) are not sufficient to improve the interaction substantially. We have mostly used the basis set designated as 6-31G*(0.25) [37], which is a variant of the very popular Pople 6-31G* basis set. However, the exponents of the d-atomic orbitals are modified to a value of 0.25 instead of the standard value 0.8, resulting in a large difference for stacking[1]. The standard 6-31G* basis set is very deficient for base stacking and the stacking of two bases would be underestimated by as much as 40–50% if that basis set is used[1b]. The 6-31G* basis set thus is to be avoided in stacking calculations. The only exception is gradient optimization of stacked complexes at the MP2 level, as here the diffuse functions of 6-31G*(0.25) could lead to imbalanced intramolecular tems[32a]. Further, the stacked dimers are in the course of MP2 gradient optimizations artificially overstabilized by the basis

set superposition error, even with 6-31G* basis set (see below). Thus, the dispersion-deficient basis set compensates for the BSSE artifact. On the other hand, single point stacking energies evaluated with the 6-31G*(0.25) basis set are reasonable close to values obtained with the aug-cc-pVDZ basis set[34]. The stacking energies must in any case be corrected for an artifact that is known as the "basis set extension effect" or "basis set superposition error" (BSSE; see also above, the discussion for H-bonding) using the standard full counterpoise procedure[38]. The basis set extension effect is very large for any stacked cluster, larger than for H-bonded pairs, and BSSE-uncorrected calculations are of little value. Recent studies convincingly confirm that the full counterpoise procedure is an exact solution to eliminate the basis set extension effect independent of its magnitude[39]. Note that for a stacked cytosine dimer the BSSE can amount up to 10 kcal/mol (mostly affecting the MP2 component of stacking), being as large as the actual stabilisation of the dimer in its global minimum.

Aromatic base stacking does not fundamentally differ from non-aromatic stacking. Let us now briefly outline the physical origin and magnitude of the gas phase base stacking interaction energy as finally revealed by *ab initio* calculations with inclusion of electron correlation[1]. Base stacking is primarily determined by three contributions: dispersion attraction, short-range repulsion, and electrostatic interaction. No specific π–π interactions have been evidenced. The stabilization of base stacking is dominated by the dispersion attraction, which is rather isotropic and proportional to the geometrical overlap of the bases. The distance between stacked bases and base pairs is invariably around 3.3 Å, being determined by the balance between dispersion attraction and short-range repulsion present between the adjacent nucleobases. Finally, the mutual orientation of bases and their displacement are primarily determined by the electrostatic attraction, that is by the interaction between the electric fields of the two monomers with dominating molecular dipole–molecular dipole interaction[1]. Of course, this picture of base stacking will be different when solvent effects are considered because the response of the polar solvent to the electric field of the bases would lead to a destabilization of the most stable gas phase arrangements[13]. However, this is common to all molecular interactions and does not indicate any substantial change of the physical origin of the base–base interaction itself. Due to the quality of the calculations, which has been verified on numerous other systems, we believe that this description of stacking is ultimate and will not qualitatively change by further improvements in the quality of quantum-mechanical calculations with the advance of more sophisticated hardware and software in future.

Accuracy of other QM and force field approaches. Comparison with other methods shows a complete failure of semiempirical methods for base stacking. Semiempirical methods ignore the dispersion attraction and fail to provide even a good angular dependence of electrostatic part of base stacking[36a]. DFT methods are capable of providing good angular dependence of stacking, however, concerning the stability of stacking these methods fail as they basically ignore the dispersion energy[40]. DFT methods give for vertical separation to nucleic acid bases only repulsive curves while correlated *ab initio* techniques yield a deep energy minimum; for example, in the case of guanine dimer of more than 10 kcal/mol. Some work has been recently reported to attempt to deal with this limitation[41]. Presently, however, the most promising way appears to be to combine the DFT method with an empirical dispersion energy[32c,40]. Recently we introduced the first DFT technique successfully covering the London dispersion energy[32c]. This was achieved by adding simple atom–atom London dispersion energy term proportional to the sixth power of the reciprocal distance. As the DFT technique is based on self-consistent-charge, density-functional tight-binding method, it is suitable for calculation of extended DNA base clusters containing several dozens of atoms. The pilot calculations showed that the method describes H-bonding and stacking of DNA bases reasonably well[32c].

The use of local the MP2 technique for aromatic stacking is not advised[34b]. On the other, hand recently released empirical potentials for biomolecular modeling show a very reasonable accuracy for base stacking. A more detailed comparison can be found in literature[36a]. To our surprise, we did not obtain a correct description of electrostatic part of stacking when utilizing the distributed multipole analysis (DMA)[40]. Unfortunately, so far this result has not been commented or noticed by any group working and/or developing DMA-based potentials while no other group tested DMA for aromatic stacking.

Stacking of protonated and modified bases. We have also studied stacking of protonated bases[30] and stacking of thio bases[31a]. In the case of protonated bases one has to consider polarization effects[30] while studies of thio bases clearly demonstrated the anisotropic nature of short-range repulsion. None of these contributions is included in current force fields[30,31a]. The anisotropic short-range repulsion means that the atoms do not always look spherical. Interestingly, anisotropy of the short range repulsion has been found also in some parts of the potential energy surface of a stacked cytosine dimer and certainly exists also for other stacks[40a]. We have also characterized another interaction occurring in DNA, the sugar base stacking[8a]. The calculations convincingly demonstrated that this interaction should be

classified as a dispersion-controlled contact, so its physical nature is close to regular base stacking.

Base stacking in DNA geometries. Besides mapping the gas phase conformational space of stacked dimers, we have also studied stacking arrangements observed in nucleic acids (Figs 7 and 8). When evaluating stacking in DNA geometries, we established the following principles: (i) The sequence-dependent variability of stacking energy along the double helix is within the range of only a few kcal/mol [8]. (ii) Individual base-pair steps show large

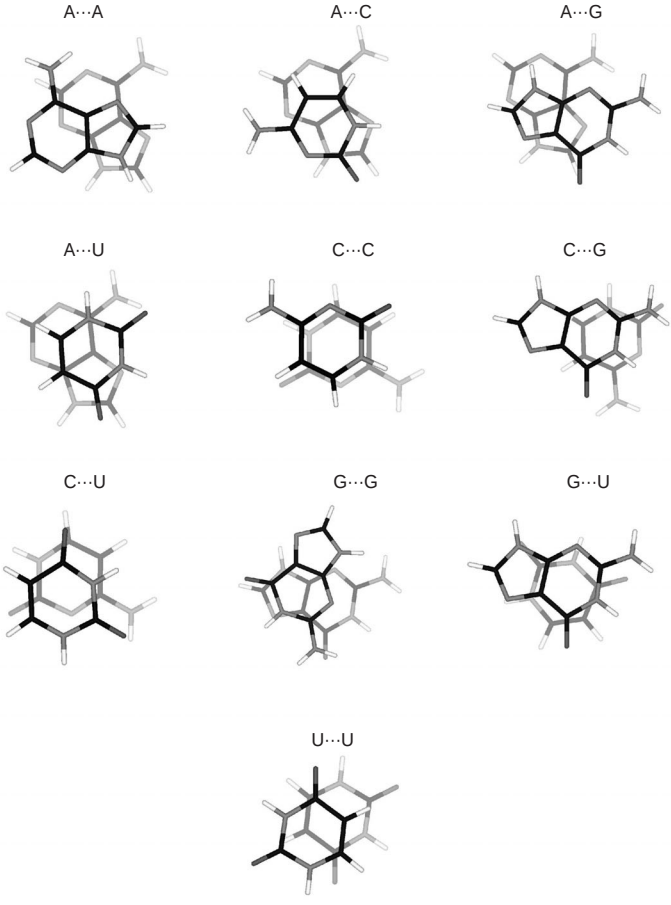

FIG. 7

Optimal base dimer face to back stacking geometries, as predicted *via* search with MP2-adjusted force field. The corresponding MP2/6-31*(0.25) stacking energies (in kcal/mol) are: AA –8.8, AC –9.5, AG –11.2, AU –9.1, CC –8.3, CG –9.3, CU –8.5, GG –11.3, GU –9.5, UU –6.5 [1a]

variabilities in interstrand and intrastrand contributions to stacking, mutually compensating. (iii) Also, large variations and mutual compensations are observed for dispersion and electrostatic contributions to stacking.

Nonadditivity of base stacking. We have recently estimated the base stacking nonadditivity and evaluated the stacking energy in *ca* 30 base-pair step geometries in two ways[8]. First, as the sum of two intrastrand and two interstrand base–base contributions (*i.e.*, assuming additivity), and then as the interaction between two base pairs. The difference between the two evaluations is the nonadditivity of base stacking, that is the four-body term in this particular case (Eq. (*1*), Fig. 9; symbol A-B|C-D stands for stacking between base pairs A-B and C-D).

$$\Delta E^{\text{A-B|C-D}} = \Delta E^{\text{A|C}} + \Delta E^{\text{A|D}} + \Delta E^{\text{B|C}} + \Delta E^{\text{B|D}} + \Delta E^{4} \qquad (1)$$

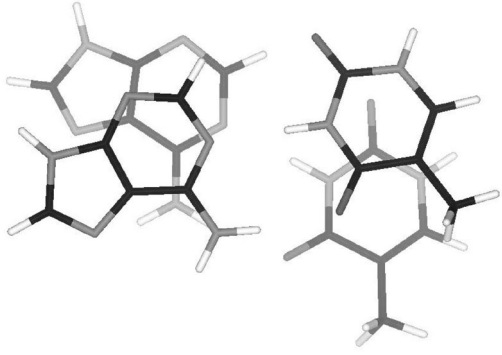

FIG. 8
Typical base-pair stacking in B-DNA. The ApA(TpT) step. Base stacking energy in such a large system can these days be routinely evaluated within 10 days on a single 2GHz Pentium processor with a high degree of accuracy (for example, using the RI-MP2/aug-cc-pVDZ method)

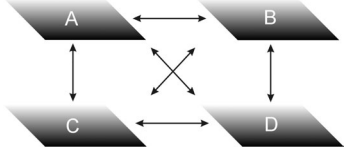

FIG. 9
Nonadditivity of base stacking. The base stacking between two DNA base pairs can be evaluated as a sum of four base–base stacking energy terms AC, BD, AD and BC and a four-body correction

Base stacking appears to be additive when considering stacking of consecutive A-T base pairs. However, the nonadditivity of stacking can reach a value of *ca* 2–3 kcal/mol in base-pair steps consisting of G-C base pairs. This is a significant portion of the base-pair stacking energy. The electronic structure of base pairs is influenced by the environment and the bases respond to the external electric field by polarization. This can be particularly well documented for the G-C base pair. Its electronic structure is different depending on whether it is surrounded by basically nonpolar A-T base pairs, or stacked between very polar G-C base pairs.

Local conformational variations in DNA do not improve base stacking. After careful analysis of base stacking in several high-resolution crystals, we have recently concluded, in contrast to widespread earlier opinions, that local conformational variations observed in DNA crystals do not improve the intrinsic base stacking energy terms[8b]. The stacked base-pair steps have wide low-energy regions and in the crystals any of these geometries can actually occur, depending on "external" factors including crystal packing[8b]. This view is perhaps best documented in our recent study on a unique high-resolution B-to-A DNA duplex intermediate containing a central G-tract capped with flexible CpA segments (Tables XI–XIII). Even in this molecule no improvement of stacking associated with local conformational variations is seen. The study also indicates that the magnitude of stacking in B- and A-DNA forms is similar. This is in contrast to a literature suggestion of a poor stacking in the A-form[42], which was likely caused by utilizing fiber diffraction geometries for evaluation of stacking. As the A-DNA fiber model contains vertically compressed steps (an error frequently occurring for fiber structural models lacking atomic resolution), the calculated A-form stacking energies are biased.

Base stacking energies are extremely sensitive to small vertical compression or extensions of the interacting extended planar systems. Interestingly, it has been found that one base-pair step of the A-to-B intermediate crystal is deformed by crystal packing forces and actually for this base-pair step the computations revealed a modest deterioration of the stacking energy[8b].

Towards final values of base stacking. Our calculations of base stacking are mostly based on systematic potential energy surface searches carried out *via* a series of single points at the MP2/6-31G*(0.25) level. Reference MP2 calculations with larger basis sets supplemented with CCSD(T) correction have been carried out first for model aromatic stacking complexes[34a] and then for stacked nucleobase dimers with inclusion of a complete basis set (CBS) extrapolation with the MP2 method[34b]. These calculations indicate that the MP2/6-31G*(0.25) data are close to the actual (unknown) values, since the

MP2 method moderately overestimates the stabilization originating in intermolecular electron correlation effects, compared with the CCSD(T) method. This compensates for the limited size of the basis set. From Table XIV it is first evident that all stabilization comes from electron correlation with repulsive interaction energies evaluated at the HF level. Further, it is evident that stacking energy is more dependent on the quality of the A-O basis set than H-bonding energy. Passing from the 6-31G* level to higher levels leads to with dramatic stacking energy improvement which unambig-

TABLE XI

Base-pair step stacking energies $\Delta E^{A\text{-}B|C\text{-}D}$ (in kcal/mol; symbol | stands for stacking, *cf.* Fig. 9) evaluated at the MP2/6-31G*(0.25) level along the high-resolution d(CATGGGCCCATG)$_2$ crystal structure of a unique A-B intermediate duplex[8b], standard base-pair step geometries, and some other high-resolution crystal geometries. The values in parentheses show the HF component of the interaction energies

Base-pair step	A-B crystal[a]	Standard B-DNA geometry	Other B-DNA crystals[a]
C1-G24|A2-T23	–11.9 (+8.0)	–12.1 (+4.2)	–11.7 (+1.1)[b]
A2-T23|T3-A22	–12.6 (+8.0)	–10.7 (+11.6)	
T3-A22|G4-C21	–11.6 (+9.9)	–12.1 (+4.2)	
G4-C21|G5-C20	–7.8 (+13.4)	–9.5 (+12.2)	–8.9 (+12.9)[c]
G5-C20|G6-C19	–8.4 (+13.8)	–9.5 (+12.2)	
G6-C19|C7-G18	–14.0 (+5.2)	–13.2 (+8.5)	–11.5 (+8.5)[d]
C7-G18|C8-G17	–8.4 (+11.6)	–9.5 (+12.2)	
C8-G17|C9-G16	–7.6 (+12.2)	–9.5 (+12.2)	
C9-G16|A10-T15	–11.5 (+10.1)	–12.1 (+4.2)	
A10-T15|T11-A14	–12.7 (+9.5)	–10.7 (+11.6)	
T11-A14|G12-C13	–9.0 (+7.8)	–12.1 (+4.2)	
Total 11 steps	–115.5[e]	–121.0	
total C-G|C-G steps	–32.2	–38.0	

[a] The base-pair steps were constructed with *ab initio* optimized geometries of base pairs by shifting the bases into proper intermolecular positions using distances and angular parameters taken from the crystal. Thus the crystal bases are overlaid by QM-optimized ones. Note that the monomer geometries directly taken from the PDB X-ray files are unrelaxed and thus not suitable for accurate calculations. [b] 5′-C2A3-3′(C2pA3) step of d(CCAAGATTGG)$_2$ decamer 1.5 Å crystal structure, high twist – high slide – low roll geometry. [c] G4pG5 step of d(CCAGGCCTGG)$_2$ decamer 1.6 Å resolution crystal structure. [d] G5pC6 of step d(CCAGGCCTGG)$_2$. [e] The crystal stacking geometries do not bring any systematic improvement compared to the standard geometries[8b].

TABLE XII
Decomposition of selected base-pair step stacking energies into the individual terms, the same structures as in Table XI

A-B|C-D	$\Delta E^{A\|C}$	$\Delta E^{B\|D}$	$\Delta E^{A\|D}$	$\Delta E^{B\|C}$	ΔE^{4}	$\Delta E^{A-B\|C-D}$
A2-T23|T3 A22	–6.4	–6.5	+0.1	–0.4	+0.6	–12.6
T3-A22|G4-C21	–2.7	–4.4	–0.5	–5.0	+1.1	–11.6
G4-C21|G5-C20	–2.6	–1.7	–3.4	–2.5	+2.5	–7.8
G5-C20|G6-C19	–2.1	–1.7	–3.2	–3.8	+2.4	–8.4
G6-C19|C7-G18	–9.6	–9.8	+2.1	+2.8	+0.6	–14.0
T11-A14|G12-C13	–2.3	–4.3	–0.5	–3.1	+1.2	–9.0
T-A|G-C[a]	–5.4	–2.1	–1.7	–2.6	+0.1	–11.7
G-C|G-C[b]	–3.5	–1.1	–4.1	–2.8	+2.0	–9.5
G-C|C-G[b]	–9.0	–9.0	+1.4	+2.5	+0.9	–13.2
T-A|G-C[b]	–4.7	–3.8	–0.9	–3.5	+0.8	–12.1
A-T|T-A[b]	–5.1	–5.1	–0.8	+0.4	+0.0	–10.7

[a] High twist-high slide-low roll geometry[8b]. [b] Standard geometry.

TABLE XIII
Decomposition of the electrostatic stacking energy (in kcal/mol) terms into the individual base–base contributions (note that the many-body contribution to the Coulombic term is zero)

A-B|C-D	$E^{el,A\|C}$	$E^{el,B\|D}$	$E^{el,A\|D}$	$E^{el,B\|C}$	E^{el}			Overlap stacking[a]
					intra-strand	inter-strand	total	
A2-T23|T3-A22	–0.7	–0.8	+0.5	+1.5	–1.5	+2.0	+0.5	–13.7
T3-A22|G4-C21	+0.7	+0.1	–0.4	+0.9	+0.8	+0.5	+1.3	–14.0
G5-C20|G6-C19	+4.6	+2.0	–2.8	–0.3	+6.6	–3.1	+3.5	–14.3
G6-C19|C7-G18	–4.0	–3.1	+4.7	+3.6	–7.1	+8.3	+1.2	–15.8
T-A|G-C[b]	–1.7	+2.0	+0.3	–1.5	+0.3	–1.2	+0.9	–12.7
G-C|G-C[c]	+4.3	+3.5	–3.1	–1.3	+7.8	–4.4	+3.4	–14.9
G-C|C-G[c]	–2.9	–2.9	+4.1	+3.2	–5.8	+7.3	+1.4	–15.5
T-A|G-C[c]	+1.0	0.0	–0.4	–0.4	+1.0	–0.8	+0.2	–13.0
A-T|T-A[c]	+0.6	+0.6	+1.0	+0.8	+1.2	+1.8	+3.0	–13.7

[a] Overlap stacking energy shows an especially small variability with sequence and structure, and is calculated as the difference between total stacking and its electrostatic component. For more details see ref.[8b]. [b] High twist-high slide-low roll geometry. [c] Standard geometry.

TABLE XIV
Stacking interaction energies (in kcal/mol) of selected geometries of cytosine, uracil and guanine dimers, and guanine···cytosine stack (G|C)[34b]

Basis set	Dimer	HF	MP2	CCSD(T)	Δ|CCSD(T)-MP2|
6-31G*	cytosine	0.33	–5.27	–4.16	1.11
	uracil	0.99	–6.01	–4.75	1.26
	guanine	–0.64	–8.06		
6-31G*(0.25)	cytosine	–0.04	–8.27	–7.15	1.11
	uracil	0.14	–8.94	–8.01	0.93
	guanine	–0.87	–11.19	–9.34	1.85
6-31G*(0.25,0.15)	cytosine	0.18	–8.39	–7.39	1.00
	uracil	0.29	–9.01	–8.20	0.81
6-31G*(0.25) + bf1[a]	uracil	0.13	–9.35	–8.58	0.77
6-31G*(0.25) + bf2[a]	cytosine	–0.01	–8.78	–7.95	0.83
cc-pVDZ(0.25,0.15)	cytosine	0.14	–8.51	–7.55	0.96
	uracil	0.48	–8.98	–8.15	0.83
	G|C		–16.20	–14.3	1.8
6-31G**(0.25,0.15) + bf2[a]	cytosine	0.05	–8.97	–8.20	0.78
6-31++G**(0.25,0.15)	cytosine	–0.59	–9.75	–8.87	0.88
	uracil	0.19	–9.62	–8.94	0.69
aug-cc-pVDZ	cytosine	–0.02	–10.15		
	uracil	0.57	–10.46		
	guanine	–1.13	–13.22		
	G|C		–18.50		
aug-cc-pVDZ + bf1[a]	uracil	0.57	–10.87		
aug-cc-pVDZ + bf2[a]	uracil	0.54	–10.99		
aug-cc-pVTZ	cytosine	–0.06	–10.76		
	uracil	0.57	–11.52		
	G|C		–20.00		
aug-cc-pVTZ + bf2[a]	uracil	0.55	–11.68		

[a] bf indicates that bond functions were placed between stacked subsystems.

uously indicates the importance of extrapolation to the complete basis set limit. In the case of cytosine dimer we obtain more than a 100% stabilization energy change when passing from 6-31G* to aug-cc-pVTZ basis set. Further, the higher correlation energy contributions are significant and the CCSD(T) – MP2 difference can be anywhere between 0.7 to 1.9 kcal/mol. In the case of purine···pyrimidine stacked base pairs this contribution can be even larger and can reach about 4 kcal/mol (not shown). Table XIV shows that the CCSD(T) – MP2 difference is always positive, *i.e.* passing to CCSD(T), the stabilization is reduced. Contrary to H-bonded DNA base pairs, these differences are systematically significant and can never be neglected. Similar results are also indicated by other groups[34c]. This clearly means that obtaining reliable relative energies for H-bonding and stacking (which is of key importance for describing the stability and dynamics of DNA and RNA) requires the use of CCSD(T) level of calculations at least for the correction of the MP2/CBS data. All our data further suggests that the MP2/aug-cc-pVDZ stacking interaction energies are the most accurate low-cost method to estimate aromatic base stacking while the MP2/6-31G*(0.25) method still remains qualitatively correct[34b].

2.5. Potential Energy Surface (PES) and Free Energy Surface (FES) of DNA Base Pairs and Microhydrated DNA Base Pairs

The PES of DNA base pairs is very complex and contains a large number of energy minima[43]. The global minimum can be theoretically determined if the complete PES is known. The number of energy minima increases very rapidly with the cluster size. While there is just one minimum on the water dimer PES, there are 11 minima on the uracil dimer PES and more than 1000 minima on the adenine···thymine···$(water)_2$ PES. The localization of all minima is tedious, if not impossible, by standard methods based on experience and chemical intuition. It is necessary to use some effective searching technique and methods based on molecular dynamics (MD) simulations in combination with quenching (Q) technique are very promising. Performing the MD/Q calculations for a longer time, we can obtain in addition to localization of all the energy minima also their population. This population is proportional to the change of the Gibbs free energy of cluster formation.

The PESs and FESs of all 10 canonical and methylated nucleic acid base pairs were studied by MD/Q technique in combination with the Cornell *et al.*[43c] force field and by correlated *ab initio* calculations. More than a dozen energy minima were located on the PES of each base pair. The global

and first local minima of a nonmethylated base pairs have systematically a planar H-bonded structure, while T-shaped and stacked structures were less stable. The MD/Q search sometimes reveals an unexpected structure of the global energy minimum. For example the global minimum of the adenine···thymine base pair corresponds neither to Watson–Crick nor to Hoogsteen type of bonding[43d]. Entropy does not play an important role and relative order of individual structures on the PES and FES does not differ too much. However, methylations at purine N9 and pyrimidine N1 bring dramatic changes because now the entropy plays an important role. The structure of the global minimum does not usually correspond to the most populated structure; frequently, it is the stacked structure which is the most populated.

The role of microhydration. Microhydration plays an important role and changes the structure of a global minimum of a base pair[43e]. The presence of one water does not affect the structure of any H-bonded base pair. An equal population of H-bonded and stacked structures of adenine···adenine, adenine···guanine and adenine···thymine pairs is reached if as few as two water molecules are present, while obtaining equal population of these structures in the case of adenine···cytosine, cytosine···thymine, guanine··· guanine and guanine···thymine dimers required the presence of four water molecules, and in the case of guanine···cytosine pair even six water molecules. A comparable population of H-bonded and stacked structures for cytosine···cytosine and thymine···thymine base pairs was only obtained if at least eight water molecules hydrated the nucleobase dimer. The data give a clear evidence that the preferred stacked structure of DNA base pairs in a water solution is due to the hydrophilic interaction of a small number of water molecules, thus without hydrophobic interaction between bulk of water and the base pair.

2.6. Interactions of Nucleic Acid Bases and Base Pairs with Metal Cations

After clarifying the nature of molecular interactions in H-bonded and stacked base pairs, the focus of *ab initio* studies has shifted to investigations of interactions between nucleobases and metal cations. Properties of nucleic acids are significantly influenced by metal cations. With the increasing power of computers we are in a position to study the interactions of bases and base pairs using reliable quantum-chemical methods. These studies illustrate all advantages and weaknesses of the *ab initio* approach. Complexes with metal cations are characterized by very large nonadditivities of interactions[28a]. The total amount of these effects increases with the charge

of the cation and is substantially larger for transition metal elements than for alkaline metals. Therefore, to study such complexes one cannot use pair-additive force fields and a method considering the electronic structure of the complexes must be utilised. This gives a unique position to *ab initio* quantum chemistry. Nonadditivity of interactions is exceptionally important to properly account for the metal cation coordination and hydration, and to describe the balance of interactions of a given cation with various chemical groups in biopolymers. Quantum-chemical calculations are these days the basic tool to study the interactions of metal cations with biopolymers, or fragments of biopolymers. *Ab initio* techniques can be used to study important issues such as specific differences among cations.

On the other hand, metal cation-containing complexes are mostly non-neutral. For non-neutral ionic systems the gas phase limitations of the *ab initio* technique can create almost insurmountable problems concerning the comparison with the situation present in condensed phase. While the ionic electrostatic effects dominate in the gas phase, they are almost eliminated in polar solvents and in crystals. This can lead to a profound difference in the views adopted by computational and bio-inorganic chemists[28a,44]. Thus, it is essential to make quantum-chemical calculations for rather extended complexes[28a]. Inappropriate interpretation of gas phase data (theoretical as well as experimental) can easily lead to conclusions which bear no relevance to condensed phase situation and nucleic acids. For example, binding of divalent metals to the N7 position of guanine at first sight could induce a very high rate of a spontaneous formation of ion-pair structure of the G-C base pair, due to a substantial destabilization of the H1 guanine ring hydrogen. This would be a highly mutagenic process; however, proper consideration of environment shows that this process is basically always eliminated by environment[28a].

The effect of metal binding on base pairing. Extended *ab initio* calculations have been reported on metalated H-bonded base pairs[28a,44–47]. The calculations showed that direct (inner-shell) binding of a cation to the N7 atom of guanine significantly influences the strength of the guanine-containing base pairs and provides about 5–10 kcal/mol of additional polarization stabilization. Table XV shows that stabilization of the Watson–Crick G-C base pair is substantially affected by polarization of the aromatic system. The polarization is included in the three-body term ΔE^3. Even larger effect was found for the reverse Hoogsteen G/G base pair. The base-pair strength (interaction energy) of isolated reverse Hoogsteen G/G base pair is around –18 kcal/mol. When a hydrated divalent metal cation binds to the N7 position of guanine, the energy necessary to disrupt the base pair is twice as

large than without the cation[46c]. Approximately half of this metal-ion induced stabilization (cooperativity) is due to polarization effects, the rest stems from the electrostatic attraction between the hydrated cation and the nonmetalated guanine. Interestingly, the cation binding to N7 has no influence on the stability of the adenine–adenine reverse Hoogsteen base pair. Inclusion of the sugar-phosphate backbone into the calculations reduces the polarization strengthening by less than 50% [28a]. Strengthening of base pairing by cation binding has been recently confirmed by condensed phase experiments on platinated base pairs in DMSO [48]. Our recent unpublished data also show that N7 metal binding can substantially affect the base stacking energies.

Zinc vs magnesium difference. QM calculations rationalize the difference in binding of Zn^{2+} and Mg^{2+} cations in nucleic acids. Both cations have the same charge and approximately the same ionic radius. Thus, they would have very similar properties when treated by pair-additive empirical force fields. On the other hand, quantum-chemical calculations clearly show the difference between these two cations and also where this difference originates[28a,46]. Compared to magnesium, zinc has a much larger affinity to

TABLE XV

Interaction energies (in kcal/mol) in the solvated cation (M)–guanine (G)–cytosine (C) complexes (ΔE_{GM}, pairwise interaction energy between the guanine and the solvated cation; ΔE_{CM}, pairwise interaction energy between cytosine and the solvated cation; ΔE_{GC}, pairwise base-pair interaction energy; ΔE^3, the three-body term; ΔE^T, total interaction energy, *i.e.*, the sum of the previous contributions[a])

Cation	ΔE_{MG}	ΔE_{CM}	ΔE_{GC}	ΔE^3	ΔE^T
Mg^{2+}	–89.3(–198.7)	–1.5(–)	–26.4(–26.0)	–8.1(–)	–125.4(–243.8)
Ca^{2+}	–82.6(–133.9)	–1.7(–3.0)	–26.3(–25.8)	–5.2(–10.1)	–115.8(–172.7)
Sr^{2+}	–76.0(–)	–2.1(–)	–25.8(–)	–4.4(–)	–108.5(–)
Ba^{2+}	–71.2(–118.3)	–7.7(–2.0)	–23.2(–25.6)	–2.1(–9.6)	–104.1(–156.1)
Zn^{2+}	–93.8(–237.2)	–1.5(–)	–26.4(–)	–8.7(–)	–130.4(–285.4)
Cd^{2+}	–87.9(–192.6)	–1.1(–)	–26.3(–26.0)	–8.0(–)	–123.3(–237.2)
Hg^{2+}	–94.3(–208.0)	–1.3(–)	–26.2(–25.9)	–8.7(–)	–130.5(–253.9)
$MgOH^+$	–57.6(–)	+0.4(–)	–27.0(–)	–4.8(–)	–89.0(–)

[a] Evaluated with inclusion of the electron correlation at the MP2/6-31G*//HF/6-31G* level. Deformation energies of monomers were not included. The values in parentheses were obtained for the G-C base pair interacting with the unsolvated cation.

bind to nitrogen sites while both metal cations have rather similar affinities to interact with oxygen atoms. The partial covalent bonding involving the 3d electrons of zinc and nitrogen lone electron pairs is the most important contribution. It is well illustrated in Table XVI showing different balance of Zn^{2+} and Mg^{2+} interactions with water and guanine N7. The first two rows compare the interactions between cation and guanine and between cation and a single water molecule. Zinc has a stronger interaction with both water and nucleobase, however, the zinc-magnesium difference is much more pronounced for the cation···nucleobase complex. This difference then influences all the other contributions. Thus, the next row provides the hydration energies of zinc and magnesium in hexahydrate complexes. The hydration energies are rather similar with a 10 kcal/mol difference in favor of Zn^{2+}. The next row of Table XVI presents hydration energy of the cations bound to a nucleobase (hydration of a metalated base, *i.e.*, interaction of the metal–base complex with 5 molecules of waters), indicating that the

TABLE XVI

Interaction energies (in kcal/mol) in selected complexes involving zinc and magnesium divalent cations (M, metal). Direct (inner-shell) binding of pentahydrated M to guanine N7 is considered. The last three rows show different ways how the complex between guanine and pentahydrated cation can be divided into subsystems to highlight the difference between zinc and magnesium

Complex	Zn^{2+}	Mg^{2+}	Difference (Zn^{2+}-Mg^{2+})
G...M^{2+}	–185.1	–148.9	–36.2
H_2O...M^{2+}	–94.9	–82.9	–12.0
M^{2+}...$6H_2O$[a]	–339.8	–329.6	–10.2
G-M^{2+}...$5H_2O$[b]	–203.2	–220.3	+17.1
G...(M^{2+} + $5H_2O$)[c]	–94.4	–89.9	–4.5
G...M^{2+}...$5H_2O$[d]	–388.4	–373.7	–14.7

[a] Hexahydration of the cation (seven subsystems). [b] Hydration of the "metalated base" G-M^{2+} (six subsystems: five water molecules and the metalated base). Note the substantial reversal of the zinc-magnesium difference compared with the preceding line. Replacement of one water molecule by guanine N7 in the primary hydration shell substantially reduces the hydration energy of zinc compared with Mg^{2+}. [c] Interaction between the hydrated cation and base (two subsystems). Although this term appears to be similar for both cations, when considering the preceding two lines we clearly see the different balance of water–cation and base–cation interactions for zinc and magnesium. [d] Interaction energy of the whole complex (seven subsystems).

difference between Zn^{2+} and Mg^{2+} is sharply (by almost 30 kcal/mol) reversed. The reason is the repulsive contribution originating in the weakening of the cation–base binding upon hydration. This contribution is much larger in the case of Zn^{2+}. In contrast, similar interaction energies are obtained when the interaction is formally treated as interaction between a base and a hydrated cation (fifth row of the Table). The last row of the Table shows the total interaction energies of the guanine–metal-hydration shell complex. Here the energy difference between zinc and magnesium complexes increases only slightly with respect to the corresponding value for hexahydrated cations (row 3), again because of the larger reduction of Zn^{2+}–base interaction by hydration. Summarizing, the complex consisting of the guanine molecule and a hydrated cation can be viewed in two ways: (i) As a complex between the hydrated cation and a base (the hydrated cation is taken as one subsystem), and (ii) as hydration of a metalated base (G-M^{2+} is taken as one subsystem). With Zn^{2+} the system is shifted significantly more towards the second type of interaction. The key energy contribution is the more covalent nature of the N7(base)···Zn^{2+} interaction.

In order to further illustrate the zinc *vs* magnesium difference, we carried out an additional set of calculations. First, we have optimized (HF/6-31G* level) the complex of guanine with the hydrated cation (cytosine has been removed from the base pair). Then, we have fixed the N7–M distance to be by 0.1 and 0.2 Å longer (shorter) with respect to the optimized structure while the system has been fully optimized. It corresponds to a variation of the cation–base distance upon a full relaxation of the hydration shell. It is easier to increase the N7–Mg^{2+} distance (the energy penalty of an increase of 0.2 Å is +0.7 kcal/mol) than the guanine–Zn^{2+} distance (+1.7 kcal/mol). Thus, it is easier to separate the Mg^{2+} cation away from the base into the solvent. On the other hand, a compression of the cation–base distance by 0.2 Å requires +5.8 kcal/mol for Mg^{2+}, but only +3.9 kcal/mol for Zn^{2+}. This means that once the Zn^{2+} cation approaches the N7 position of guanine, it tends to be bound there more firmly than Mg^{2+} while its hydration shell is very flexible. On the other hand, Mg^{2+} can be easier released back to solvent and its hydration shell is less flexible with respect to the cation. Thus QM calculations explain why zinc and other transition metal elements interact frequently with nucleobases in DNA while magnesium tends to bind to the anionic oxygens of the phosphate groups.

Metal binding to a nucleotide. Table XVII summarizes partial decomposition of the interaction between hydrated divalent cations and nucleotides, assuming inner-shell direct binding of the metal to the N7 position of the nucleobase. There are five water molecules in the first coordination shell,

two of them bridging the metal with the anionic phosphate oxygens (Fig. 10). The dominating contribution is the pairwise nucleotide–cation interaction (first column of Table XVII). The many-body term (total nonadditivity) is highly repulsive and shows the screening of the nucleotide–cation interaction by the water shell. The nonadditivity in absolute value reaches 17–28% of the total interaction energies. The nonadditivity is considerably larger for Zn^{2+} compared with Mg^{2+}. The sum of the pairwise nucleotide–water interactions is around 0 kcal/mol. Pairwise interactions for the water–backbone bridges are modestly attractive, approximately –3 to –4 kcal/mol (not shown). In reality, however, these water molecules are highly polarized and form very strong water bridges linking the cation with the backbone. The pair contributions do not reveal the actual strength of

TABLE XVII

Interaction energies (in kcal/mol; MP2/6-31G* method) between hydrated cations and nucleotides ($\Delta E^{N,M}$, pairwise nucleotide–metal cation interaction; $\Sigma\Delta E^{N,W}$, sum of the pairwise nucleotide–water interactions; ΔE^{mb} , many-body term; $\Delta E^{N,M+5W}$, total interaction between the nucleotide and the hydrated cation[a])

Complex	$\Delta E^{N,M}$	$\Sigma\Delta E^{N,W}$	ΔE^{mb}	$\Delta E^{N,M+5W}$
DpG-(Mg+5W)	–315.5(–314.0)	–0.3(5.3)	45.0(37.5)	–270.8(–271.3)
DpG-(Zn+5W)	–352.1(–337.3)	0.3(6.0)	77.2(55.6)	–275.1(–275.7)
DpA-(Mg+5W)	–291.6(–288.7)	–0.3(5.8)	47.6(38.6)	–254.3(–244.3)

[a] The data in parentheses shows the HF component of the interaction energy.

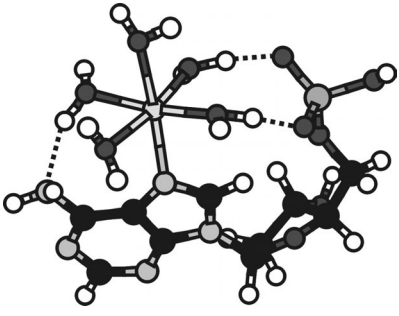

FIG. 10

Inner-shell binding of a hydrated divalent metal cation to the N7 of adenine nucleotide. Note the formation of the amino acceptor interaction between the N6 amino group and the polarized water molecule from the cation hydration shell

these H-bonds, and a major part of the strength of the water bridges is included in the nonadditivities, in this particular cases polarization of the water molecules by the metal cation. Due to the definition of the non-additivity, we cannot separate the actual strength of the water bridges from the other contributions. It is clear, however, that the polarized water molecules belonging to the cation hydration shell are capable of forming H-bonds which are several times stronger than those usually formed by water and much stronger than indicated by the pairwise terms. It is illustrated by enormous elongation of the O–H bonds with a red shift of *ca* 900 cm^{-1} at the HF/6-31G* level. Note that the HF/6-31G* level still underestimates the flexibility of monomers rather significantly. Thus, the energetics of this system can be described only by methods properly accounting for non-additivities and polarization effects.

Electrostatic vs non-electrostatic effects. A frequent assumption in structural biology literature is that interactions of cations with nucleic acids can be explained solely using electrostatic effects and size of the cation. It is likely that, in the first approximation, one can consider only the electrostatics. However, there would be no differences between divalent cations such as zinc and magnesium within this approximation. Recent molecular dynamics simulations of quadruplex DNA molecules indicate that even for monovalent sodium cations the neglect of non-electrostatic effects can lead to artefacts (albeit local) of the simulated structures[50]. Bifurcated H-bonding of guanine quartets and underestimated mobility of the cations in the central channel of the quadruplex stem were observed[50]. The inaccuracy of a leading force field for sodium binding to guanine O6 in G-DNA is illustrated in Table XVIII. The force field underestimates the strength of the binding while the cation appears oversized.

The metal-cation binding is associated with substantial polarization and charge transfer contributions that cannot be described by a simple pair-additive electrostatic term or empirical potential. These contributions are often quoted as non-electrostatic effects in the literature although, strictly speaking, polarization originates in coulombic terms of the Hamiltonian. This is in order to distinguish them from pair-additive long-range electrostatic effects well described by coulombic empirical potential with constant point charges. Note that a vast majority of computational studies of biomolecular systems are carried out using simple force fields neglecting polarization and charge transfer. It is also well established that the long-range electrostatic effects, especially in ionic systems, are mostly fully counterbalanced by polar solvents relevant for biomolecular systems. All the non-electrostatic effects (as defined above) are considerably less affected by the

environment and remain expressed in bio-inorganic experiments. That is another reason why to separate long-range electrostatic and non-electrostatic effects as suggested above.

The effect of metal binding on protonation, deprotonation and tautomerism. Gas phase vs condensed phase trends. Nucleobase metalation may affect tautomeric equilibria of nucleobases, their proton affinities, and ability to form mismatch base pairs[44,51]. The computational studies[44,52] were systematically designed to complement experiments carried out in condensed phase and in crystals[53]. The major issue we were trying to address was the relevance of the gas phase picture revealed by the quantum chemistry (showing the intrinsic trends when the studied systems are in isolation) with respect to the "real" experimental situations. These studies were, among other reasons, motivated by disagreements between the quantum-chemical and bio-

TABLE XVIII
Interaction energy ΔE (in kcal/mol) between guanine O6 and Na^+ and K^+ metal ions depending on the O6–M^+ distance (in Å). The M–O6–C6 angle was frozen at 135° (MP2 *ab initio vs* AMBER force field data)

Monovalent metal ions M^+	O6–M^+ distance, Å	ΔE_{MP2} *ab initio*	ΔE_{AMB} empirical potential
Na^+	1.749	–10.0	+69.5
	1.849	–21.2	+20.8
	1.949	–27.6	–2.3
	2.049	–30.9	–12.8
	2.149[a]	–32.2	–17.3
	2.349	–31.5	–19.0
	2.549	–28.8	–17.6
	2.749	–25.5	–15.6
K^+	2.288	–14.5	–5.8
	2.388	–17.8	–10.9
	2.488	–19.6	–13.1
	2.588[b]	–20.3	–14.0
	2.788	–20.0	–13.8
	2.988	–18.4	–12.6
	3.188	–16.6	–11.3

[a] Optimal O6–M^+ distance obtained by the HF/6-311+G(d,p) method. [b] Optimal O6–M^+ distance obtained by the HF/6-31G(d) and effective core pseudopotential method.

inorganic views on the metal–nucleobase interactions. This can be well illustrated with a systematic study of protonation energies of platinated adenines, which provides insights into the way how the gas phase and condensed phase pictures complement each other[44b]. Table XIX shows the gas phase protonation energies of different Pt(II) adducts with variable charges attached to different sites of adenine (Fig. 11). Not surprisingly, the gas phase energetics of the protonation is dominated by the total charge on the

TABLE XIX
Protonation energies (in kcal/mol) of adenine metalated by $[Pt(NH_3)_3]^{2+}$, *trans*- and *cis*-$[PtCl(NH_3)_2]^+$, *trans*- and *cis*-$[PtCl_2(NH_3)]$ and $[PtCl_3]^-$ entities (Becke3LYP/6-31G* method)

Position of Pt	N1		N3		N7	
Position of the proton	N3	N7	N1	N7	N1	N3
$[Pt(NH_3)_3]^{2+}$	–85.6	–96.0	–89.1	–95.0	–98.7	–98.1
trans-$[PtCl(NH_3)_2]^+$	–149.0	–152.7	–151.5	–152.0	–160.2	–158.35
cis-$[PtCl(NH_3)_2]^+$	–155.3	–159.5	–157.3	–161.3	–166.4	–163.6
trans-$[PtCl_2(NH_3)]$	–224.2	–219.3	–226.3	–223.6	–231.7	–228.1
cis-$[PtCl_2(NH_3)]$	–216.3	–214.5	–217.1	–219.7	–226.5	–222.4
$[PtCl_3]^-$	–285.0	–273.8	–286.6	–281.3	–288.1	–283.5

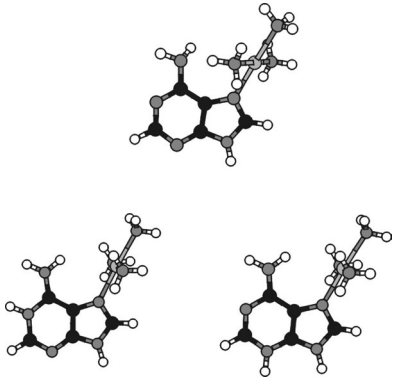

FIG. 11
Platination of N7 of adenine affects the protonation energies and solution pK_a values of N1 and N3 ring positions *via* electrostatic and non-electrostatic effects[44b]

metal entity, with a huge slope of the proton affinity *vs* charge dependence. In a sharp contrast, almost no influence of the charge of the Pt entity on the adenine pK_a values is seen in aqueous solution. The evident lack of the charge dependence shows that the environment very efficiently compensates for the long-range electrostatic effects. However, when subtracting the net ionic contribution from the gas phase data and comparing with adenine multiple protonation data, we have shown that the "non-electrostatic" part of the gas phase trends is nicely reflected by the solution data[44b]. For example, comparing only systems with equal charges (all with +2, *etc*. ...), gas phase and solution data predict very similar relative protonation preferences.

Two recent studies evaluated the correlation between base-pairing association constants of metalated and modified bases (in DMSO) and guanine N1 acidity (in water) with the corresponding gas phase trends[52]. The calculations have demonstrated that although there is a fundamental difference in gas phase deprotonation energies and aqueous solution pK_a data, the condensed phase trends may be well rationalized upon the assumption that they are determined by the non-electrostatic effects with complete screening of the major ionic electrostatics. On the other hand, no correlation could be established between gas phase base-pairing energies and the solution association constants for base pairing. This was explained by the presence of specific effects affecting the experimental values, such as direct (though undesired) interaction of the studied system with counter-anions (*e.g.*, nitrate) present in the solutions.

Metal-assisted tautomers. A salient case is represented by the so called metal-assisted rare tautomers of DNA bases[44a]. In this particular case one of the amino hydrogens of C or A is replaced by a metal entity. After this, proton is observed at the imino N1 (A) or N3 (C) positions. This proton shift is caused by substantial changes of the electronic structure after metalation, accompanied by a substantial increase in proton affinity of bare ring nitrogens. This molecular orbital (non-electrostatic) effect is equally expressed in the gas phase as in the polar solvents.

Cation–π interactions of DNA bases? *Ab initio* calculations are useful for studies of local interactions observed in crystal structures. Thus, we have recently analyzed the possibility of cation–π interactions between hydrated divalent cations and DNA bases. Such interactions have been suggested in B-DNA crystals (see ref.[44c] and references therein). The calculations demonstrated that aromatic base rings are capable of forming, in principle, cation–π complexes of a similar strength as with benzene (Fig. 12). However, in contrast to benzene, aromatic rings of bases have nitrogen and oxy-

gen sites with lone pairs[44c]. Thus, in contrast to benzene, in-plane binding of cations to these nucleobase sites is always highly preferred over any cation–π stacking. Cation–π complexes could only be enforced by some major external contributions[44c]. We have also demonstrated that the nucleobase–cation contacts presently seen in the crystals do not represent any cation–π complexes, as the actual base aromatic system–cation interactions are very weak in the crystal geometries. The observed mutual orientation of hydrated cation with respect to the aromatic plane of cytosines is a simple consequence of a conventional strong in-plane cation binding to neighboring guanines[44c].

In conclusion, although the gas phase calculations on metal–DNA interactions often provide results that at first sight are entirely different from solution and X-ray bio-inorganic experiments, deeper analysis shows that clear correlation exists. The gas phase data provide a picture that is complementary to the bio-inorganic experiments and shows the intrinsic interactions of the metals without perturbation of the environment. This comparison is often needed to properly interpret the experiments and to extrapolate their results towards nucleic acids as the bio-inorganic experiments, in fact, are also done mostly for model complexes.

3. CONCLUSIONS

In the above paragraphs, we have briefly summarized the outcome of high-level *ab initio* calculations on interactions of nucleic acid bases carried out in last several years by our group. Among the most important results was clarification of the nature of base stacking, discovery of nonclassical interactions involving nonplanar amino groups of bases, description of interactions between nucleobases and metals, surprising stabilization of unusual

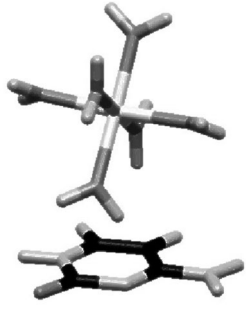

FIG. 12
Possibility of cation–π interactions in DNA duplexes has been investigated by QM methods[44c]

rare tautomers by microhydration as well as by bulk water, *etc.* All these findings contributed to our knowledge of important aspects of nucleic acids structure and dynamics. Our calculations furnish a major reference data set for verification and parametrization of other computational tools. As stated in the introduction, we did not intend to provide a complete overview of the literature and we concentrated mainly on our papers. Many other notable quantum-chemical studies devoted to base stacking, H-bonding and related interactions can be found in recent literature[54], and are discussed in our preceding review papers[1b,16–20].

It is to be noted that qualitative improvements of computational methods occurred also in other areas of theoretical studies of nucleic acids. The most important in our opinion was the advance of large-scale explicit solvent simulations of nucleic acids with accurate treatment of long-range electrostatic forces. This breakthrough was achieved in 1995, when the first simulations appeared[55,56], that is about at the same time when the first electron correlation QM studies were published. The key point in getting successful outcome of simulations of highly charged nucleic acids is a proper treatment of long-range electrostatic forces. With the old cut-off methods, the simulated molecules are unstable on a time scale of hundreds of picoseconds due to a cumulation of errors along the trajectories. With modern methods, stable trajectories are achieved. The most widely used method to treat long range electrostatic forces properly is the particle mesh Ewald (PME) method[57], though also alternative approaches exist. Although there exist studies attempting to reveal possible artifacts of the PME method[58], no convincing failure has been reported on the presently used time scale of *ca* 1–20 ns. There has also been a major effort to reparametrize the pair-additive biomolecular force fields, though their analytical expression remained unchanged[59]. More complex polarization force fields are currently being developed; however, it is not yet clear when a new generation of force fields providing improved results will be available and what level of improvement could be expected. Nevertheless, the presently available force fields appear to be quite robust and well balanced. We are not aware of any report of a major instability of simulations starting from atomic resolution experimental data. We have for example reported major instability during a simulation of a DNA gaunine quadruplex with lateral thymidine loops starting from crystal data[50a]: however, very recent crystallographic studies of the same molecule indicate that the original crystal structure reported an incorrectly determined overall fold of the whole quadruplex[60]. On the other hand, trajectories starting from model structures often show disintegration or large structural changes indicating that the simulations are not

artificially stabilized[50b]. Thus, contemporary simulations are well suited to basically reproduce the existing atomic resolution structures although, of course, local structural differences sometimes occur and may be attributed mostly to the force-field approximations. The simulations can complement the experiments in a number of aspects. For example, our simulations have recently revealed that complex RNA folds such as pseudoknots and 5S rRNA loop E extended non-Watson–Crick motif are stabilized by long-residency water molecules with binding times of individual water molecules 10–40 times longer compared with common hydration sites[61]. The simulations do not yet allow predicting molecular structures of nucleic acids *de novo*.

In contrast to the QM calculations, the simulations do not provide energy information in a straightforward way. This obviously is a major limitation. There have been substantial efforts to overcome this drawback and the most promising approach appears to be the method abbreviated as MM_PBSA [62]. In the MM_PBSA technique, the free energy of the simulated molecule is derived a posteriori from the simulated trajectory (using series of snapshots) with the explicit solvent replaced by continuum described by the Poisson–Boltzmann method. In this way, one can in principle compare free energies of two folds of the same molecule without simulating the transitions, calculate binding free energies, *etc.* However, the technique has also a wide range of limitations. These stem from the limited sampling on the nanosecond time scale, force field approximations (these two limitations are inherent to the simulation itself), neglect of specific bound water and cations (nevertheless, one can include a subset of tightly bound ions and water explicitly), estimates of the solute entropy and sensitivity of the calculated results to the atomic radii used in the PB calculation. The last limitation is quite significant and reminds of difficulties in continuum solvent QM methods. We have recently applied the MM_PBSA method to study binding of DAPI to DNA [25c] and substates involved in guanine stem quadruplex formation[62]. Unfortunately, the MM_PBSA method was not capable of reproducing absolute binding energies of DAPI. Nevertheless, the results regarding relative free energies were considerably more promising and other groups have similar experience.

In conclusion, we believe that the next decade will bring additional substantial improvements in computational studies of nucleic acids, we are excited by the recent progress and look forward to further developments in the field. We believe that the best outcomes can be obtained when advanced computational tools are combined in an integrated way as each of the available methods has its advantages and limitations. In our opinion, the power of QM methods to understand molecular interactions in nucleic

acids has not yet been fully utilized. This is partly due to prevailing negative opinion about computational methods in many structural biology experimental laboratories. Also, many groups utilizing large-scale MD methods tend to undervalue and downplay the message of QM data, perhaps because the QM methodology often clearly points to substantial deficiencies in the force fields. Pairwise empirical potential is certainly not the end of the efforts and relevant discussion of the force fields in light of the QM data should be a routine part of the MD research, in contrast to the present cover-up practice. The other side of the problem is that quite often QM studies are too specialized and lacking any messages to structural biologists. We would like to emphasise that QM technique is currently one of the leading tools in physical chemistry of molecular interactions widely accepted by the physical chemistry experimental community. As we assume a significant expansion of the modern physical chemistry methods into the structural biology field, the role of QM methods is likely to increase substantially.

This study was supported by the following grants: LN00A032 by the Ministry of Education, Youth and Sports of the Czech Republic (to J. Šponer), A4040904 by the Grant Agency of the Academy of Sciences of the Czech Republic (to P. Hobza) and Wellcome Trust International Senior Research Fellowship for Biomedical Science in Central Europe GR067507 (to J. Šponer).

4. REFERENCES

1. a) Šponer J., Leszczynski J., Hobza P.: *J. Phys. Chem.* **1996**, *100*, 5590; b) Hobza P., Šponer J.: *Chem. Rev.* **1999**, *99*, 3247.
2. Ramakrishnan V., Moore P. B.: *Curr. Opin. Struct. Biol.* **2001**, *11*, 144.
3. Schneider B., Berman H. M.: *Biophys. J.* **1995**, *69*, 2661.
4. Drew H. R., Dickerson R. E.: *J. Mol. Biol.* **1981**, *151*, 535.
5. a) Lippert B.: *Coord. Chem. Rev.* **2000**, *200*, 487; b) Sigel A., Sigel H. (Eds): *Metal Ions in Biological Systems*, Vol. 32. Marcel Dekker, New York 1996.
6. Hobza P., Sandorfy C.: *J. Am. Chem. Soc.* **1987**, *109*, 1302.
7. a) Bugg C. E., Thomas J. M., Sundaralingam M., Rao S. T.: *Biopolymers* **1971**, *10*, 175; b) Hunter C. A.: *J. Mol. Biol.* **1993**, *230*, 1025.
8. a) Šponer J., Gabb H. A, Leszczynski J., Hobza P.: *Biophys. J.* **1997**, *73*, 76; b) Šponer J., Florian J., Ng H. L., Šponer J. E., Špačková N.: *Nucl. Acids Res.* **2000**, *28*, 4893.
9. a) Yanson I. K., Teplitsky A. B., Sukhodub L. F.: *Biopolymers* **1979**, *18*, 1149; b) Sukhodub L. F.: *Chem. Rev.* **1987**, *87*, 589; c) Desfrancois C., Abdoul-Carime H., Schulz C. P., Schermann J. P.: *Science* **1995**, *269*, 1707; d) Schnier P. D., Klassen J. S., Stritmatter E. F., Williams E. R.: *J. Am. Chem. Soc.* **1998**, *120*, 9605; e) Hoyau S., Norrman K., McMahon T. B., Ohanessian G.: *J. Am. Chem. Soc.* **1999**, *121*, 8864; f) Nir E., Kleinermanns K., de Vries M. S.: *Nature* **2000**, *408*, 949; g) Nir E., Plutzer C., Kleinermanns K., de Vries M. S.: *Eur. Phys. J. D* **2002**, *20*, 317; h) Nir E., Janzen C., Imhof P., Kleinermanns K., de Vries

M. S.: *Phys. Chem. Chem. Phys.* **2002**, *4*, 740; i) Dong F., Miller R. E.: *Science* **2002**, *298*, 1227.

10. a) Kratochvíl M., Engkvist O., Šponer J., Jungwirth P., Hobza P.: *J. Phys. Chem. A* **1998**, *102*, 6921; b) Kratochvíl M., Šponer J., Hobza P.: *J. Am. Chem. Soc.* **2000**, *122*, 3495.
11. a) Kuechler E., Derkosch J.: *Z. Naturforsch., B: Anorg. Chem., Org. Chem.* **1966**, *21*, 209; b) Pitha J., Jones R. N., Pithova P.: *Can. J. Chem.* **1966**, *44*, 1045; c) Ts'o P. O. P, Melvin I. S., Olson A. C.: *J. Am. Chem. Soc.* **1963**, *85*, 1289; d) Gray D. M.: *Biopolymers* **1997**, *42*, 783; and references therein; e) SantaLucia J., Jr.: *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460; f) Newcomb L. F., Gellman S. H. L.: *J. Am. Chem. Soc.* **1994**, *116*, 4993; g) Sartorius J., Schneider H.-J.: *J. Chem. Soc., Perkin Trans. 2* **1997**, 2319; h) Mizutani M., Kubo I., Jitsukawa K., Masuda H., Einaga H.: *Inorg. Chem.* **1999**, *38*, 420; i) Allawi H. T., SantaLucia J.: *Nucl. Acids Res.* **1998**, *26*, 2694; j) Nakano S., Fujimoto M., Hara H., Sugimoto N.: *Nucl. Acids Res.* **1999**, *27*, 2957; k) Bommarito S., Peyret N., SantaLucia J.: *Nucl. Acids Res.* **2000**, *28*, 1929.
12. a) Cramer C. J., Truhlar D. G.: *Chem. Rev.* **1999**, *99*, 2161; b) Luque F. J., LopezBes J. M., Cemeli J., Aroztegui M., Orozco M.: *Theor. Chem. Acc.* **1997**, *96*, 105.
13. Florian J., Šponer J., Warshel A.: *J. Phys. Chem. B* **1999**, *103*, 884; and references therein.
14. a) Pohorille A., Burt S. K., MacElroy R. D.: *J. Am. Chem. Soc.* **1984**, *10*, 402; b) Danilov V. I., Tolokh I. S.: *J. Biomol. Struct. Dyn.* **1984**, *2*, 119; c) Cieplak P., Kollman P. A.: *J. Am. Chem. Soc.* **1988**, *110*, 3334; d) Friedman R. A., Honig B.: *Biophys. J.* **1995**, *69*, 1528; e) Arora N., Jayram B.: *J. Phys. Chem. B* **1998**, *102*, 6139; f) Luo R., Gilson H. S. R., Potter M. J., Gilson M. K.: *Biophys. J.* **2001**, *80*, 140; g) Florian J., Goodman M. F., Warshel A.: *J. Phys. Chem. B* **2000**, *104*, 10092.
15. a) Barsky D., Kool E. T., Colvin M. E.: *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119; b) Sivanesan D., Subramanian V., Nair B. U., Ramasami T.: *Indian J. Chem., Sect. A: Inorg., Bio-inorg., Phys., Theor. Anal. Chem.* **2000**, *39*, 132; c) Kabeláč M., Ryjáček F., Hobza P.: *Phys. Chem. Chem. Phys.* **2000**, *2*, 4906; d) Sivanesan D., Babu K., Gadre S. R., Subramanian V.: *J. Phys. Chem. A* **2000**, *104*, 10887.
16. Šponer J., Berger I., Špačková N., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn. Conversation 11* **2000**, *2*, 383.
17. Šponer J., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn.* **1996**, *14*, 117.
18. Šponer J., Leszczynski J., Hobza P. in: *Computational Chemistry – Reviews of Current Trends* (J. Leszczynski, Ed.), p. 185. World Scientific Publisher, Singapore 1996.
19. Šponer J., Hobza P. in: *Encyclopedia of Computational Chemistry* (P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, Eds), p. 777. John Wiley & Sons, Chichester 1998.
20. Šponer J., Leszczynski J., Hobza P. in: *Theoretical Computional Chemistry* (J. Leszczynski, Ed.), Vol. 8, p. 85. Elsevier, Oxford 1999.
21. a) Trygubenko S., Bogdan T. V., Rueda M., Orozco M., Luque F. J., Šponer J., Slavíček P., Hobza P.: *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192; b) Hanus M., Ryjáček F., Kabeláč M., Kubar T., Trygubenko S., Bogdan T. V., Hobza P.: *J. Am. Chem. Soc.* **2003**, *125*, 7678; c) Hanus M., Ryjáček F., Kabeláč M., Rejnek J., Hobza P.: *Phys. Chem. Chem. Phys.* **2003**, *5*, in press.
22. a) Leszczynski J.: *Int. J. Quantum Chem., Quantum Biol. Symp.* **1992**, *19*, 43; b) Šponer J., Hobza P.: *J. Phys. Chem.* **1994**, *98*, 3161; c) Bludsky O., Šponer J., Leszczynski J., Špirko V., Hobza P.: *J. Chem. Phys.* **1996**, *105*, 11042; d) Šponer J., Hobza P.: *Int. J. Quantum Chem.* **1996**, *57*, 959; e) Dong F., Miller R. E.: *Science* **2002**, *298*, 1227.

23. Šponer J., Hobza P.: *J. Am. Chem. Soc.* **1994**, *116*, 709.
24. Šponer J., Florian J., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.
25. a) Luisi B., Orozco M., Šponer J., Luque F. J., Shakked Z.: *J. Mol. Biol.* **1998**, *279*, 1123; b) Vlieghe D., Šponer J., van Meervelt L.: *Biochemistry* **1999**, *38*, 16443; c) Špačková N., Cheatham T. E., Ryjáček F., Lankas F., van Meervelt L., Hobza P., Šponer J.: *J. Am. Chem. Soc.* **2003**, *125*, 1759; d) Šponer J., Mokdad A., Šponer J. E., Špačková N., Leszczynski J., Leontis N. B.: *J. Mol. Biol.* **2003**, *330*, 967.
26. Ryjáček F., Kubar T., Hobza P.: *J. Comput. Chem.* **2003**, *24*, 1891.
27. Šponer J., Kypr J.: *Int. J. Biol. Macromol.* **1994**, *16*, 3.
28. a) Šponer J., Sabat M., Gorb L., Leszczynski J., Lippert B., Hobza P.: *J. Phys. Chem. B* **2000**, *104*, 7535; b) Soliva R., Laughton C. A., Luque F. J., Orozco M.: *J. Am. Chem. Soc.* **1998**, *120*, 11226; c) Štefl R., Koča J.: *J. Am. Chem. Soc.* **2000**, *122*, 5025.
29. Šponer J., Leszczynski J., Hobza P.: *J. Phys. Chem.* **1996**, *100*, 1965.
30. Šponer J., Leszczynski J.,Vetterl V., Hobza P.: *J. Biomol. Struct. Dyn.* **1996**, *13*, 695.
31. a) Šponer J., Leszczynski J., Hobza P.: *J. Phys. Chem. A* **1997**, *101*, 9489; b) Šponer J., Burda J. V., Mejzlík P., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn.* **1997**, *14*, 613.
32. a) Hobza P., Šponer J.: *Chem. Phys. Lett.* **1998**, *288*, 7; b) Hobza P., Šponer J., Cubero E., Orozco M., Luque F. J.: *J. Phys. Chem. B* **2000**, *104*, 6286; c) Elstner M., Hobza P., Frauenheim T., Suhai S., Kaxiras E.: *J. Chem. Phys.* **2001**, *114*, 5149; d) Hobza P., Špirko V.: *Phys. Chem. Chem. Phys.* **2003**, *5*, 1290.
33. Toczylowski R. R., Cybulski S. M.: *J. Phys. Chem. A* **2003**, *107*, 418.
34. a) Šponer J., Hobza P.: *Chem. Phys. Lett.* **1997**, *267*, 263; b) Hobza P., Šponer J.: *J. Am. Chem. Soc.* **2002**, *124*, 11802; c) Leininger M. L., Nielsen I. M. B., Colvin M. E., Janssen C. L.: *J. Phys. Chem. A* **2002**, *106*, 3850.
35. Brameld K., Dasgupta S., Goddard III W. A.: *J. Phys. Chem. B* **1997**, *101*, 4851.
36. a) Hobza P., Kabeláč M., Šponer J., Mejzlík P., Vondrášek J.: *J. Comput. Chem.* **1997**, *18*, 1136; b) Jurečka P., Hobza P.: *J. Am. Chem. Soc.*, **2003**, *125*, in press; c) Hobza P., Riehn C., Weichert A., Brutschy B.: *Chem. Phys.* **2002**, *283*, 331; d) Šponer J., Hobza P.: *J. Phys. Chem. A* **2000**, *104*, 4592.
37. Kroon-Batenburg L. M. J., van Duijneveldt F. B.: *J. Mol. Struct.* **1985**, *121*, 185.
38. Boys S. F., Bernardi B.: *Mol. Phys.* **1970**, *19*, 553.
39. van Duineveldt F. B., van Duineveldt-van de Rijdt J. G. C. M., van Lenthe J. H.: *Chem. Rev.* **1994**, *94*, 1873.
40. Šponer J., Leszczynski J., Hobza P.: *J. Comput. Chem.* **1996**, *17*, 841.
41. a) Wesolowski T. A., Parisel O., Ellinger Y., Weber J.: *J. Phys. Chem. A* **1997**, *101*, 7818; b) Kurita N., Araki M., Nakao K., Kobayashi K.: *Int. J. Quantum Chem.* **2000**, *76*, 677; c) Williams H. L., Chabalowski C. F.: *J. Phys. Chem. A* **2001**, *105*, 646; d) Perez-Jorda J. A., San-Fabian E., Perez-Jimenez J.: *J. Chem. Phys.* **1999**, *110*, 1916.
42. Alhambra C., Luque F. J., Gago F., Orozco O.: *J. Phys. Chem. B* **1997**, *101*, 3846.
43. a) Kratochvíl M., Engkvist O., Vacek J., Jungwirth P., Hobza P.: *Phys. Chem. Chem. Phys.* **2000**, *2*, 2419; b) Gervasio F. L., Procacci P., Cardini G., Guarna A., Giolitti A., Schettino V.: *J. Phys. Chem. B* **2000**, *104*, 1108; c) Kabeláč M., Hobza P.: *J. Phys. Chem. B* **2001**, *105*, 5804; d) Kratochvíl M., Šponer J., Hobza P.: *J. Am. Chem. Soc.* **2000**, *122*, 3495; e) Kabeláč M., Hobza P.: *Chem. Eur. J.* **2001**, *7*, 2067.
44. a) Šponer J., Šponer J. E., Gorb L., Leszczynski J., Lippert B.: *J. Phys. Chem A* **1999**, *103*, 11406; b) Šponer J. E., Glahe F., Leszczynski J., Lippert B., Šponer J.: *Inorg. Chem.* **2000**, *269*, 2868; c) Šponer J., Šponer J. E., Leszczynski J.: *J. Biomol. Struct. Dyn.* **2000**, *17*, 1087.

45. Burda J. V., Šponer J., Hobza P.: *J. Phys. Chem.* **1996**, *100*, 7250.
46. a) Burda J. V., Šponer J., Leszczynski J., Hobza P.: *J. Phys. Chem. B* **1997**, *101*, 9670; b) Šponer J., Burda J. V., Sabat M., Leszczynski J., Hobza P.: *J. Phys. Chem. A* **1998**, *102*, 5951; c) Šponer J., Sabat M., Burda J. V., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn.* **1998**, *16*, 139.
47. Šponer J., Burda J. V., Leszczynski J., Hobza P.: *J. Biomol. Struct. Dyn.* **1999**, *17*, 61.
48. a) Sigel R. K. O., Lippert B.: *Chem. Commun.* **1999**, 2167; b) Sigel R. K. O., Freisinger E., Lippert B.: *J. Biol. Inorg. Chem.* **2000**, *5*, 287.
49. Han W. H., Dlakic M., Zhu Y. W. J., Lyndsay S. M., Harrington R. E.: *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10565.
50. a) Špačková N., Berger I., Šponer J.: *J. Am. Chem. Soc.* **1999**, *121*, 5519; b) Štefl R., Špačková N., Berger I., Šponer J.: *Biophys. J.* **2001**, *80*, 455; c) Špačková N., Berger I., Šponer J.: *J. Am. Chem. Soc.* **2001**, *123*, 3295.
51. Burda J. V., Šponer J., Leszczynski J.: *J. Biol. Inorg. Chem.* **2000**, *5*, 178.
52. a) Schmidt K. S., Reedijk J., Weisz K., Janke E. M. B, Šponer J. E., Šponer J., Lippert B.: *Inorg. Chem.* **2002**, *41*, 2855; b) Burda J. V., Šponer J., Hrabáková J., Zeizinger M., Leszczynski J.: *J. Phys. Chem. B* **2003**, *107*, 5349.
53. Lippert B.: *J. Chem. Soc., Dalton Trans.* **1997**, 3971; and references therein.
54. a) Florian J., Leszczynski J.: *J. Am. Chem. Soc.* **1996**, *118*, 3010; b) Bertran J., Oliva A., Rodriguez-Santiago L., Sodupe M.: *J. Am. Chem. Soc.* **1998**, *120*, 8159; c) Chandra A. K., Nguyen M. T., Uchimaru T., Zeegers-Huyskens T.: *J. Phys. Chem. A* **1999**, *103*, 8853; d) Kawahara S., Uchimaru T.: *Chem. Phys. Phys. Chem.* **2000**, *2*, 869; e) Plokhotnichenko A. M., Radchenko E. D., Stepanian S. G., Adamowicz L.: *J. Phys. Chem. A* **1999**, *103*, 11052; f) Meyer M., Suhnel J.: *J. Biomol. Struct. Dyn.* **1997**, *15*, 619; g) Brandl M., Meyer M., Suhnel J.: *J. Am. Chem. Soc.* **1999**, *121*, 2605; h) Raimondi M., Famulari A., Gaininetti E.: *Int. J. Quantum Chem.* **1999**, *74*, 259; i) Kawahara S., Wada T., Kawauchi S., Uchimaru T., Sekine M.: *J. Phys. Chem. A* **1999**, *103*, 8516; j) Alagona G., Ghio C., Giolitti A., Monti S.: *Theor. Chem. Acc.* **1999**, *101*, 143; k) McCarthy W., Plokhotnichenko A. M., Radchenko E. D., Smets J., Smith D. M. A., Stepanian S. G., Adamowicz L.: *J. Phys. Chem. A* **1997**, *101*, 7208; l) Guerra C. F., Bickelhaupt F. M., Snijders J. G., Bearends E. J.: *J. Am. Chem. Soc.* **2000**, *122*, 4117; m) Bondarev D. A., Skawinski W. J., Venanzi C. A.: *J. Phys. Chem. B* **2000**, *104*, 815; n) Gago F.: *Methods Enzymol.* **1998**, *14*, 277; o) Colson A.-O., Sevilla M. D.: *Int. J. Radiat. Biol.* **1995**, *67*, 627; p) Gu J. D., Leszczynski J.: *J. Phys. Chem. A* **2000**, *104*, 6308; q) Meyer M., Steinke T., Brandl M., Suhnel J.: *J. Comput. Chem.* **2001**, *22*, 109; r) Famulari A., Moroni F., Sironi M., Raimondi M.: *Comput. Chem.* **2000**, *24*, 341; s) Guerra C. F., Bickelhaupt F. M., Snijders J. G., Baerends E. J.: *Chem. Eur. J.* **1999**, *5*, 3581.
55. Cheatham III T. E., Miller J. L., Fox T., Darden T. A., Kollman P. A.: *J. Am. Chem. Soc.* **1995**, *117*, 4193.
56. a) Cheatham T. E., Young M. A.: *Biopolymers* **2000**, *56*, 232; b) Kollman P. A., Massova I., Reyes C., Kuhn B., Huo S. H., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D. A., Cheatham T. E. L.: *Acc. Chem. Res.* **2000**, *33*, 889; c) Srinivasan J., Cheatham T. E., Cieplak P., Kollman P. A., Case D. A.: *J. Am. Chem. Soc.* **1998**, *120*, 9401.
57. Darden T., York D., Pedersen L.: *J. Chem. Phys.* **1993**, *98*, 10089.
58. Hunenberger P. H., McCammon J. A.: *Biophys. Chem.* **1999**, *78*, 69.

59. a) Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A.: *J. Am. Chem. Soc.* **1995**, *117*, 5179; b) MacKerell A. D., Workiewicz-Kuczera J., Karplus M.: *J. Am. Chem. Soc.* **1995**, *117*, 11946.
60. Haider S., Parkinson G. N., Neidle S.: *J. Mol. Biol.* **2002**, *320*, 189.
61. a) Csaszar K., Špačková N., Štefl R., Šponer J., Leontis N. B.: *J. Mol. Biol.* **2001**, *313*, 1073; b) Reblová K., Špačková N., Štefl R., Csaszar K., Koča J., Leontis N. B., Šponer J.: *Biophys. J.* **2003**, *84*, 3564.
62. Štefl R., Cheatham III T. E., Špačková N., Fadrna E., Berger I., Koča J., Šponer J.: *Biophys. J.* **2003**, *85*, 1787.