Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical *Ab Initio* Calculations

Pavel Hobza* and Jiří Šponer

J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic

Received March 2, 1999 (Revised Manuscript Received June 29, 1999)

Contents

I. Introduction

Living organisms contain a set of instructions that specifies every step required for the organism to construct a replica of itself. This information is stored in deoxyribonucleic acid, DNA, and the respective molecule is thus one of the most important molecules in our life. The discovery of DNA was made in 1869 by Miescher[1] and it took more than 70 years to demonstrate[2] that DNA was the molecule that carried genetic information. A further important step in the study of DNA was made in 1953 by Watson and Crick,[3] who proposed a three-dimensional, right-handed double-helical B-DNA model of the DNA structure. This model was discovered on the basis of analysis of fiber diffraction data. More advanced experimental studies revealed that DNA can adopt other double-helical forms, such as A-DNA, and many other three-dimensional forms, including triple[4] and quadruple helices.[5−10] Another biologically very important molecule closely related to DNA is ribonucleic acid, RNA. DNA and RNA possess an enormous conformational variability.

Further improvement of our knowledge of the DNA structure has been achieved due to the advance of high-resolution oligonucleotide crystal structure studies.[11−14] It has been established that the B-DNA molecules (as well as all other forms of nucleic acids) are nonuniform molecules, exhibiting a significant conformational variability at the base-pair step level. The sequence-dependent variability is believed to be determined, among other factors, by base stacking.[15−22] The conformational variability of nucleic acids plays the key role in transferring the information through biomolecular recognition processes and how the biomacromolecules read each other. Nevertheless, intense investigation of the relationship between sequence and the fine aspects of B-DNA variability still did not provide the ultimate picture of sequence−local structure relationships. These relationships are sensitive to the quality of the refined crystal structures, and it has been convincingly demonstrated that only the highest quality oligonucleotide crystal structures are accurate enough to unambiguously separate local conformational variations from the refinement data and/or errors.[22b,23]

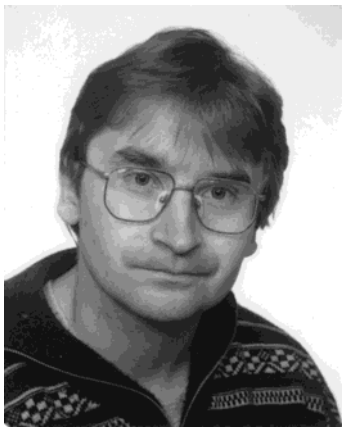
On the other hand, crystallographic data revealed (and continue to reveal) a number of new exciting structures. The first one was left-handed Z-DNA, which usually diffracts to very high resolution.[24,25] Another noticeable molecule is a four-stranded G-DNA; the recently published structure of parallel G-DNA is the highest resolution DNA molecule published so far.[9,10] Many other DNA forms including triplexes, however, have not been crystallized so far.

Significant progress has been made also in NMR studies of DNA structure.[26−29] An important advantage of this technique is that it studies DNA in solution. Although the resolution of the NMR methods is behind the crystallographic data, it provides unique information. NMR techniques can characterize many structures which could not be crystallized. Their power has been demonstrated by the recent discovery of one of the most exciting, unusual DNA forms, four-stranded intercalated hemiprotonated

* Corresponding author. Tel: (+4202) 6605 2056. Fax: (+4202) 858 2307. E-mail:hobza@indy.jh-inst.cas.cz.

10.1021/cr9800255 CCC: $35.00
Published on Web 10/13/1999

Pavel Hobza was born in 1946 in Přerov, Czechoslovakia, and graduated from Czech Technical University in 1969. In 1974 he received his Ph.D. (with Professor R. Zahradník) and in 1988 the D.Sc. degree; after 1989 he became Associated Professor of Charles University in Prague. Since 1989 he has been employed at the J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague. After postdoctoral study with Professor C. Sandorfy at the Université de Montréal, Montréal, he spent several periods as a visiting professor and a visiting scientist at Université de Montréal, Montréal; Friedrich-Alexander-Universität Erlangen-Nürnberg, and Technische Universität München. Since 1998 he has been head of the Department of Chemical Physics at the J. Heyrovský Institute. He has been a fellow of the Learned Society of the Czech Republic since 1998. Dr. Hobza has authored or coauthored about 180 papers and two books. These studies dealt mainly with molecular interactions and their role in physical chemistry and biodisciplines.



Jiří Šponer was born in 1964 in Brno, Czechoslovakia, and graduated at the Masaryk University in 1987. In 1992 he received his Ph.D. degree at the Masaryk University. Essential for his early scientific career was his research at the Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno (since 1987) and at the J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Prague (since 1991). In 1993 he became a researcher at the Institute of Biophysics and later that year he moved to work with Pavel Hobza at the J. Heyrovský Institute of Physical Chemistry. Presently he is a researcher at the J. Heyrovský Institute of Physical Chemistry while still keeping a part-time faculty position at the Institute of Biophysics. He has published about 60 scientific papers including several review articles. His main research interests are structure and dynamics of nucleic acids and interactions of nucleic acid bases.

i-DNA,[30] confirmed later by X-ray crystallography.[31,32]

Another important area of intense research is analysis of DNA bending,[33] which was first characterized by anomalous electrophoretic gel retardation of certain sequences (for a recent review concerning DNA bending, see ref 33e).

From the experimental techniques directly related to the present review, the gas-phase experiments of nucleic acid (NA) bases and their complexes should be mentioned. An important advantage of such studies is the fact that they allow determination of structure and properties of clusters studied without the influence of a solvent.[34−39]

Studies of the structure and dynamics of DNA are vital for understanding the mechanisms of how the genetic code is expressed, processes of DNA replication and transcription, DNA−protein recognition, DNA−drug interactions and others. Various experimental techniques are used to study nucleic acids. The experimental techniques are more and more supplemented by state-of-the-art computational techniques. Rapid development of computer hardware and software at the end of the 1980s led to extended application of computational chemistry to biodisciplines including various aspects of nucleic acids and their structures. We have evidenced an impressive improvement of the quality of all-atom molecular dynamics simulations of hydrated nucleic acids[40−54] due to the advances in accurate treatments of electrostatic interactions using the particle mesh Ewald (PME) method[40−42] and also partly due to improvement of force fields for nucleic acids (for a review, see ref 54).

Another area of tremendous development is *ab initio* quantum chemistry. The high-level *ab initio* calculations expanded into most areas of science during the past few years, playing a major role in many cases. The recent success and credit of *ab initio* quantum chemistry have been reflected by the 1998 Nobel Prize awarded to J. A. Pople and W. Kohn. Quantum chemistry is one of the main tools for parametrization of empirical force fields these days. Because the *ab initio* theory is deductive, it can provide experimentalists with a consistent set of various properties of NA bases and their complexes prior to any experiment being performed or even designed. *Ab initio* calculations were intensely applied to study nucleic acid bases and their hydrogen-bonded (H-bonded) and stacked complexes. Base stacking[55−57] and H-bonding of bases[58] were first investigated by the *ab initio* methods in 1986−1988. In the first pioneering studies,[55−58] the important role of dispersion (correlation) energy was recognized. Many other studies had been published in the next years.[59−67] (A brief overview of some additional older studies can be found in our previous reviews.[68−70]) A real breakthrough in the quality of *ab initio* studies of bases and base pairs occurred around 1994−1995, when the first high-level *ab initio* calculations with consistent treatment of electron correlation effects became feasible. These calculations improved our knowledge of the structure and energetics of NA bases[71−75] and NA base pairs.[76−93] These calculations basically provided the ultimate picture of the origin and magnitude of interactions of DNA bases in the gas phase.

Nevertheless, to make the theory directly comparable with gas phase[34−39] and condensed phase experiments[94−97] we must provide more complete information about DNA interactions. Computer experiments and mainly molecular dynamics (MD) simulations represent a tool for passing from isolated NA bases or their gas-phase complexes at tempera-

ture of 0 K to interactions in a polar solvent and nonzero temperature, implying the importance of the entropy. Conditio sine qua non for performing the computer experiments represents the knowledge of empirical potential describing all intermolecular and intramolecular degrees of freedom of the system under study. The quality of the empirical potential used predetermines the quality of calculated results; combination of high-level computer experiments with a poor empirical potential necessarily leads to poor results. Empirical potentials were originally parametrized almost exclusively using experimental characteristics, parameters of recently introduced potentials are obtained from experimental and theoretical (mostly *ab initio*) data. We believe that the future parametrizations of empirical potentials will be based almost solely on theoretical grounds, though some empirical corrections toward condensed phase might still be necessary. Performance of various widely used empirical potentials to describe interactions of NA bases was recently tested by comparing with high level *ab initio* calculations for NA base pairs.[98,99]

Nonzero temperature brings entropy into consideration. The potential energy surface (PES) and free energy surface (FES) of NA base pairs (and more generally of all molecular clusters) may differ substantially. Entropy can favor one structural arrangement of a cluster over another; specifically, it can influence the relation between H-bonded and stacked NA base pairs. Proper evaluation of the gas-phase thermodynamics is essential to understanding and interpreting gas-phase experiments. Thermodynamic characteristics can be estimated using classical statistical thermodynamic treatment [i.e. rigid rotor−harmonic oscillator−ideal gas approximation (RR−HO−IG)][100] or more accurately evaluated by performing computer experiments.[101] A recent MD study[101] clearly indicates that most likely the situation observed in the gas phase experiments on NA base pairs corresponds to a mixture of several significantly populated isomers including stacked structures.

The present review concerns mainly with the gas-phase ineractions of DNA base pairs. It should be noted, however, that any investigation of base pairing and stacking of nucleobases in condensed and solid states requires us first to understand and reproduce the gas-phase interaction between bases. In classical empirical potential molecular dynamics studies, for example, pair additive force fields are used. This means that in such studies solvent molecules are added without any explicit consideration of the perturbation of the gas-phase electronic structure of the solute due to solvent. Thus the energy is calculated as a sum of solute−solute, solute−solvent, and solvent−solvent terms while the solute−solute term is calculated exactly as in the gas phase. Other models can consider explicitly polarization of solute; however, even in this case, one needs of course to know accurate gas-phase solute−solute energetics before starting to think how this term can be modified by the solvent. The importance of the gas-phase calculations (or experiments) is further underlined by the fact that it impossible to separate (extract) the solute−solute potential energy based on any condensed-phase data.

The aim of the present review is to orient the reader in the latest developments in the theoretical calculations of NA bases and base pairs and to help him or her to carry out similar calculations correctly and with reasonable computer demands.

II. Calculations

II.1. Interaction Energy

The interaction energy of the NA base pairs is evaluated almost exclusively by using the variational supermolecular method. The other theoretical procedure, the perturbational method, is applied only to smaller molecular complexes and its use is tedious. It mostly provides a benchmark for supermolecular calculations because it is free of the main obstacle of the variational calculation, the basis set superposition error (see later discussion). Interaction energy (ΔE) is within the variational approach determined as a difference between energy of a complex (E) and energies of isolated subsystems forming the complex (the energy of the ith subsystem is denoted as E_i):

$$\Delta E = E - \sum E_i \tag{1}$$

Values of E and $\sum E_i$ typically differ only by several kcal mol^{-1}, while the value of E is many orders of magnitude larger than that of ΔE. Therefore, all energies must be calculated with high accuracy—to 8 or more valid digits. Because the energy of a system is usually expressed as a sum of the Hartree−Fock (HF) and correlation (COR) energies, the interaction energy ΔE can be expressed as a sum of the HF interaction energy (ΔE^{HF}) and the correlation interaction energy (ΔE^{COR})

$$\Delta E = \Delta E^{HF} + \Delta E^{COR} \tag{2}$$

The correlation interaction energy can be separated into intrasystem and intersystem components and a small coupling term. According to the perturbational theory, the HF interaction energy is roughly identical to a sum of Coulombic, induction, and exchange−repulsion terms, and the intersystem correlation energy corresponds to the dispersion energy. The intrasystem correlation energy is either attractive or repulsive and corresponds to the correlation correction to the Coulombic and exchange energies. The intrasystem correlation energy is important if the dipole moments of the subsystems determined at the HF and correlated levels differ considerably.

There are several advantages of the supermolecular approach which are profitable for studies of NA base pairs and larger clusters: (i) it is applicable to any type of a molecular cluster; (ii) it yields a wave function which can be used to derive various properties of the system; (iii) highly accurate interaction energies can be obtained, provided that sufficiently large basis sets of atomic orbitals are used and a major portion of the correlation energy is included; and (iv) many-body interactions and charge transfer effects are explicitly taken into account.

There are, however, also drawbacks of the supermolecular approach and two inconsistencies should be mentioned: the size inconsistency and the basis set inconsistency. The size inconsistency does not represent any problem nowadays, since the most efficient methods for calculation of correlation energy are size-consistent (the HF method is size-consistent as well). The basis set inconsistency is a serious obstacle and leads to the basis set superposition error (BSSE). The BSSE is a purely mathematical artifact due to the fact that different basis sets are used for energy evaluations of the supersystem and the subsystems. The supersystem possesses a larger basis set which (artificially) increases its stabilization energy, and this effect is referred to as BSSE. Jansen and Ross,[102] and Boys and Bernardi[103] independently introduced the function counterpoise (CP) method, which eliminates the BSSE. The principle of the method is simple: subsystems are not treated in their own basis sets but in the basis set of the whole complex.

It has been repeatedly suggested that full CP overcorrects the BSSE at a correlated level. Recent comprehensive investigations concluded, however, that the full CP procedure is rigorously correct for closed-shell interactions and that the overcorrection argument is fundamentally wrong.[104−109] We strongly discourage following any suggestion not to apply the CP procedure (or to consider only a fraction of BSSE) at correlated levels based on claims that CP worsens the agreement with some reference (experimental) data or expected trends. These disagreements are always due to other substantial inaccuracies in the calculations opposing the BSSE, namely, the insufficient size of the basis set used for electron correlation. When the CP correction is not applied, some low-quality calculations may seem closer to experiment, but this compensation of errors may not work the same way for other systems. The magnitude of BSSE differs for small and large complexes, depends on the structure (e.g., is different for H-bonded and stacked NA base dimers), and is highly dependent on the basis set. Let us note that in some cases the BSSE-corrected correlation interaction energy can be close to zero or even repulsive. It does not indicate any overcorrection of the CP procedure; it is a correct result caused by the above-mentioned reduction of electrostatic attraction when considering the correlation of electron motions.[79]

The full CP procedure is straightforward when making a single-point interaction energy calculations, and most codes allow application of the correction routinely. The structure of larger clusters cannot be, however, determined using the point-by-point method and gradient optimization is to be used. Full gradient optimization methods relax simultaneously the intra- and intermolecular degrees of freedom. However, currently none of the commercially available codes allow correction of the BSSE during an optimization. Thus the complexes are optimized using the standard supermolecular gradient optimization and the BSSE correction is applied only to the subsequent evaluation of interaction energies, i.e., after the optimization is finished. This is certainly not correct, since the structure of a complex is optimized on the standard PES and not on the theoretically more justified CP-corrected PES. The problem of CP-corrected gradient optimization was solved recently[110] and the respective method allows evaluation of the CP-corrected gradient and Hessian of a complex at any *ab initio* level. The first calculations[110−115] showed that the CP-corrected and standard optimizations might lead to different structures and demonstrated the necessity of using the CP-corrected optimization.

The stabilization energy of a complex is also influenced by the deformation of the monomers upon the formation of the complex. This can be accounted for by subtracting the energy of the optimized isolated monomers (E_{Ai}, E_{Bi}) from the energies of the monomers in the dimer geometry (E_{Ad}, E_{Bd}). The respective deformation energy E^{DEF} is always repulsive:

$$E^{DEF} = (E_{Ai} - E_{Ad}) + (E_{Bi} - E_{Bd}) \quad (3)$$

The total complexation energy ΔE^{T} is thus defined as

$$\Delta E^{T} = \Delta E + E^{DEF} \quad (4)$$

The inclusion of deformation energy is required only if the interaction energy is determined using the standard gradient optimization (taking all degrees of freedom into account) and adopting "a posteriori" the Boys and Bernardi procedure.[103] Using the theoretically more justified CP-corrected gradient optimization, the deformation energy is fully covered in the final interaction energy.

II.2. Correlation Energy

The correlation energy is much smaller than the HF energy (it amounts to a few percent of the total electronic energy); nevertheless, the role of correlation interaction energy is topical and cannot be neglected. The first treatment really applicable to large systems is the Møller−Plesset (MP) perturbational theory.[116] The correlation energy is determined as a sum of the second, third, fourth, and higher contributions:

$$\Delta E^{COR} = \Delta E^{2} + \Delta E^{3} + \Delta E^{4} + \cdots \quad (5)$$

The most economical and thus widely used is the simplest version of the MP method, the second-order MP theory (MP2). It is applicable to extended complexes (up to several dozens atoms) and gives surprisingly good estimates of the correlation energy. This is mainly due to mutual compensation of the neglected higher order contributions. It is known[117] that ΔE^{MP2} is overestimated for some clusters, while for other clusters it is close to the actual value of the correlation interaction energy [mostly refers to reference CCSD(T) calculation]. If the MP theory is performed up to the fourth order of perturbational treatment, then a substantial part of the correlation energy is included, though it is very demanding considering computer time, disk space, and memory. The main problem connected with the use of MP4 and higher levels of perturbational theory concerns con-

vergence of the expansion, which may oscillate. A practical solution to this problem is the coupled-cluster (CC) theory.[118−122a] This method includes a larger portion of the correlation energy compared to the MP methods and currently it seems to be the most promising tool for accurate calculations of the correlation energy. The original CC method includes all possible single and double electron excitations (CCSD). In the language of the perturbational theory, single and double excitations are covered up to the infinite order, and the same is true for a part of the quadruple and hexatuple excitations. The next step in improving[119,120] the CCSD method lies in considering the effect of triple excitations, because these are of topical importance for molecular clusters. The CCSDT method calculates single, double, and triple excitations iteratively, and all these contributions are therefore included up to the infinite order. The CCSDT correlation energies are close to those obtained from the full CI method; the computer demands of the CCSDT calculations are, however, prohibitively high, and the method is impractical for large clusters, such as NA base pairs. A compromise between economy and accuracy are the CCSD + T(CCSD)[121] and CCSD(T) methods,[122a] which evaluate triple excitation in a noniterative way after performing a CCSD calculation. The CCSD(T) method also includes some fifth-order contributions. The CCSD(T) method is iterative but nonvariational, size-consistent, and represents the most robust tool for calculation of the correlation energy of large systems including the NA base pairs. Because of the importance of triple electron excitations for molecular interactions,[122b] we do not recommend the use of either MP4SDQ or CCSD methods.

The use of all the methods mentioned is based on the assumption that a single reference description is possible; this requirement is, indeed, fulfilled in the case of NA base pairs.

II.3. Density Functional Theory

Density functional theory (DFT)[123] does not include any empirical parameters and thus belongs to the family of *ab initio* methods. In the DFT approach, the exact exchange term used in the Hartree−Fock method is replaced by a more general expression: the exchange-correlation functional. The simplicity of the DFT stems from the fact that it uses a functional of electron density to model exchange and correlation. Real progress in the application of DFT to molecular systems began with the introduction of nonlocal (gradient) corrections.[124] What makes the use of DFT attractive in the realm of large systems is the fact that DFT energy includes, besides the exchange contributions, also some portion of the correlation energy. Further, DFT requires shorter computer times when compared with conventional *ab initio* methods with inclusion of electron correlation. Care should be taken, however, when applying DFT to molecular clusters where the dominant part of the stabilization energy comes from the dispersion energy. There are several papers demonstrating[80,125−130] that this energy contribution is not covered in the DFT interaction energy, which leads to a large underestimation of the stabilization. Other DFT functionals may dramatically fail in the completely opposite direction for some other reasons. Thus, the use of DFT theory for intermolecular interactions is very insecure and there is currently no DFT method capable of studying, for example, base stacking. The widely used B3LYP and BLYP methods fail completely for base stacking.[80] Considerable effort has recently been devoted to develop DFT techniques which will allow description of van der Waals complexes,[129,130] while fast methods are being developed to treat large fragments of nucleic acids.[131−134] It seems, however, that inclusion of dispersion energy in the DFT techniques in a consistent way will be very difficult if not practical at all. One of the possibilities for solving the problem is to add an empirical dispersion energy to the DFT energy.[80]

It must be kept in mind that the DFT technique is the only *ab initio* method which allows study of larger fragments of nucleic acids. The basic advantage of a variational *ab initio* method over an empirical potential approach is the fact that besides structure and energy also the wave function is generated. Further, *ab initio* techniques cover many-body interactions and charge transfer. These effects are not included in the currently available empirical potentials, though the first successful attempts to cover them already exist.[135,136]

Two new promising DFT techniques introduced recently should be mentioned in this respect. The first is the self-consistent charge density functional tight binding (SCC-TB) method. [133] This method is derived from the DFT theory by a second-order expansion of the DFT total energy functional with respect to the charge density fluctuations at a given reference density. The second approach, called SIESTA,[134] utilizes a linear scaling, fully self-consistent DFT method with standard norm-conserving pseudopotentials and flexible LCAO basis sets. The SIESTA method presently allows consideration of fragments of DNA with as many as 700 atoms. The former method is less time-consuming, and even larger fragments of DNA can be calculated. The possibility of studying larger fragments of DNA by the DFT technique is quite important, since none of the currently available semiempirical quantum-chemical methods (such as AM1 and PM3) is suitable for studies of complexes of nucleic acid bases.[98,99] The approximations used in these methods do not allow any reasonable description of nucleic acid interactions, not only for stacked structures but also for H-bonded ones.[98,99]

II.4. Vibrational Frequencies

Harmonic vibrational frequencies are evaluated for the optimized structures using Wilson FG analysis.[137] The respective Hessians are obtained from the standard and quite recently also CP-corrected calculations. However, the vibrational frequencies and especially the intermolecular vibrational frequencies of any molecular complex including NA base pairs are not harmonic, and anharmonic effects should be considered, although it is really demanding. While harmonic vibrational frequencies are routinely ob-

tained by any *ab initio* code, multidimensional anharmonic treatment requires the use of state of the art methods and is only rarely reported in the literature.[74,138]

Recently two multidimensional anharmonic analyses have been performed on systems relevant to the scope of this review. First, intermolecular vibrational modes of the H-bonded adenine···thymine Watson−Crick base pair were obtained[138] by solving the vibrational Schrödinger equation. Relying on a Born−Oppenheimer-like separation of the fast and slow vibrational motions, the complete multidimensional vibrational problem was reduced to a six-dimensional subproblem in which all rearrangements between adenine and thymine can be described. The potential energy surface was determined at the *ab initio* Hartree−Fock level, and an analytical potential energy function was obtained by fitting to the *ab initio* data. This function was used to calculate vibrational energy levels, disregarding the role of the kinematic and potential (in-plane)−(out-of-plane) interactions. The original six-dimensional space is thus split into two three-dimensional subspaces. The eigenvalue problems were solved variationally by diagonalizing the vibrational Hamiltonians in the form of matrixes over basis set functions which were expressed as products of the eigenfunctions of the corresponding uncoupled one-dimensional Schrödinger equations.

The other problem treated was the aniline amino group motion.[74] The nonharmonicity of the amino group motion of aniline (modeling the NA bases) was investigated since direct spectroscopic data are available only for aniline and not for NA bases. Inversion splitting of aniline was evaluated by solving a two-dimensional vibrational Schrödinger equation for large-amplitude inversion (including the nitrogen out-of-plane motion) and torsion motions, while respecting the role of small-amplitude C−N stretching and H−N−H bending motions. A five-dimensional aniline potential energy surface was described at the Hartree−Fock level with the 6-31G* basis set, since for this particular system the HF/6-31G* data are in good agreement with estimates obtained at MP2 and CCSD(T) levels with extended basis sets.[74]

II.5. Gas-Phase Thermodynamic Characteristics of Nucleic Acid Base Pairs

Gas-phase thermodynamic characteristics of NA base pairs can be evaluated in two ways.

Statistical Thermodynamic Treatment[100]

The equilibrium constant (K_T) of the formation of NA base pair at temperature T is related to the standard change in the Gibbs energy, $\Delta G°$, by the following equation:

$$\Delta G°_T = -RT \ln K_T \qquad (6)$$

The $\Delta G°$ term can be determined from knowledge of the enthalpy and entropy of the base pair formation, $\Delta H°$ and $\Delta S°$, using the usual equation

$$\Delta G°_T = \Delta H°_T - T\Delta S°_T \qquad (7)$$

The thermodynamic functions of the formation of NA base pairs were evaluated from the total partition function given as product of translational, rotational, and vibrational partition functions. Partition functions were evaluated within the rigid rotor−harmonic oscillator−ideal gas approximation and were computed from the Hartree−Fock constants (equilibrium geometry and harmonic vibrational frequencies).

Molecular Dynamic Calculations[101]

This state-of-the-art technique has been applied for uracil dimer. The populations of various structures of the dimer were extracted from long runs of NVE microcanonical ensemble molecular dynamics simulations combined with the quenching method. The potential energy surface of the dimer was determined with the Amber potential using the force field of Cornell et al.[139] Constant-energy MD simulations were performed assuming rigid uracil monomers (quaternion formalism). A 1 fs integration step was used. In the course of the MD simulation, after a certain number of steps, the simulation was stopped and minimization using the Amber potential was performed: Intermolecular and intramolecular coordinates were optimized. The optimal energy and coordinates were stored, and the MD started again from the point where it was stopped. Populations of various structures were determined after 1 μs; quenches were made after each 10 ps. Convergence was checked by dividing the MD simulation into five parts while the population distribution for each part was calculated separately. The calculation of populations (proportional to ΔG values) from quenching is possible in a rather narrow energy interval. The energy should be sufficiently high to allow a high frequency of interconversions among different isomers. The basic advantage of this technique is that it allows determination of the relative populations for all structures on the PES, i.e., to pass from the PES to FES. In the case of NA base pairs, it allows comparison of the relative populations of H-bonded and stacked structures.

III. Structure and Properties of Isolated NA Bases

This review paper deals with various aspects of H-bonding and stacking of NA bases. Nevertheless, the molecular interactions depend on properties of isolated NA bases. Let us, therefore, first comment on the most important properties of isolated bases and briefly summarize the current status of quantum chemical theory in this area.

III.1. Tautomeric Equilibria of Bases

Different tautomers of bases are obtained when considering different hydrogen positions around the base.[140] Rare tautomers of bases are assumed to participate in some biochemical processes including point mutations.[81] A cytosine imino tautomer has recently been proposed to stabilize some three-dimensional structures of nucleic acids.[46,82,93,141] Tautomers are rarely observed in oligonucleotide crystals[142] and for most biochemical processes probably only major tautomers of bases are involved. There

are several reasons for that. First, there are nucleobases such as uracil or thymine for which there is a very large energy gap between the major form and minor tautomers. For some other bases (guanine, cytosine) there are several energetically acceptable tautomers. However, the major tautomer forms are still the only ones that appear in nucleic acids under normal circumstances. Many rare tautomers are destabilized by solvent effects, or they do not lead to a pairing compatible with the nucleic acids architecture. The N_7H_7 tautomer of guanine cannot be involved because there is a sugar attached to N_9 in DNA (the other position competing with N_7 in gas-phase experiments). It is reasonable to assume that the minor tautomers were eliminated during evolution processes to ensure the stability of the genetic code.

Quite an interesting possibility is a formation of so-called metal-cation-assisted tautomers of bases frequently observed in crystals of metalated bases and nucleotides.[143−147a] In the gas phase, the charge carried out by the metal group is crucial to influence the tautomerism of bases and protonation or deprotonation processes. However, in a polar solvent and in crystals, the electrostatic effects can be completely suppressed by the solvent screening and counterions. Therefore, comparison of the gas phase *ab initio* data with solid state and condensed phase experiments is not straightforward, especially for ionic systems. The gas phase and experimental trends can often be opposite. Nevertheless, the electrostatic effects can, for example, rationalize that positively charged Pt^{II} adducts attached to N_7 of guanine frequently lead to deprotonation of N_1 of guanine in crystals, leading to mispairs.[144] The reason is that the charge of +2 residing on the metal group enhances the aciditity of the N_1 site of guanine. This effect is very strong in the gas phase and has nothing to do with the specific covalent bond between Pt and N_7. It is a long-range electrostatic effect, and all positively charged metal groups will exert a similar influence on the N_1 acididity. When the H_1 hydrogen of guanine is transferred to a ring nitrogen of the other base in the mispair, the ion-pair structure is formed and it has better electrostatic interaction. In solid-state and condensed-state experimental situations, the extent to which the electrostatic contribution is expressed depends on the environment of the metalated base. Similarly, when negatively charged $ZnCl_3^-$ is attached to the same position of adenine, the adenine is protonated at the N_1 position, which also leads to mispairing.[147] The rationalization of why this cation binding can promote a protonation of adenine can be similar. Since the $ZnCl_3$ group bears a formal charge of −1, its binding increases the basicity of the N_1 position of adenine, making its protonation more likely.[147b] On the other hand, protonation of N_3 of cytosine and N_1 of adenine has been reported after a replacement of an amino group proton by a metal adduct with charge +1, hinting at possible nonelectrostatic changes of electronic structure of bases due to the metalation.[145,146]

Theoretical evaluation of gas-phase tautomeric equilibria of nucleobases is not difficult and numerous papers can be found in the literature. Let us reference a few recent papers which we have found notable, some of them combining theoretical and experimental approaches.[148−162] Semiempirical techniques are unsuitable to study tautomerism. Calculations employing the MP2 level of theory with medium-sized polarized basis sets of atomic orbitals provide quite good results for tautomeric equilibria of bases. The ability of density functional theory methods to predict tautomeric equilibria of bases seems to be limited. As usual with DFT, the results depend on which functional has been used. Currently available reference studies show that inclusion of higher electron correlation contributions with inclusion of triple electron excitations and large basis sets of atomic orbitals are desirable to guarantee quantitative predictions.[154,159,161] Recent developments in this field were reviewed by Leszczynski.[162]

Tautomeric equilibria of nucleobases can be greatly influenced by solvent effects. Realistic inclusion of solvent is quite difficult[163] and thus a rather small number of papers dealing properly with this subject can be found.[164−166] It is assumed that the intramolecular proton transfer leading to a formation of tautomers is mediated by water molecules.[164] A series of papers combining high-level *ab initio* calculations with state-of-the-art solvent models (good-quality quantum chemical self-consistent reaction field technique, classical Monte Carlo free energy perturbation technique, and Poisson−Boltzmann calculations to estimate the effect of DNA environment) were reported by Orozco, Luque and co-workers.[165,166] They convincingly ruled out that the mutagenic properties of substituted uracils (5-bromouracil) can be due to a formation of mispairs involving the enol form of uracil.[166] This contradicts a widespread and likely incorrect opinion in the biological literature about the mechanism of the mutagenetic action of 5-substituted uracils.[167]

III.2. Protonization Energies of Bases

Protonated bases appear more frequently in nucleic acids compared to the minor tautomers.[28,30−32,168−173] In fact, even the above-mentioned metal-assisted tautomers of bases should be in many cases considered protonated metalated bases rather than tautomers. Protonated cytosine stabilizes the three-dimensional architecture of four-stranded i-DNA and appears in pyrimidine−purine−pyrimidine triplexes (where it causes a marked destabilization if consecutive protonated bases are present). i-DNA is one of the most exciting systems for studies of molecular interactions.[46] Its cytosine-rich system consists of consecutive, protonated cytosine−cytosine pairs. There is a fast proton transfer and/or delocalization of the inner (H_3) proton in the pair, since NMR and X-ray data show that the cytosines in the pair are equivalent.[28,30−32] Further, this molecule has repulsive net stacking interactions which contrasts all other nucleic acid forms.[46,82]

Gas-phase protonization energies are relatively easy to calculate,[82,174,175] though they do not necessarily reflect the actual protonation in nucleic acids. While NA bases should not be protonated at neutral

pH, the molecular interactions of the environment can change the free energy of protonation drastically in favor of the protonated structure.[82] Bases can then be protonated around neutral pH in DNA because a protonated base forms exceptionally stable base pairs with highly polar neutral bases.[82,176] Protonation of cytosine is favorable compared to protonation of adenine. However, when a protonated adenine··· cytosine base pair is formed in the DNA, the proton is transferred to adenine to optimize the molecular ion−molecular dipole interaction,[82] as has also been evidenced by experimental data.[177]

III.3. Electrostatic Potential around NA Bases

The electrostatic field around NA bases is of primary importance for all molecular interactions of bases, since it governs the electrostatic interactions.[80,82,84,178−180] Electrostatic interactions determine the strength of H-bonding, hydration, and dominate also the binding of small or polyvalent cations to bases.[181−187] What experimentalists call the basicity of various sites of bases can be rationalized through molecular electrostatic potential calculations. It should nevertheless be noted that cation binding is associated with significant polarization and charge-transfer effects which are responsible for very specific differences among the cations and determine their different biological roles.[181−189] Electrostatic interactions even govern orientation of bases in stacked complexes, although their stability originates in dispersion attraction.[84]

Inclusion of electron correlation is vital for accurate evaluation of the electrostatic potential. Hartree−Fock approximation overestimates the polarity of the bases; further, the ratio of the HF data with respect to the correct values is not uniform for different bases. Thus there is no unique scaling factor applicable. As demonstrated by our earlier calculations,[80,84] a medium-sized polarized basis set provides already quite converged dipole moments of bases, molecular electrostatic potentials (MEP), and MEP-derived charges (also called ESP charges); the respective data are given in Table 1. Atomic charges and dipole moments obtained with extended aug-cc-pVDZ basis set (containing diffuse s-, p- and d-functions) represent benchmark data. From Table 1 it is also evident that charges and dipole moments obtained with the considerably smaller 6-31G* (0.25) basis set differ from previous ones only slightly, which gives confidence to the use of the 6-31G* (0.25) basis set (see also section IV.3).

A very important result bolstering the current molecular modeling tools was obtained recently through extensive comparison of empirical potentials and full *ab initio* data.[84] It shows that atom-centered point charges derived from MEP provide a very reasonable approximation of MEP. Therefore, no additional charge centers (π, lone-pair charges, etc.) are required to describe base stacking.[80,84]

The most salient properties of the electric field of bases are magnitude and the direction of the molecular dipole moment.[80,84,190−193] Obviously, the electric field of bases is not represented by the dipole only, and for proper treatment of the interactions, MEP must be considered.

Table 1. Atomic Charges and Dipole Moments of NA Bases[a,b] Derived by Fitting to the MP2 Molecular Electrostatic Potential Determined

base	basis set	atomic charges for various atoms																μ(D)
cytosine		C_2	N_1	C_6	C_5	C_4	N_3	O_2	N_4	H_1	H_6	H_5	H_{41}	H_{42}[c]				
	6-31G*(0.25)	0.881	−0.601	0.207	−0.653	0.970	−0.757	−0.577	−1.067	0.354	0.129	0.224	0.442	0.449				6.27
	aug-cc-pVDZ	0.967	−0.634	0.237	−0.740	1.055	−0.817	−0.619	−1.100	0.359	0.138	0.244	0.451	0.458				6.49
guanine		O_6	C_6	N_1	C_2	N_3	C_4	C_5	N_7	C_8	N_9	N_2	H_1	H_8	H_9	H_{21}	H_{22}[c]	
	6-31G*(0.25)	−0.499	0.602	−0.747	0.841	−0.654	0.394	0.051	−0.498	0.236	−0.434	−0.977	0.340	0.091	0.356	0.420	0.419	6.45
	aug-cc-pVDZ	−0.500	0.524	−0.710	0.856	−0.713	0.480	0.194	−0.593	0.302	−0.602	−1.053	0.393	0.105	0.415	0.458	0.444	6.65
adenine		N_3	C_2	N_1	C_6	C_5	C_4	N_9	C_8	N_7	N_6	H_8	H_2	H_9	H_{61}[c]	H_{62}		
	6-31G*(0.25)	−0.659	0.452	−0.662	0.638	0.013	0.538	−0.503	0.277	−0.522	−0.844	0.047	0.089	0.369	0.392	0.375		2.55
	aug-cc-pVDZ	−0.715	0.465	−0.694	0.603	0.128	0.591	−0.568	0.243	−0.579	−0.897	0.072	0.129	0.402	0.421	0.399		2.56
uracil		O_4	C_4	N_3	C_2	N_1	C_6	C_5	O_2	H_5	H_3	H_1	H_6					
	6-31G*(0.25)	−0.519	0.744	−0.571	0.681	−0.455	0.106	−0.481	−0.519	0.198	0.340	0.334	0.144					4.10
	aug-cc-pVDZ	−0.555	0.817	−0.609	0.749	−0.505	0.133	−0.554	−0.554	0.222	0.349	0.351	0.155					4.37
thymine		O_4	C_4	N_3	C_2	N_1	C_6	C_5	O_2	C_M	H_3	H_1	H_6	HM_1	HM_2	HM_3		
	6-31G*(0.25)	−0.487	0.586	−0.573	0.666	−0.475	−0.059	0.055	−0.520	−0.576	0.354	0.351	0.173	0.177	0.151	0.177		4.01
	aug-cc-pVDZ	−0.531	0.683	−0.618	0.744	−0.520	−0.043	−0.041	−0.559	−0.474	0.361	0.364	0.186	0.158	0.133	0.158		4.31

[a] Cf. Figure 1. [b] Taken from Šponer et al. *J. Phys. Chem.* **1996**, *100*, 5590. [c] This hydrogen atom is H-bonded in the Watson−Crick base pairs.

Guanine and cytosine are very polar bases. Therefore, the GC base pair is also very polar. Thymine has smaller dipole moment and adenine is even less polar (cf. Table 1). As a result, the whole AT base pair is much less polar than the GC one. Hypoxanthine (=6-oxopurine, i.e., guanine lacking an amino group. This base is sometimes called inosine, though inosine should rather be used for the whole nucleotide) has the same orientation of dipole as guanine, though the dipole is somewhat reduced.[79] 2-Aminoadenine has an exceptionally small dipole moment.[79] Sometimes the simplest consideration of polarities of bases is very helpful and can eliminate widespread misconcepts. For example, one should never argue that the number of H-bonds determines the strength of a base pair. The 2-aminoadenine··· thymine base pair is almost two-times less stable than the GCWC base pair (see later discussion), despite having three H-bonds.[79] Nonpolar 2-aminoadenine does not form strong H-bonds. The GG1 base pair (see below) is more than twice as stable as the GG4 base pair, despite the fact that both base pairs have two H-bonds. It is again due to the orientation of molecular dipoles. Comparison of G···C WC, A···T WC, and I···5metC WC base pairs (I = hypoxanthine; see above) indicates why the latter base pair behaves similarly to AT under certain circumstances and similarly to the GC base pair in other situations. The I···5metC base pair has a similar van der Waals contour as the AT base pair, including lack of the sterically important minor groove amino group. On the other hand, its electric properties are very close to those of the GCWC base pair.[79] Therefore, it can also behave similarly to the GCWC base pair. That behavior is more likely if the I···5metC pair is surrounded by polar GC base pairs or consecutive I···5metC base pairs run along the double helix, because then the electrostatic similarity between the GC and I···5metC appears. The electrostatic polarity of I···5metC is expected to be much less visible when surrounded by less polar AT base pairs. This simple consideration correlates well with recent studies on the influence of the I···5metC base pair on the DNA structure and bending propensity in B-DNA.[194] The crystal structure investigation[194] indicates that while the mixed AIA sequence has similar properties as the net A-tract with high propeller twist, the I-tract region shows low propeller and variable stacking. This was originally rationalized (besides the shape similarity between I···5mC and A···T) by considering the interactions of exocyclic groups.[194] However, one can also imagine that I···5mC base pair stacked between two A···T base pairs behaves similarly to A···T base pairs, not only because of the molecuar shape similarity between AT and I···5mC but also due to a reduced role of electrostatic interaction in this sequence. This is because A···T base pairs are rather nonpolar, so the high polarity and GC-like electrostatics of the I···5mC base pair is not expressed when surrounded by AT pairs on either side. On the other hand, I-tract sequences should have very similar and strong electrostatic part of the stacking interaction, like the sequence of consecutive guanines (G-tract). This could prevent A-tract twist geometry even though it would be sterically affordable as for the A-tract.

III.4. Polarizability

Another important property of bases is their polarizability.[193,195,196] Polarizability influences the induction and dispersion contributions to the perturbation dimerization energy. Proper evaluation of polarizability is very difficult, and inclusion of electron correlation and extended diffuse basis sets is required. Current reference values of polarizabilities of bases are those reported by Cybulski et al.[193] It should be mentioned, however, that the role of models based on polarizabilities for estimating the accuracy of interaction correlation energy cannot be overestimated. Usually, only a dipole polarizability is determined, while the correlation interaction energy includes not only the second-order dispersion energy but also higher order dispersion contributions which are proportional to higher order polarizabilites. Further, the intersystem correlation energy is affected (mostly partially compensated) by the intrasystem term. Finally, the widely used MP2 procedure seems to overestimate the aromatic stacking stabilization compared to the more accurate CCSD(T) data (cf. section IV.3). Medium-sized diffuse basis sets therefore yield reliable MP2 aromatic stacking stabilization energies.[196]

III.5. Gas-Phase Electron Affinities (EA) and Ionization Potentials (IP)

IP and EA of bases can be obtained through molecular beam experiments.[35−37] Such experiments can be directly compared with *ab initio* results.[197,198] Proper evaluation of these properties of NA bases is not easy and there still are doubts in the literature concerning their values. Evaluation of EA and IP is quite important to the understanding electron-transfer processes along the DNA chain and how they are influenced (terminated) by proton transfer between bases in base-pair radical cations. These processes are related to radiation damage of DNA.[199−206] Base pair radical cations are not reviewed in detail in the present paper and the reader can find enough information in the cited literature, mainly in the thorough study of Bertran et al.[202] and the review by Colson and Sevilla.[200]

III.6. Nonplanarity of DNA Bases

For many years isolated nucleic acid bases were believed to be perfectly planar. However, recent *ab initio* studies demonstrated that the amino groups of bases are nonplanar due to the partial sp^3 hybridization of the amino group nitrogen atom. *Ab initio* calculations carried out at the HF level with polarized basis sets of atomic orbitals provided the first respectable suggestion that a weak nonplanarity of amino groups of bases could occur.[71,207−209] Nevertheless, it was still believed that the effect was very weak and biologically unimportant. This view changed when unexpectedly abundant close mutual interstrand amino groups contacts were found in B-DNA crystal structures.[75,210] This observation contradicted

Table 2. Nonplanarity of Amino Group of Isolated NA Bases Determined at Various Theoretical Levels (data taken from Šponer et al. *J. Biomol. Struct. Dyn.* 1996, *14,* 117)

base	level	$XCNH_1$[a]	$XCNH_2$[b]	HNX[c]	ΔE[d] (kcal/mol)
cytosine	HF/6-31G**	−5.5	3.3	359.5	0
	MP2/6-31G*	−26.2	14.1	348.8	−0.38
	MP2/6-311G(2df,p)	−21.4	12.6	351.9	−0.15
adenine	HF/6-31G**	−5.0	4.6	359.6	0
	MP2/6-31G*	−21.1	18.7	349.3	−0.34
	MP2/6-311G(2df,p)	−15.3	16.5	352.9	−0.13
guanine	HF/6-31G**	−11.1	28.6	348.6	−0.34
	MP2/6-31G*	−11.8	43.2	338.1	−1.63
	MP2/6-311G(2df,p)	−13.3	39.2	339.6	−1.12
isocytosine	HF/6-31G**	−10.8	27.1	349.6	−0.27
	MP2/6-31G*	−12.3	40.2	340.2	−1.22
	MP2/6-311G(2df,p)	−13.6	35.7	342.3	−0.78

[a] Amino group dihedral angle; X = C_5 of cytosine and adenine and N_3 of guanine and isocytosine. [b] Amino group hydrogn dihedral angle; X = N_3 of cytosine and N_1 of adenine, guanine, and isocytosine. [c] Sum of the HNC and HNH amino group hydrogen valence bond angles. This quantity equals 360° for a planar molecule and is less than 360° if the amino group is pyramidal. [d] The inversion barrier for pyramidalization, i.e. the energy difference between nonpolar and polar optimized molecules.

the usual view assuming that amino group contacts are purely repulsive steric clashes,[15,16] and it was suggested that amino group pyramidalization allows or stabilizes these contacts.[75] Simultaneous quantum chemical calculations including electron correlation effects have shown that the pyramidalization was underestimated by the earlier SCF calculations.[71−75] It has then been demonstrated that the very flexible amino groups participate in a number of interactions in nucleic acids. Proper treatment of amino group nonplanarity needs inclusion of electron correlation effects. From Table 2 it is evident that amino group nonplanarities obtained at the HF level are underestimated. Correlated calculations with medium-sized polarized basis sets of atomic orbitals somewhat overestimate the pyramidalization. Current reference values are MP2/6-311G(2df,p) optimizations (cf. Table 2).[69] MP2 and CCSD(T) calculations are assumed to provide the same results for amino group nonplanarity.[74]

Interestingly, the pyramidalization of guanine is significantly more pronounced compared to that of adenine and cytosine.[73] Further, the pyramidalization of the guanine amino group is highly asymmetrical due to the repulsion between the H_1 ring hydrogen atom and the adjacent amino group hydrogen atom. Some asymmetry is observed also for the cytosine amino group due to the proximal $H_5(C_5)$ hydrogen atom.

The amino group hydrogens can participate in out-of-plane H-bonds, where the hydrogens are bent away from the molecular plane of bases.[83] Further, the amino group nitrogen atom can serve as a weak H-acceptor.[75,211] The strongest amino acceptor identified so far was found in studies of complexes between hydrated cations and base pairs, where a polarized water molecule from the hydration shell donates hydrogen to the amino group nitrogen atom of adenine.[187] Amino acceptor interactions of bases were identified in crystals of nucleobases.[211] Out-of-plane and amino acceptor interactions can be arranged in a concerted manner, since bending of the hydrogens supports the sp^3 pyramidalization, strengthens the negatively charged region above the nitrogen atom, and eases the access of a H-donor to the nitrogen.[83] Clear direct experimental evidence about the nonplanarity of isolated bases is still missing due to the resolution of the available experimental techniques, although there is a certain indirect evidence from different experiments supporting the quantum-chemical data[211−213] (see later discussion).

The high-level calculations described above give clear evidence that amino groups in NA bases are nonplanar. Nevertheless, due to the potential biological importance of interactions involving nonplanar amino groups of bases, it is still highly desirable to verify the theoretical predictions on the amino group nonplanarity by comparison with a gas-phase experiment. Such experiments are not available for NA bases, but good-quality spectra were reported especially for the aniline molecule.[214−216] The geometry characteristics are, however, not available, and their determination from the most accurate spectroscopic experiments requires a solution of the pertinent internuclear dynamics Schrödinger equation. Thus geometries are obtained from the spectroscopic data in an indirect way by calculating the spectra and assuming a number of serious approximations. It must be stressed at this point that the harmonic approach is not justified for the amino group pyramidalization, since the potential opposing the amino group rocking vibration is of a double-minimum character. Moreover, due to a rather strong coupling of the rocking mode with the remaining modes, the amino group vibrational dynamics may be tractable only by means of multidimensional models. The nonharmonicity of the amino group motion has been almost completely ignored in theoretical studies. Only recently a one-dimensional study of the amino group rocking motion for 2-aminopyrimidine has been reported.[217] As mentioned above, because of the strong coupling of rocking motion with other vibrational modes, one-dimensional modeling may not be adequate.

The aniline molecule has 36 vibrational degrees of freedom, two of which correspond to large-amplitude motions (inversion and torsion). A complete dynamic study of such a system is clearly unfeasible. We have found[74] that the full vibrational problem of aniline may be reduced to a two-dimensional problem for the inversion (including the nitrogen out-of-plane motion) and torsion motions parametrically depending on two

Table 3. Calculated and Experimental Inversion−Torsion Frequencies for Aniline and Aniline-ND_2[a]

frequency (in cm^{-1})	aniline		aniline-ND_2	
	calcd	exp[b]	calcd	exp[b]
ν_i[c]	43	41	10	13
$2\nu_i$	444	424	341	338
ν_t[d]	298		209	
$\nu_t + \nu_i$	332	304	217	
$\nu_t - \nu_i$	256	237	199	

[a] Taken from Bludský et al. *J. Chem. Phys.* **1996**, *105*, 11042. [b] For references see the text. [c] Inversion frequency. [d] Torsion frequency.

small-amplitude displacements describing the C−N distance and H−N−H angle. All the remaining small-amplitude motions were kept frozen in the calculations. The five-dimensional PES calculated at the HF/6-31G* level of theory was used to fit the potential energy function. The HF/6-31G* method was selected, because for this particular molecule, it provided data very close to the results of CCSD(T) and MP2 calculations with extended basis sets of atomic orbitals.[74] The calculated and experimental inversion−torsion frequencies for aniline and aniline-ND_2 are given in Table 3. Evidently, the agreement between theoretical and experimental data for fundamental, overtone, and combination modes is very good. This may be partially due to compensation of errors (the size of the basis set and inclusion of correlation energy). Nevertheless, the very good agreement between the theoretical and experimental inversion−torsion frequencies for aniline as well as for aniline-ND_2 gives us a clear message that the theoretical HF/6-31G* PES is reliable. Let us recall that the vibrational frequencies are very sensitive to the quality of the PES. From the PES we can easily extract values of barrier height (534 cm^{-1}) and equilibrium inversion angle (43.1°), which should be very close to the actual values. The good match between experimental and theoretical inversion frequencies for aniline, and aniline-ND_2 and the similarity of the HF, MP2, and CCSD(T) aniline inversion barriers and amino group nonplanarities give us confidence that the MP2 inversion barriers and amino group nonplanarites calculated with large basis sets for NA bases at the MP2 level[69,74] are close to their actual values, which are experimentally unknown.

IV. Structure, Energetics, and Properties of NA Base Pairs

IV.1. H-Bonded NA Base Pairs

Let us now discuss the gas-phase structures and energies of H-bonded base pairs. The geometries of H-bonded NA base pairs (Figure 2) were determined using gradient optimization at the HF level with the 6-31G** basis set first under C_s symmetry[79] and later without any geometrical constraint.[83] Stabilization energies were determined for the optimized structure using correlated MP2 level with 6-31G*(0.25) basis

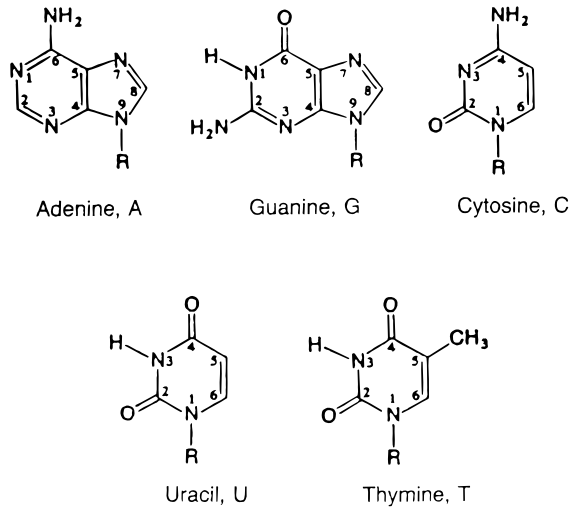


Figure 1. Structures of isolated NA bases.

set with diffuse d-polarization functions with an exponent of 0.25.[79] The MP2 stabilization energy was corrected for the BSSE and for the deformation energy of monomers. Interaction energies and their components are summarized in Table 4. Let us point out that the interaction energies and geometries in the original paper[79] are reported for structures obtained assuming C_s symmetry for all base pairs. Table 4 in this review complements the previously published reference interaction energy data[79] and presents MP2 energy data re-evaluated for the nonplanar base pairs.[83]

The total MP2/6-31G*(0.25)//HF/6-31G** stabilization energies of the base pairs range from 23.8 to 9.4 kcal/mol. The GCWC base pair with three H-bonds is the most stable pair, while one of the adenine··· adenine pairs with two H-bonds is the weakest one. The stability of the GG1 pair is surprisingly large. This pair having two H-bonds is almost as stable as the GCWC one with three H-bonds. As explained above, the number of H-bonds is not the only factor determining the stability of base pairing; the magnitude and mutual orientation of molecular dipole moments should also be considered. The stabilization of H-bonded pairs originates from the Hartree−Fock contribution. This is due to the electrostatic nature of stabilization in the H-bonded pairs (electrostatic energy is included in the HF stabilization energy). This finding is not surprising in light of the rather large dipole moments of NA bases (cf. Table 1). The positive (destabilizing) value of ΔE^{COR} found for the GG1 pair is due to the fact that the repulsive contribution caused by reduction of the dipole moments of monomers due to electron correlation is larger in absolute value than the dispersion attraction. Nevertheless, it can be assumed that when performing the optimization at the MP2 level and considering the larger basis set of AO, the correlation interaction energy would be attractive even in this case. The correlation contribution is important especially for the weaker pairs, where it amounts to 30−40% of the total stabilization energy. However, neglecting the correlation contributions would not change the overall stabilization of H-bonded pairs dramatically, mainly for the most stable pairs. The

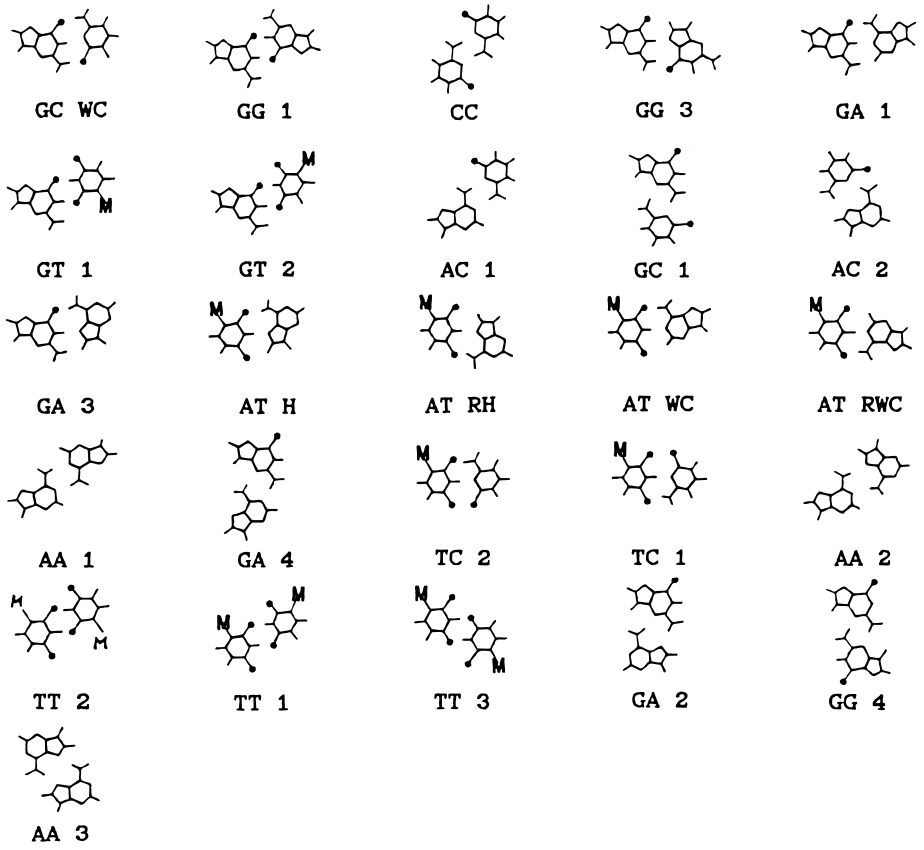


Figure 2. Structures of H-bonded NA base pairs; points represent oxygen atoms and M means methyl group.

Table 4. Calculated Interaction Energies and Their Components (in kcal/mol) of H-Bonded NA Base Pairs Obtained at the HF/6-31G and HF/MINI-1 (in parentheses) Optimized Geometries[a]**

structure[b]	ΔE^{HF} [c]	ΔE^{DFT} [d]	E^{DEF} [e]	ΔE^{COR} [f]	ΔE^{T} [g]
GCWC	−24.6 (−32.2)	−26.5	2.1 (6.6)	−1.2	−23.8 (−25.6)
GG1	−25.1 (−33.6)	−24.8	2.5 (10.5)	+0.4	−22.2 (−23.1)
CC	−16.1 (−21.8)	−18.4	1.4 (3.5)	−2.7	−17.5 (−18.3)
IC	−19.3	−19.2	1.4	−0.9	−18.0
GG3	−15.9 (−16.2)	−16.8	0.9 (1.7)	−1.8	−16.8 (−16.2)
GA1	−12.2 (−18.6)	−14.5	1.1 (3.6)	−3.1	−14.2 (−15.0)
GT1	−14.1 (−19.9)	−14.7	1.3 (5.3)	−0.9	−13.7 (−14.6)
GT2	−13.8 (−20.7)	−14.2	1.2 (5.9)	−0.9	−13.4 (−14.8)
AC1	−10.8 (−16.0)	−13.7	0.9 (2.7)	−3.5	−13.5 (−13.3)
GC1	−11.6 (−18.6)	−14.2	1.0 (3.2)	−2.7	−13.4 (−15.4)
AC2	−10.4 (−14.9)	−13.4	0.9 (2.2)	−3.7	−13.2 (−12.7)
GA3	−11.6 (−16.1)	−13.7	1.2 (2.8)	−3.3	−13.7 (−13.3)
TAH	−10.4 (−14.3)	−12.5	0.7 (2.1)	−2.9	−12.7 (−12.4)
TARH	−10.3 (−14.8)	−12.6	0.7 (2.2)	−2.9	−12.6 (−12.6)
TAWC	−9.7 (−15.9)	−11.9	0.7 (2.8)	−2.7	−11.8 (−13.1)
TARWC	−9.6 (−16.3)	−11.4	0.7 (3.0)	−2.7	−11.7 (−13.3)
AA1	−7.8 (−14.4)	−10.7	0.6 (2.1)	−3.7	−11.0 (−12.3)
GA4	−7.8 (−14.8)	−10.5	0.6 (2.5)	−3.4	−10.6 (−12.3)
TC2	−8.9 (−15.6)	−11.0	0.9 (3.6)	−2.7	−10.7 (−12.0)
TC1	−9.1 (−16.1)	−11.2	0.9 (3.9)	−2.7	−10.9 (−12.2)
AA2	−7.2 (−12.4)	−10.3	0.6 (1.9)	−3.8	−10.3 (−10.5)
TT2	−9.3 (−13.7)	−9.8	0.6 (3.2)	−1.3	−10.0 (−10.5)
TT1	−9.3 (−14.3)	−9.9	0.6 (3.5)	−1.3	−10.0 (−10.8)
TT3	−9.3 (−15.2)	−9.6	0.6 (4.0)	−1.3	−9.9 (−11.2)
GA2	−7.1 (−12.5)	−9.5	0.7 (2.4)	−3.5	−9.9 (−10.1)
GG4	−6.6 (−13.5)	−9.0	0.7 (3.2)	−3.5	−9.4 (−11.3)
AA3	−6.3 (−9.8)	−8.9	0.6 (1.6)	−3.7	−9.4 (−8.2)

[a] Taken from refs 79, 83, and 65. [b] Cf. Figure 2. [c] 6-31G*(0.25), (MINI-1). [d] B3LYP/6-31G**. [e] HF/6-31G**, (MINI-1); deformation energies of bases with respect to planar optimized monomers. [f] MP2/6-31G*(0.25) correlation interaction energy. [g] Total interaction energy, $\Delta E^{HF} + E^{DEF} + \Delta E^{COR}$, evaluated at MP2/6-31G*(0.25) level; E^{DEF} is determined at HF/6-31G** level; in the case of MINI-1, it is the sum of ΔE^{HF} and E^{DEF}.

deformation energies range from 0.6 to 2.5 kcal/mol and are largest for the three most stable pairs.

The same set of calculations was also carried out[66] at the HF level using the minimal basis set MINI-

1.[218] We already pointed out[219] in 1982 the good performance of this basis set for H-bonded and ionic complexes, and our experience in the following years[220,221] confirmed this fact. We do not present any data with inclusion of correlation energy for the MINI-1 basis set, since minimal basis sets cannot be used for a meaningful evaluation of electron correlation; this contribution would be strongly underestimated. The HF/MINI-1 results[66] are close to those determined with the 6-31G** basis set[79] and the stability order of the NA base pairs is mostly similar. The important difference between 6-31G** and MINI-1 results concerns the deformation energies, which are much smaller at the former level. In other words, the MINI-1 basis set exaggerates the intramolecular relaxation of monomers in the complex, which on the other hand increases the intermolecular interaction energy. Both contributions compensate each other, and this leads to rather reasonable total complexation energies. Nevertheless, the MINI-1 results are promising because the respective HF interaction energy can be determined for very large clusters.

Structures

Optimizations carried out at the HF/6-31G** level without geometrical constraints show that energy-optimized structures of many pairs are nonplanar.[83] One of the most prominent sources of nonplanarity of pairs is the pyramidalization of the amino groups of bases. Quite significant nonplanarity is apparent for all GA mismatch pairs; also TC pairs, several GG pairs, and 2-aminoadenine···thymine are nonplanar. On the other hand, the GCWC, all AT, AC, CC, and TT pairs are planar. Here the amino group is planarized as a result of the formation of H-bonds. The energy difference between nonplanar and planar base pairs is usually small. For example, for GA1, GA2, GA3, and GA4 pairs, it is 0.8, 0.5, 1.8, and less than 0.1 kcal/mol, respectively.[83] In the case of TC1, TC2, and AA3 pairs it is 0.2 kcal/mol, and for GG4 pair it amounts to 0.7 kcal/mol. The nonplanar base pairs are frequently propeller-twisted and buckled.[83,222] As will be shown later, the propeller and buckle motions correspond to the lowest intermolecular vibrations (below 40 cm^{-1}) in base pairs. Thus the base pairs are extremely flexible with respect to both out-of-plane deformations, and it explains why nonplanar pairs adopt propeller-twisted and buckled structures.[138] Performing the geometry optimization at the much cheaper HF/MINI-1 level, we obtain surprisingly very similar nonplanarities of base pairs as the HF/6-31G** ones; both basis sets agree as to which pairs are planar and which are not.[66] Intrinsic nonplanarity and large flexibility of base pairs can influence local DNA structure, recognition processes, and crystal packing; for more details, see ref 83.

Accuracy of the Theoretical Treatment Used

Structures of H-bonded NA base pairs were evaluated at the HF/6-31G** level using the standard gradient optimization. Hence, the results are influenced by three inaccuracies: (i) neglect of correlation energy; (ii) the use of only a medium-sized basis set; (iii) geometry optimization at the standard PES instead of using counterpoise-corrected PES. Discussion of these three inaccuracies requires a sufficient amount of reference data obtained at very high level of theory. However, there are almost no such data available at this moment due to the enormous computer requirements of such computations. Let us nevertheless discuss those higher level calculations, which have been recently reported.

Table 5. Geometrical Characteristics (in Å and deg) of H-Bonded Uracil Dimer and Cytosine Dimer Obtained at the MP2/6-31G and HF/G-31G** (in parentheses) Levels[a]**

dimer	X···Y[b]	H···Y[b]	X−H···Y[b]
UU4[c]	2.854 (2.992)	1.823 (1.920)	177.8 (177.3)
CC[d]	2.917 (3.050)	1.893 (2.049)	173.6 (173.1)

[a] From refs 79 and 223. [b] X = N_1(U) and N_4(C); Y = O_2(U) and N_3(C). [c] Cf. Figure 7. [d] Cf. Figure 2.

The neglect of electron correlation certainly influences the molecular structures of base pairs. Structures of two H-bonded NA base pairs, uracil dimer and cytosine dimer, were optimized at the MP2 level.[79,223] Table 5 summarizes some selected internal coordinates evaluated at HF and MP2 levels. Evidently, correlation energy brings both subsystems closer together. The distance of two heavy atoms in X−H···Y H-bond in both dimers is reduced by about 0.14 Å while the X−H···Y angle is not affected. Stabilization energies of both pairs determined at the MP2//HF and MP2//MP2 levels are similar. The effect of higher order correlation contributions (MP2 vs CCSD(T) data) and the effect of the size of the basis set will be discussed below in section IV.3.

Let us briefly comment on the role of BSSE in the structure and energetics of base pairs. Usually, the complex is optimized using the standard supermolecular gradient optimization and only at the very end, a posteriori, BSSE correction is added. This means that the basis set extension effect is eliminated only for the stabilization energy. For small complexes, evidence was found[224,225] that a proper consideration of basis set extension effect affects not only the interaction energy but also geometry and other properties of a complex. Simon et al.[110] recently offered a straightforward and elegant way to treat the problem. The geometry of a complex is evaluated at the CP-corrected surface, and during the respective optimization, the BSSE correction is applied in each gradient optimization cycle. Further, both intramolecular and intermolecular coordinates are simultaneously corrected for the basis set extension effects, which also means that the deformation energy is included within the stabilization energy. First applications of the method have shown[110−115] that stabilization energies from the CP-corrected PES are always larger (more attractive) than those from the standard PES (where the BSSE is added at the very end). Further, and this is more important, optimization at the standard PES might result in an inaccurate structure.

We have compared standard and CP-corrected HF/6-31G** optimizations for three NA base pairs: ATWC, TC2, and GA2. Standard HF/6-31G** gradi-

ent optimization led in the case of the first pair to a planar structure, while the other two pairs were quite significantly nonplanar.[83] Following experience from previous calculations, stabilization energies determined from CP-corrected optimizations were systematically larger (more attractive) than those from standard optimization. However, the difference for the NA base pairs investigated was very small and was equal to 0.04, 0.04, and 0.06 kcal/mol for ATWC, GA2, and TC2 pairs, respectively. Also the geometry changes introduced by the CP correction were rather modest. The CP-corrected optimization for ATWC pair again resulted in a planar structure with slightly larger intermolecular separations. The distance between the two nitrogens in the N−H···N intermolecular H-bond increased from 2.988 Å (standard optimization) to 3.043 Å, and distance between the nitrogen and oxygen in the second N−H···O intermolecular H-bond increased from 3.086 to 3.091 Å. Changes of the intramolecular bond lengths in adenine and thymine were marginal; the CP-corrected bond lengths were systematically smaller. Changes of intermolecular bond angles resulting from both optimizations were also small, mostly within 1°.

Investigating the nonplanar TC2 pair, we found only small changes in the intermolecular dihedral angles, mostly within 1°. Also, all intra- and intermolecular valence angles and intramolecular dihedral angles differed mostly by less than 1°. The intermolecular distance resulting from the CP-corrected optimization was larger with a similar extension of the H-bonds, as in the case of ATWC. Only in the case of the GA2 base pair, the geometry changes were more pronounced. The intermolecular dihedral angle N−H···N−C decreased from 134° (standard optimization) to 122° (CP-corrected optimization). Other geometrical features of both optimized structures were similar to those found in the case of other base pairs.

It can be concluded that the energetical and geometrical characteristics of ATWC, TC2, and GA2 pairs obtained by standard and CP-corrected optimizations differ only slightly; the largest difference was found for the GA2 pair. This means that standard geometry optimization as well as standard evaluation of stabilization energy (BSSE corrections added only for the final structure) is justified for H-bonded NA base pairs at the HF/6-31G** level. Considerably larger differences between CP-corrected and standard optimizations were obtained[110−115] for model systems at correlated level. Therefore, care should be paid in the future when applying gradient optimizations with inclusion of electron correlation effects to NA base pairs. Differences determined at standard and CP-corrected levels can be in these cases larger, since the BSSE is much larger when including electron correlation effects.

Vibrational Frequencies of Base Pairs: Harmonic Treatment

The HF/6-31G** and HF/MINI-1 harmonic intermolecular vibrational frequencies (no scaling factors were used) and the respective stabilization energies for 25 H-bonded NA base pairs are presented in Table 6.[138] The vibrational frequencies were evaluated for the optimized base pair structures. The MINI-1 frequencies are not presented for these pairs, where the optimal structure is nonplanar. An unambiguous assignment of individual modes is not easy, and we attempted to assign (using graphical visualization) at least the two lowest and the highest intermolecular modes. These modes correspond to the buckle, propeller, and stretching vibrations and are noted by the abbreviations b, p, and s, respectively. In ambiguous cases, no mode description is used, and sometimes, the combination of two modes is indicated due to a significant coupling between two vibrations. From Table 6 we can learn that the HF/6-31G** frequencies are practically always slightly lower than the corresponding HF/MINI-1 ones. Assignment of HF/6-31G** and HF/MINI-1 frequencies, at least in cases of b, p, and s vibrations, remains unchanged. Similarity of intermolecular vibrations, evaluated at both levels, is promising, because the HF/MINI-1 frequencies can be determined even for large complexes.

The following discussion will be based on more reliable HF/6-31G** data. From Table 6 it becomes evident that there is no correlation between stabilization energies and vibrational frequencies. This is true even for stretching frequencies, where this correlation was expected. The intermolecular vibrational frequencies are almost constant for all the pairs, while the respective stabilization energies differ considerably. For all the pairs the buckle and propeller vibrations are the lowest ones, while the stretching vibrational frequency is the highest. The lowest frequency for all the pairs lies in the narrow range 4−23 cm^{-1}, and stretching vibrations are between 83 and 169 cm^{-1}. Four pairs (AA2, GG1, GC1, GG3) possess the lowest vibration, below 10 cm^{-1}. These very low frequencies are of crucial importance for evaluation of thermodynamic characteristics, specifically of vibration entropy.

Vibrational Frequencies of Base Pairs: Nonharmonic Treatment

A natural question is whether a harmonic approach is adequate for vibrational analysis of base pairs. Let us first discuss one-dimensional anharmonic calculations[138] for the lowest buckle mode of the GCWC and GG1 pairs. From the one-dimensional HF/6-31G** potential energy curves for the buckle motion constructed for both pairs, it was immediately clear that the harmonic approach is justified in the case of the former pair, but it is completely unjustified in the latter pair. The anharmonic vibrational energy levels of the buckle motion for the GG1 base pair was obtained by solving a pertinent one-dimensional vibrational Schrödinger equation. The resulting fundamental frequency (13 cm^{-1}) is much higher than the respective harmonic frequency (6 cm^{-1}; cf. Table 6), giving clear evidence for inadequacy of the harmonic approach for the buckle motion of the GG1 pair. On the basis of this calculation, we believe that the low values of the buckle vibration in the case of AA2, GC1, and GG3 could be also explained by inadequacy of the harmonic approximation. Is it,

Table 6. Harmonic Intermolecular Vibrational Frequencies and Stabilization Energies for H-Bonded NA Base Pairs Evaluated at the Hartree−Fock Level[a]

pair	level[b]	ΔE (kcal/mol)	ν_i[c] (cm^{-1})					
GCWC	HF/6-31G**	25.5	25(b)	34(p)	69	84	116	123(s)
	HF/MINI-1	32.2	33(b)	41(p)	89	110	147	168(s)
GG1	HF/6-31G**	25.0	6(b)	31(p)	42	57	75	112(s)
	HF/MINI-1	33.6	21(b)	35(p)	81	82	96	134(s)
CC	HF/6-31G**	17.3	28(b)	38(p)	67	69	69	129(s)
	HF/MINI-1	21.8	31(bp)	39(bp)	79	83	91	168(s)
GG3	HF/6-31G**	16.8	9(b)	29(p)	43	48	61	99(s)
GA1	HF/6-31G**	12.6	17(b)	32(p)	46	67	83	94(s)
GT1	HF/6-31G**	14.0	20(b)	26(p)	53	71	72	105(s)
GT2	HF/6-31G**	13.7	18(b)	31(p)	52	62	68	117(s)
AC1	HF/6-31G**	11.9	22(b)	25(p)	57	57	75	108(s)
	HF/MINI-1	16.0	28	29	73	76	88	131
GC1	HF/6-31G**	12.7	4(b)	17(p)	49	50	68	113(s)
AC2	HF/6-31G**	11.4	22(b)	26(p)	58	66	77	108(s)
	HF/MINI-1	14.9	22(p)	28(b)	71	73	100	140
GA3	HF/6-31G**	11.0	18(b)	32(p)	49	66	84	95(s)
TAH	HF/6-31G**	10.9	23(b)	31(p)	53	67	88(s)	97
	HF/MINI-1	14.3	27	29	69	78	98	125(s)
TARH	HF/6-31G**	10.9	22(b)	33(p)	52	65	86	103(s)
	HF/MINI-1	14.8	23(b)	32(p)	67	71	99	138(s)
TAWC	HF/6-31G**	10.3	22(b)	30(p)	53	65	122	169(s)
	HF/MINI-1	15.9	27(b)	34(p)	68	80	101	130(s)
TARWC	HF/6-31G**	10.2	21(b)	32(p)	50	62	92	102(s)
	HF/MINI-1	16.3	24(b)	38(p)	66	69	99	148(s)
AA1	HF/6-31G**	8.7	14(pb)	18(pb)	46	50	74(s)	94
	HF/MINI-1	14.4	20	27	65	68	94	115
GA4	HF/6-31G**	8.8	10(p)	18(b)	39	49	72	97(s)
TC2	HF/6-31G**	9.2	19(p)	27(b)	50	70	77	111
AA2	HF/6-31G**	8.0	4(p)	17(b)	47	54	73	91(s)
	HF/MINI-1	12.4	16(p)	21(b)	64	69	94	116(s)
TT2	HF/6-31G**	9.0	17(pb)	27(pb)	56	71	72	99(s)
	HF/MINI-1	13.7	18(pb)	34(pb)	83	83	97	136
TT1	HF/6-31G**	9.1	15(b)	29(p)	55	64	70	107(s)
	HF/MINI-1	14.3	16(b)	37(p)	70	79	87	156(s)
TT3	HF/6-31G**	9.1	13(b)	33(p)	56	57	67	114(s)
	HF/MINI-1	15.2	14(b)	41(p)	62	79	88	177(s)
GA2	HF/6-31G**	7.5	18(b)	21(p)	35	51	78	91(s)
GG4	HF/6-31G**	7.3	19(p)	20(b)	38	55	57	99(s)
AA3	HF/6-31G**	6.9	17(p)	18(b)	40	63	70	83(s)

[a] Taken from Špirko et al. *J. Chem. Phys.* **1997**, *106*, 1472. [b] For nonplanar structures only HF/6-31G** values are presented. [c] b, buckle; p, propeller; s, stretch.

however, adequate to consider only one-dimensional anharmonicity and neglect couplings with other intermolecular and intramolecular modes?

To answer this question, we studied the ATWC pair using multidimensional nonharmonic treatment.[138] Relying on a Born−Oppenheimer-like separation of the fast and slow vibrational motions, the complete multidimensional vibrational problem was reduced to a six-dimensional (intermolecular) subproblem. The potential energy surface was determined at the HF/MINI-1 level, and an analytic potential energy function was obtained by fitting to *ab initio* data. Due to the symmetry of the optimized pair and by disregarding the role of the kinematic and potential interactions, the six-dimensional problem was reduced to two three-dimensional subproblems. The resulting anharmonic and also the respective harmonic frequencies are collected in Table 7. The agreement among harmonic and anharmonic in-plane and out-of-plane vibrational frequencies is good. In-plane harmonic frequencies are systematically overestimated by about 50%, while the absolute difference among harmonic and anharmonic out-of-plane vibrations is lower than 15%. The finding that anharmonic and harmonic out-of-plane frequencies, which are the lowest ones, differ by less than 15% is very promising for evaluation of vibrational frequencies and also of vibrational partition functions; harmonic frequencies are easily obtainable even for large clusters.

Table 7. Harmonic and Anharmonic Intermolecular Vibrational Frequencies of the Adenine···Thymine WC NA Base Pair Evaluated at the HF/MINI-1 Level[a]

level	ν_i[b] (cm^{-1})					
harmonic	27(b)	34(p)	68	80	101	130(s)
anharmonic	25(b)	41(p)	37	91	74	93(s)

[a] Taken from Špirko et al. *J. Chem. Phys.* **1997**, *106*, 1472. [b] b, buckle; p, propeller; s, stretch.

Literature Survey

As mentioned earlier, there have been many recently published *ab initio* quantum-chemical studies on H-bonded base pairs. Thus we would like to present a brief overview of selected studies.

Reference Values. The current reference study is the paper by Šponer and co-workers analyzing dozens

of H-bonded base pairs at the MP2/6-31G*(0.25)//HF/6-31** level under C_s symmetry.[79] These data were in detail summarized above. The paper also presents extensive density functional theory calculations and MP2 optimization on a CC base pair. Since many base pairs have been shown to have C_1 symmetry, a subsequent paper reported nonplanar HF/6-31G** geometries of base pairs and the energy differences between planar and nonplanar pairs.[83] The MP2/6-31G*(0.25)//HF/6-31** calculations have been recently verified by two groups. Brameld and coworkers optimized several base pairs at the HF level with an extended basis set [cc-pVTZ(-f)] of atomic orbitals, and the MP2 interaction energies were evaluated using local MP2 procedure with the same basis set.[86] These interaction energies agree within 1 kcal/mol with our data. This study is also noticeable since it presents a very accurate empirical force field for H-bonded base pairs giving an excellent agreement with our MP2/6-31G*(0.25)//HF/6-31** data for all base pairs. Šponer and Hobza reported reference calculations using a larger basis set (aug-cc-pVDZ).[240] Further, they have verified the MP2 procedure for H-bonded base pairs against CCSD(T) data.[240] The accuracy of the older paper by Gould and Kollman reporting MP2//HF energetics of four base pairs is influenced by the neglect of the BSSE correction at the MP2 level.[76]

Extended Systems and Modified Bases. Several recent papers reported MP2 level studies on modified or extended H-bonded systems of bases. Among them a study on protonated base pairs should be mentioned, showing the importance of molecular ion-molecular dipole and induction interactions.[82] Šponer and co-workers reported a set of high-level calculations on triplexes of DNA bases together with an estimate of many-body effects and a comparison with molecular mechanical force fields.[184] Florián and Leszczynski studied the noncanonical $\pi-k$ base pair[91] and Šponer and co-workers characterized the base-pairing properties of thioguanine and thiouracils.[89] The recent paper by Mayer and Sühnel reported the first *ab initio* characterization of hydrophobic (nonpolar) base pairs, namely adenine···difluorotoluene.[87] Cysewski characterized the base-pairing properties of 8-oxoguanine using the Becke3LYP DFT technique.[92]

Proton Transfer in Base Pairs. There are two recent very good high-level papers analyzing proton transfer processes in base pairs. Florián and Leszczynski studied mainly double-proton transfer in the GC base pair,[81] while Bertran et al. investigated radical−cation base pairs.[202] Both papers contain additional valuable information of general interest including a literature survey of previous related studies.

Hydration Effects. All above-mentioned studies have been done assuming the gas phase. It is therefore important to include solvent effects in future quantum-chemical studies on H-bonding of bases. Recently two different approaches have been used. Zhanpeisov and Leszczynski studied structures and energetics of several base pairs in a small water cluster representing the first hydration shell around the base pair.[93,226−228] Florián and co-workers studied a number of H-bonded base pairs by combining correlated *ab initio* calculations with Langevin dipole moment formalism and observed the expected destabilization of base pairing upon inclusion of polar solvent effects.[229]

Interactions of Base Pairs with Metal Cations. Recently several papers reported high-level *ab initio* studies on interactions between bare[184,185] or hydrated[186,187] cations and base pairs. These papers demonstrated that the stability of guanine-containing base pairs can be significantly enhanced upon a cation binding to N_7 of guanine due to polarization effects.[186,187] However, no such polarization base-pair enhancement has been noticed for adenine-containing base pairs.[187] The calculations also rationalized the biologically important difference between zinc and magnesium interacting with the nitrogenous sites of bases.[186] Zn^{2+} and Mg^{2+} cations have very similar ionic radii and both carry a charge of +2. However, the electronic structure of Zn^{2+} includes 3d electrons, and the 3d electrons participate in a significant covalent-type bonding with the nitrogenous sites of aromatic bases having lone pairs. At the same time both Zn^{2+} and Mg^{2+} cations have rather similar interaction with oxygen atom, including water. As a result, Zn^{2+} is bound more efficiently to the N_7 position of purines compared to Mg^{+} and the two cations have a different balance of water−cation and nucleobase−cation interactions. The picture revealed by *ab initio* calculations nicely correlates with the known propensity of zinc (and other similar divalent cations) to bind directly to bases in DNA while Mg^{2+} cations are almost exclusively localized near the phosphate units. We think that this is an important area for future quantum-chemical studies, since metal−cation complexes (characterized by large many-body effects and significant charge transfer) are difficult to treat by empirical potential methods and the metal cations are known to strongly influence the DNA molecule.

IV.2. Stacked NA Base Pairs

Base-stacking interactions are equally important to DNA structure as H-bonding of bases, and therefore, the same attention should be paid to them in theoretical studies. Nevertheless, the number of *ab initio* quantum-chemical studies devoted to base stacking is much smaller compared to studies on H-bonded base pairs. One of the reasons for the lack of studies on base stacking is that the potential energy surface of stacked pairs is more complex than that of H-bonded pairs. The PES of H-bonded pairs contains a small number of deep energy minima while the PES of a stacked dimer contains broad and shallow low-energy regions separated by low energy barriers. Further, as revealed by gradient geometry optimization, some stacked dimers appear to not have a minimum corresponding to a stacked arrangement, since there has been a spontaneous transition to a H-bonded structure.[223] In contrast to H-bonded base pairs, stacking geometries observed in nucleic acids do not coincide with optimal structures of isolated stacked dimers. Another reason for a lack of stacking papers is that stabilization of the stacked NA base

Table 8. Interaction Energies (ΔE) and Their Components, HF Interaction Energy (ΔE^{HF}), and MP2 Correlation Interaction Energy (ΔE^{COR}) of Stacking and H-Bonding NA Base Pairs Determined at the MP2/6-31G*(0.25) Level; (energies in kcal/mol)[a]

stacking[b,c]	ΔE^{HF}	ΔE^{COR}	ΔE	H-bonding[d,e]	ΔE^{HF}	ΔE^{COR}	ΔE
GG	−0.84	−10.47	−11.31	GG1	−25.08	+0.39	−24.69
GA	+1.30	−12.47	−11.16	GA1	−12.22	−3.01	−15.23
GU	−1.17	−9.45	−10.62	GT1	−14.23	−0.92	−15.15
CA	+0.85	−10.35	−9.50	CA1	−10.83	−3.51	−14.34
GC	−1.44	−7.87	−9.32	GC WC	−24.58	−1.23	−25.81
AU	+1.25	−10.33	−9.08	ATH	−10.38	−2.94	−13.32
AA	+4.01	−12.84	−8.83	AA1	−7.83	−3.72	−11.55
CC	−2.09	−6.17	−8.26	CC	−16.15	−2.66	−18.81
CU	−1.51	−7.01	−8.52	CT1	−8.68	−2.67	−11.44
UU	+0.46	−6.98	−6.52	TT2	−9.29	−1.35	−10.64

[a] Taken from Šponer et al. *J. Phys. Chem.* **1996**, *100*, 5590. [b] Cf. Figure 3. [c] MP2/6-31G*(0.25) single point optimization; optimal geometries were estimated using a combined *ab initio*/empirical potential search with rigid coplanar bases. [d] Cf. Figure 2. [e] MP2/6-31G*(0.25)//HF/6-31G**; C_s symmetry; deformation energy not included.

pairs is due to the dispersion energy, which originates in the electron correlation. Therefore, the structure and energetics of the stacked pairs cannot be investigated at the HF or DFT levels, and it is imperative to include electron correlation effects. To account properly for the dispersion energy, which dominantly stabilizes these pairs, rather large basis sets containing more sets of polarization functions are required.[230] Such calculations are, and in the near future will probably remain, prohibitively expensive. A compromise between accuracy and economy represents the use of the 6-31G*(0.25) basis set containing one set of diffuse polarization functions with an exponent of 0.25 on second row elements. This basis set provides very similar electrostatic interactions as the standard 6-31G* one. However, it improves base-stacking stabilization qualitatively (see section IV.3). In fact, an extended basis set would overestimate the stacking at the MP2 level (see below). We assume that for aromatic base stacking the MP2/6-31G*(0.25) data are very close to the actual values. The use of diffuse d-polarization functions[231] for stacking was recommended in ref 232.

Characterization of the Potential Energy Surface of Stacked Base Dimers

As explained above, studies of stacking interactions of NA bases are much more complicated than evaluation of H-bonding. It is thus not surprising that the first reliable study on base stacking appeared in 1995 and characterized the potential energy surface of a stacked cytosine dimer.[78] This study ultimately confirmed the dispersion stabilization of stacked base pairs.[78]

The preliminary study was followed a year later by extensive characterization of 10 stacked DNA base dimers.[84] This study demonstrated that the empirical $\pi-\pi$ "sandwich" model relying on the use of a set of symmetrically placed out-of-plane charges does not represent a proper description of aromatic base stacking. It also ruled out the possibility that stacking is dominated by interactions between delocalized electrons of aromatic rings and exocyclic groups of adjacent bases. Instead, this study for the first time convincingly demonstrated that a simple empirical force field, relying on a Lennard−Jones van der Waals potential and a Coulombic term employing just atom-centered point charges, is qualitatively sufficient to account for base-stacking phenomena. Our MP2-adjusted empirical potential reproduced the *ab initio* stacking energies for ca. 250 configurations of stacked gas-phase nucleobase dimers (considering adenine, uracil, cytosine, and guanine), with the accuracy of ±1.5 kcal/mol for the vertical separation between the nucleobases being 3.3 Å or more. Taking into account the simplicity of the empirical potential form, the agreement is excellent and it is not likely that a better agreement can be achieved without adding additional terms to the force fields. The most important step to get the above agreement is to properly parametrize the set of atom-centered point charges. This has been done in our case by fitting them to reproduce the molecular electrostatic field around the monomers at the MP2 level. It should be nevertheless noted that somewhat larger differences than those reported above were found for some nucleobase dimers with a vertical separation of bases below 3.3 Å. We have suggested that the lower accuracy of the potential for the "compressed" dimers is due to the anisotropy of short-range repulsion not included in the isotropic Lennard−Jones term.[80,89] The agreement between *ab initio* and empirical potential data for standard force fields currently used in the molecular modeling is of course not as good, since none of the force fields uses charges fitted to MP2 wave functions. There are considerable differences among the force field (see below).

The conclusion mentioned bolster the use of the empirical potential form currently used in molecular dynamics simulations and other empirical potential studies. An important point of both studies[78,84] was that stabilization energies of stacked pairs were determined at the same theoretical level as stabilization energies of H-bonded pairs,[79] which makes a direct comparison of both types of NA base pairs possible (cf. Table 8). Contrary to the study of H-bonded base pairs,[79] calculations of stacked pairs were made with the rigid planar bases: the MP2/6-31G*-optimized geometries of guanine, adenine, and uracil and MP2/DZ(2d)-optimized geometry of cytosine were used.[73] As explained above and in the literature,[84] gradient optimization is inconvenient for base stacking, and we did not use the gradient procedure also because the 6-31G*(0.25) basis set

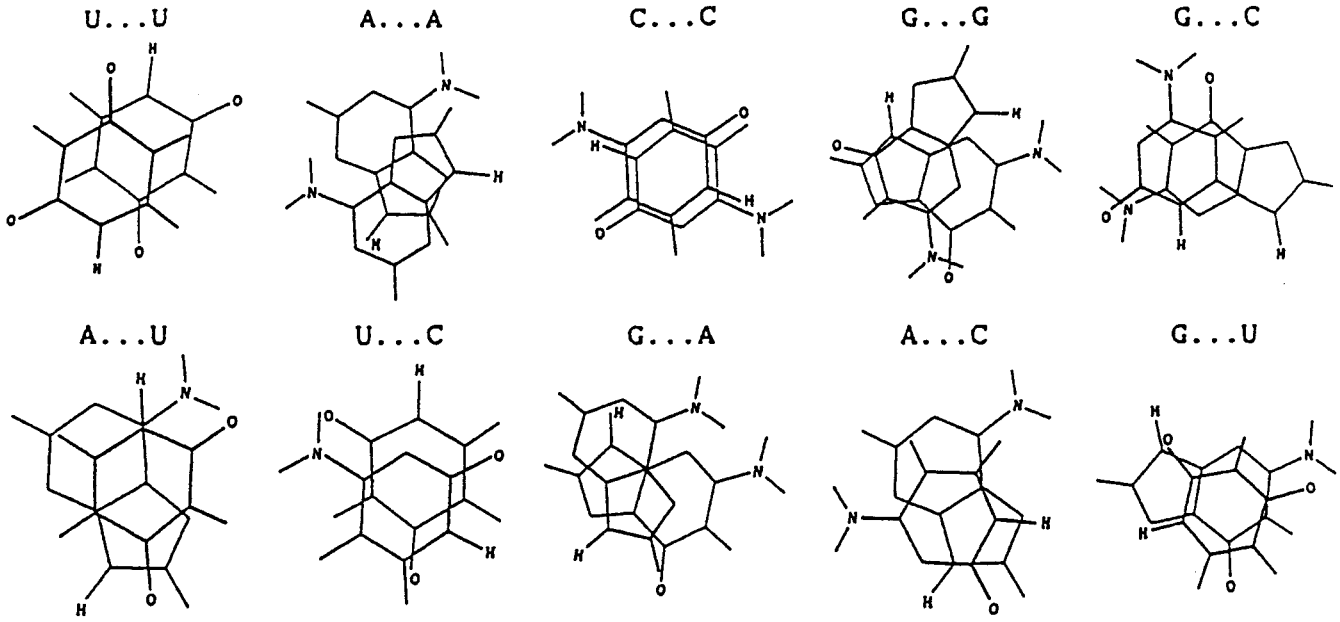


Figure 3. Structures of stacked NA base pairs.

could introduce artifacts into intramolecular degrees of freedom. Instead, we scanned the conformational space of stacked pairs by performing a set of single-point calculations. We were, of course, aware of the problems connected with such an optimization—the risk that the most important part of the PES could be completely missed.

Therefore, we have first evaluated ca. 200 points on the potential energy surface of various stacked dimers using an *ab initio* method. Then, we have parametrized an empirical potential which closely followed *ab initio* data. An essential point was the use of atom-centered point charges derived from the electrostatic field around monomers as provided by the MP2/6-31G*(0.25) wave functions, and this empirical potential is the most accurate pair-additive force field currently available for base stacking.[84,90] Then we finally made a three-dimensional search using this potential. The search included mutual displacement of the two bases and a parameter called twist; it means the mutual in-plane counterrotation of the two bases. The twist between the two bases is in our studies defined in the following way. The two bases are initially positioned in such a way that their centers of mass coincide and the glycosidic bond vectors are parallel. Then twist is a counterrotation of the two bases around the axis passing through the centers of mass of the nucleobases and perpendicular to their plane. Translational motions (vertical separation of bases and displacement) are performed after the twist rotation is done. Thus it resembles helical twist between consecutive base pairs in DNA, except that the helical twist is defined with respect to geometrical centers of base pairs. The vertical distance of bases was kept frozen at 3.3 or 3.4 Å, and both bases were assumed to be coplanar. Figure 3 shows optimal stacking geometries obtained for the 10 stacked dimers studied by the empirical potential. Table 8 summarizes MP2/6-31G*(0.25) stacking energies evaluated using these force field optimized geometries (the *ab initio* energies were quite close to the respective empirical potential values) and compare them with data for H-bonding. It is instructive to see the opposite ratio of HF and correlation components of interaction energies for these two types of clusters. It should also be noticed that in all cases the H-bonded structures are more stable in the gas phase;[84] however, as will be shown below, the mutual stability of stacked and H-bonded structures can be greatly influenced by considering entropy.

Finally, Figure 4 shows examples of the PES scan carried out for all four stacked homodimers (UU, AA, GG, CC) and illustrates the agreement between the MP2/6-31G*(0.25) method and the empirical potential.[84] The solid line shows the twist dependence of the empirical potential stacking energy for undisplaced dimers (undisplaced dimer means that the centers of monomers are stacked one above another).[84] The dashed line is the same dependence where the mutual displacement of bases is optimized. A twist of 0° corresponds to a parallel and a twist of 180° to an antiparallel arrangement of the bases in the dimer.[80] Crosses and open circles show the corresponding MP2/6-31G*(0.25) data for undisplaced and displaced structures, respectively. The curves nicely illustrate that while stability of stacking is due to the dispersion forces, the mutual orientation of bases is primarily determined by the electrostatic interaction.[84]

It should be noted that despite the overall good agreement between empirical potential and *ab initio* data, the empirical potential of course cannot fully substitute for accurate *ab initio* calculations. This has been demonstrated in the course of a subsequent study of stacked cytosine dimer.[80] The calculations revealed areas on the potential energy surface where increased anisotropic short-range repulsion decreases the agreement between the isotropic Lennard−Jones potential and *ab initio* data. Very similar effects were noticed in studies of the base stacking properties of modified thiobases;[89] here evidently the presence of the sulfur atom reduces the accuracy which can be achieved by applying a Lennard−Jones type of force field and *ab initio* calculations should be preferred.

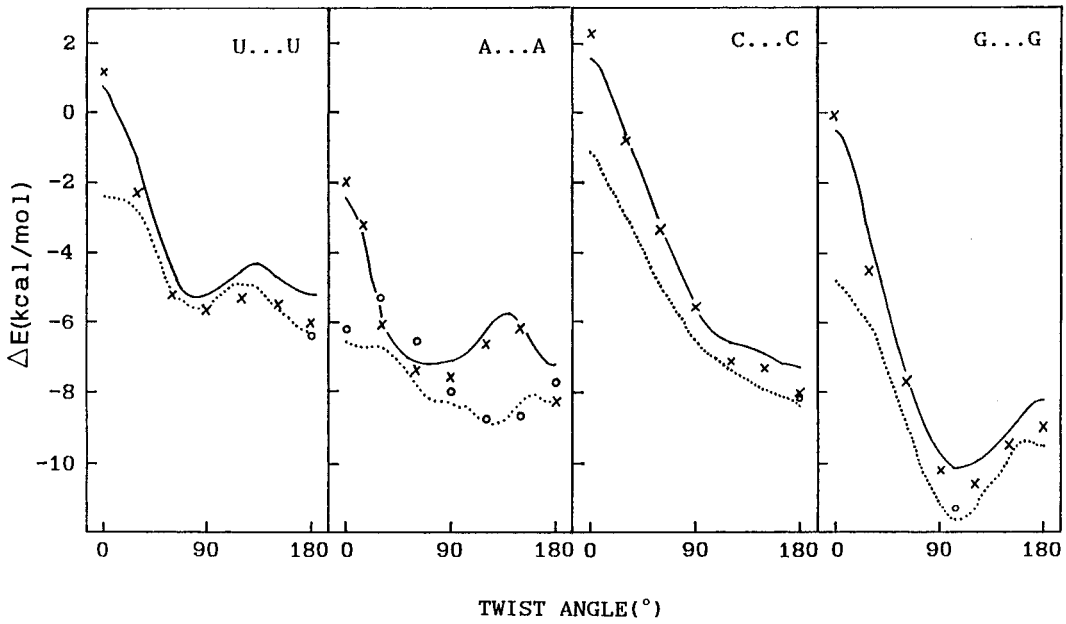


Figure 4. The dependence of the base stacking energy for CC, GG, UU, and AA stacked dimers on the twist angle calculated by the empirical potential for undisplaced dimers (solid line curves) and dimers with optimized interbase displacements (dashed line curves). The corresponding *ab initio* MP2/6-31G* (0.25) data are shown by crosses (undisplaced structures) and open circles (displaced structures). For more details, see the text and ref 84.

Another contribution which is not included in the potential is the classical induction contribution (permanent multipole−induced multipole). Induction energy is quite substantial in the case of protonated stacked complexes of nucleobases, and *ab initio* calculations show that the induction stabilization amounts to 2−4 kcal/mol.[82]

Another recent study considered base stacking of consecutive base pairs in B-DNA and Z-DNA utilizing standard and high-resolution oligonucleotide crystal geometries.[90] This study presents the first estimate of the many-body nonadditivity of base-stacking evaluated at the HF/6-31G*(0.25) level. Nonadditivity of stacking means that the base-stacking energy between two consecutive base pairs in DNA (base pair step) cannot be calculated as the sum of four base−base stacking contributions (two intrastrand and two interstrand). The stacking energy is also influenced by a many-body contribution. The many-body effects appear to be negligible for base pair steps with consecutive AT base pairs or alternating AT and GC base pairs. However, the nonadditivity of interactions influences base-stacking of base pair steps with consecutive GC base pairs and can modulate the base-stacking energy by as much as 10−20% of the total stacking energy. Although the nonadditivity effects seemed negligible in most studied configurations, they amounted to 2−2.5 kcal/mol of repulsion in GG base pair steps, which was ca. 20% of the total base-stacking energy.[90] This means that nonadditivity could significantly influence the sequence-dependence of stacking energy. Let us note that the nonadditivity can be further enhanced when considering electron correlation effects, so nonadditivity of stacking is clearly an issue which should be thoroughly scrutinized in future quantum chemical studies[90] (experimental data cannot be used for this purpose). The study indicates quite small variability of the stacking energy along the DNA helix, definitely much smaller than reported in all earlier studies.[90] This has been attributed to the fact that the calculations could be considered as physically complete, in contrast to the previous studies. However, significant variability has been noticed for intrastrand and interstrand stacking contributions which show a high degree of a mutual compensation. Their balance can contribute to the base-stacking controlled conformational variability. Similar calculations were reported for A- and B-DNA geometries by Alhambra et al.[88] However, they did not use diffuse polarization functions which leads to an underestimation of stabilization energies (see section IV.3). It explains why they reported weaker base pair step stacking in B-DNA compared to the other studies[84,90] despite having considered larger systems for the calculations with methylated N_1/N_9 positions of bases where the sugar rings are attached in DNA.

There have been recently two *ab initio* papers reporting calculations of base stacking in a polar solvent utilizing continuous models of a solvent. The first study reported the effect of a polar solvent on a twist dependence of stacking in a cytosine dimer, using the Onsager spherical cavity model.[233] The other study[229] considered all stacked base dimers and combined MP2/6-31G*(0.25) calculations with the Langevin-dipoles model.[234] Large competition between gas-phase and solvation energetics has been predicted, with an overall preference for complexes having twist angles around 30°. Of the other papers, studies estimating the influence of stacking phenomena on the ionization potential of bases in DNA should be mentioned.[203−206]

Gradient Optimizations of Stacked Base Pairs

To complement the picture obtained from studies using rigid monomers of bases, we performed[223] also in the case of stacked pairs the unconstrained MP2 gradient optimizations. The following pairs were

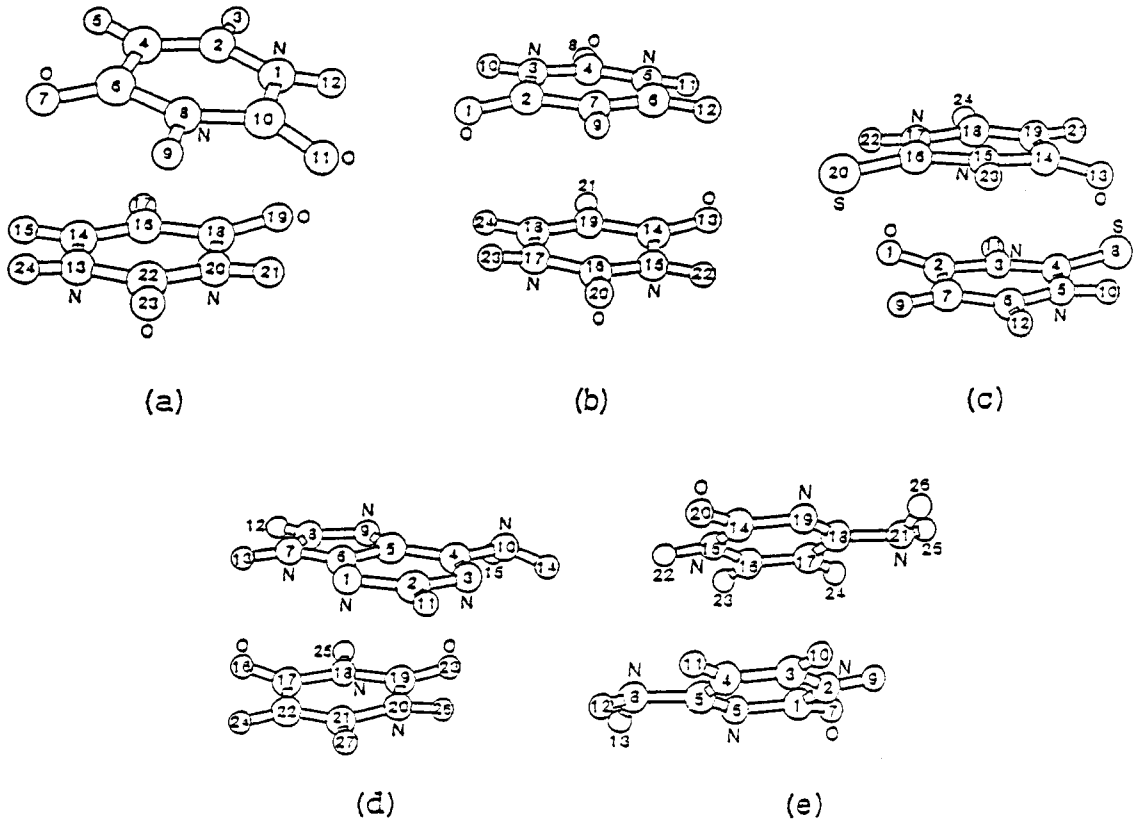


Figure 5. Structures of stacked NA base pairs determined by MP2 gradient optimization: uracil dimer [face to face (a), face to back (b)], thiouracil dimer (c), adenine-uracil pair (d), and cytosine dimer (e).

studied: uracil dimer (face-to-face and face-to-back structures), thiouracil dimer (face-to-back structure), cytosine dimer, and adenine···uracil (face-to-back structures). No stacking minimum has been found for the cytosine···uracil dimer due to a transition to a H-bonded pair. Selection of a basis set for such studies is not straightforward. As mentioned previously, the 6-31G*(0.25) basis set, used for stacked pairs (see above), could not be applied for gradient optimization. We finally decided to use the standard 6-31G* basis set. This basis set in combination with the MP2 procedure properly describes intrasystem and deformation energies but yields underestimated stabilization energies. On the other hand, the gradient procedure is not corrected for BSSE, which introduces another bias (artificial overstabilization) into the calculations. The calculations thus give reliable intrasystem geometries, but the intermolecular geometries may be slightly inaccurate. Therefore, the final stabilization energy was determined for these structures using the MP2 single point calculations with the 6-31G*(0.25) basis set standardly applying the full CP correction. In this way we corrected most of the inaccuracies, and the results are thus fully comparable with those for H-bonded base pairs.

The geometrical characteristics of the uracil dimers, thiouracil dimer, cytosine dimer, and adenine···uracil pair are summarized in Table 9 and shown in Figure 5. Further, the stabilization and deformation energies of these clusters are presented in Table 9. It must be mentioned that in this case the ZPVE is not included, because we were not able to evaluate the MP2 frequencies. At first sight, rings of the stacked

Table 9. Energetical (in kcal/mol) and Selected Geometrical Characteristics of the Stacked NA Base Pairs Determined at the MP2/6-31G* Level[a]

	ΔE	E^{DEF}	α (°)		R(Å)	
U···U[b]	−9.21	1.47	4−6−8−10	347	11−21	2.48
face-to-face			6−8−10−1	12	9−23	2.48
			6−8−10−11	192	19−12	2.95
U···U[b]	−9.07	1.63	5−4−3−2	346	11−13	2.90
face-to-back			4−3−2−7	15	1−23	2.90
			7−6−5−11	192	20−9	3.15
tU···tU[b]	−10.05	1.20	4−3−2−7	348	1−22	2.97
face-to-back			5−4−3−2	10	13−10	2.97
			4−3−2−1	167	20−9	3.40
C···C[b]	−10.41	1.43	6−1−2−3	355	9−21	2.98
			13−8−5−6	19	11−20	2.98
			7−1−2−3	174		
A···U[b]	−10.75	0.71	17−18−19−20	7	23−10	3.09
			3−4−10−14	20	23−15	2.69
			17−18−19−23	186	23−14	3.29

[a] Taken from Hobza, P.; and Šponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7. [b] See Figure 5.

dimers strongly deviate from planarity, especially for the uracil and thiouracil dimers. This is a very surprising result. Due to the fact that the stabilization of stacked pairs in the gas phase originates mainly from dispersion energy (electrostatic energy determines the orientation of subsystems in the pair), we expected that the optimal stacked structure should be characterized by planar subsystem geometries (with the exception of amino groups). From Table 9 it is evident that the nonplanarity of the subsystem rings in the case of both isomers of uracil dimer is significant (about 15°); for other stacked pairs, it is slightly smaller. The deformation of ring atoms in the cytosine stacked dimer is considerably

smaller and did not exceed 6°. The smallest ring nonplanarity (less than 5°) was found for adenine in the A···U stacked pair. Also the exocyclic atoms deviate from planarity and it concerns both oxygens in uracil and nitrogen in amino groups of adenine and cytosine. The nonplanarity of sulfur in thiouracil is lower. Worth mentioning is the fact that while the nonplanarity of the exocyclic groups in U···U, tU···tU, and A···U pairs leads to a formation of intermolecular H-bonds, the amino group hydrogens in the cytosine dimer are oriented away from the other cytosine. In this case, the lone pair region of one cytosine amino group is exposed to interact with the hydrogen of the N−H group in the adjacent cytosine, so this is yet another example of the amino acceptor interactions of bases.

What is the driving force leading to the significant geometry deformations of subsystems in the supersystem? The answer is unambiguous: the geometry deformation (associated with monomer energy increase) decreases the energy of the dimer. In the case of planar H-bonded systems with X−H···Y H-bonds, the elongation of the X−H bonds yields more favorable electrostatic dipole−dipole interactions between the proton donor and proton acceptor (e.g., see ref 235). In the case of stacked pairs, the ring deformation allows for more favorable contacts. Besides the dispersion energy increase, it also includes formation of intersystem H-bonds. For example, the intermolecular H-bonds in the stacked uracil dimer are by about 0.8 Å shorter than the separation of the same atoms in the stacked dimer with rigid subsystems. It is the absence of these H-bonds in the stacked cytosine dimer which is responsible for the smaller deformation of ring atoms in the dimer.

The large deformability of uracil and thiouracil in the stacked dimers is facilitated by the large deformability of the pyrimidine rings. It was shown[236,237] that the energy required to deform the pyrimidine ring in thymine is small, and we expect that it will be very similar for uracil. On the other hand, the ring deformation energy in adenine is more than three times larger. Consistently, the lowest ring deformation harmonic vibrational frequency of the isolated uracil is significantly lower than that of isolated adenine (134 and 295 cm^{-1}, respectively).

Comparing the present stabilization energies (stacked pairs with the deformed subsystems) with results obtained for stacked pairs with rigid planar subsystems (see above), we found that they do not differ dramatically.[223] Enhancement of the stabilization energies due to the formation of intermolecular H-bond is compensated by the repulsive intramolecular deformation energy. It can be therefore concluded that the stacking energy data obtained using rigid monomers[80,82,84,90] are quite realistic, and in addition they allow characterization of the whole conformational space. The nonplanarities revealed by the gradient optimization are nevertheless important in the gas-phase stacked dimers. Such nonplanarities might be important also in the DNA, if the overall distribution of various groups around the base results in a force supporting such a deformation.

Table 10. MP2, MP4, and CCSD(T) Interaction Energies (in kcal/mol) of Stacked Complexes Evaluated with Various Basis Sets

dimer	basis set	ΔE^{MP2}	ΔE^{MP4}	$\Delta E^{CCSD(T)}$
$(benzene)_2$[a]	6-31G	+0.39		
	6-31G*	−0.16		
	6-31G**	−0.43		
	6-31G*(0.25)	−1.76		
	cc-pVDZ/2s[b]	−1.77	−0.98	−0.46
	6-311G(2d,2p)	−2.05		
	cc-pVDZ(0.25,0.15)[b]	−2.21	−1.47	−0.91
	aug-cc-pVDZ/2s [b]	−2.56	−1.73	−1.12
	aug-cc-pVDZ	−2.80		
$(cytosine)_2$[c]	6-31G*	−5.36		
	6-31+G**	−7.64		
	6-31G*(0.25)	−8.23		
	6-311G(2d,2p)	−8.26		

[a] C_{6h} sandwich structure with vertical separation of 3.9 Å. [b] Taken from Hobza, Selzle, and Schlag *J. Phys. Chem.* **1996**, *100*, 18790. [c] C_i antiparallel undisplaced stacked structure with vertical separation of 3.3 Å.

IV.3. Reliability of MP2 Calculations

H-bonded NA base pairs were studied at the HF level. The stabilization energy was determined for the HF/optimized structures using the MP2/6-31G*(0.25) level of theory.[79] For a few H-bonded NA base pairs the full MP2 optimization was performed, and the correctness of the previous procedure was basically confirmed.[79,223] In the case of stacked NA base pairs, the calculations were performed at the MP2 level, first with rigid subsystems[80,82,84,90] and later using unconstrained MP2 gradient optimizations.[223] The question of how accurate these MP2 calculations are arises. The performance of the MP2 procedure for H-bonded NA base pairs will be discussed first. It was shown previously[238] that in the case of small H-bonded clusters, the MP2 and CCSD(T) stabilization energies (evaluated with the same basis set) are very similar, due to certain compensation of errors. To confirm this also for much larger H-bonded NA base pairs, the MP2 and CCSD(T) calculations were performed[239] for the H-bonded structures of the uracil dimer and the H_3-isocytosine dimer. Using the modified cc-pVDZ basis set ($\alpha_d = 0.25$; polarization functions on hydrogens were omitted), the following values of MP2 and CCSD(T) stabilization energies (in kcal/mol) were obtained: uracil dimer, 9.68 and 9.98; isocytosine dimer, 22.59 and 22.93.[240] Similarly to the nonaromatic H-bonded complexes, the MP2 and CCSD(T) procedures provide nearly identical results.[239,240] On the other hand, calculations on H-bonded base pairs and model systems show a very slow convergency of the correlation part of the interaction energy with the size of the basis set; higher angular momentum functions are important. However, this does not influence the final results qualitatively due to the dominating HF contribution.[239,240]

The requirements for the choice of the basis set for aromatic stacking calculations are more complicated and are demonstrated in Table 10 with stacking energies of a benzene dimer. The table compares MP2 interaction energies for a wide range of basis sets starting from the 6-31G (no polarization functions) up to the aug-cc-pVDZ one containing two sets of

d-functions, the second set being very diffuse. Further, data are shown for selected basis sets MP4SDTQ and CCSD(T). Let us first comment on the MP2 data. The MP2/aug-cc-pVDZ stacking energies can be considered as benchmark values, though a moderate increase of the stacking stabilization by further improvement of the flexibility and diffusivity of the basis set can still be expected. A basis set without polarization functions fails to describe the stacking attraction. The standard 6-31G* and 6-31G** basis sets are also very deficient. However, the modified 6-31G* (0.25) basis set, used in all our stacking calculations, gives stacking energies close to the MP2/aug-cc-pVDZ values. There is thus a *qualitative difference* in stacking energy calculations carried out with the 6-31G* and 6-31G* (0.25) basis sets. Let us once again reiterate that there is a clear explanation as the standard polarization functions in the 6-31G* and 6-31G** basis sets do not cover the space between the interacting monomers properly.

However, one should also consider the higher order correlation contributions, and here the CCSD(T) values represent the current benchmark. The CCSD(T) procedure consistently provides reduced stabilization compared with MP2 (by ca. 1.3 kcal/mol for all basis sets) for the present dimer. Let us to point out that the same behavior was found for all aromatic stacking clusters studied so far (see later). When considering and extrapolating these results the MP2/6-31G* (0.25) data appear to be very close to the actual (unknown) stacking energies, while the MP2/6-31G* stacking energies remain considerably underestimated. The MP2/aug-cc-pVDZ data are expected to overestimate the aromatic stacking.

Table 10 also shows that there is a difference between MP4SDTQ and CCSD(T) data, hinting at the importance of the higher order expansion of single, double, and quadruple electron excitations.

From Table 10 we can learn that the 6-31G*(0.25) basis provides accurate stabilization energies also for stacked NA base pairs. The structure studied is the antiparallel undisplaced cytosine dimer. The table again shows the quality of the diffuse 6-31G* (0.25) basis set for stacking. It gives better stacking values than the 6-31+G** basis set and is comparable with the large 6-311(2d,2p) basis set. Again, the standard 6-31G* basis set is far from the correct values of stacking and its use leads to an underestimation of the stacking energy.

CCSD(T) calculations on the stacked NA base pairs are much more complicated than these calculations for H-bonded pairs and are still impractical. Thus CCSD(T) calculations were performed for six model dimers containing aromatic heterocycles: $(\text{pyrrole})_2$, $(\text{pyrimidine})_2$, $(\text{triazine})_2$, $(\text{aminotriazine})_2$, $(\text{2-aminopyrimidine})_2$, and $(\text{4-aminopyrimidine})_2$.[240] All the dimers were studied in the optimal antiparallel geometry (rotation of subsystems by 180° around its normal; starting geometry is parallel undisplaced geometry) visualized in Figure 6. Calculations were made again with the modified cc-pVDZ basis set and the resulting stabilization energies are given in Table 11. From this table it is seen that the MP2 correlation interaction energies are overestimated (in absolute value) for all the dimers and all vertical distances with respect to the CCSD(T) data; this overestimation is similar to that found for $(\text{benzene})_2$, and C_2H_4···Ar and benzene···Ar clusters.[241−243]

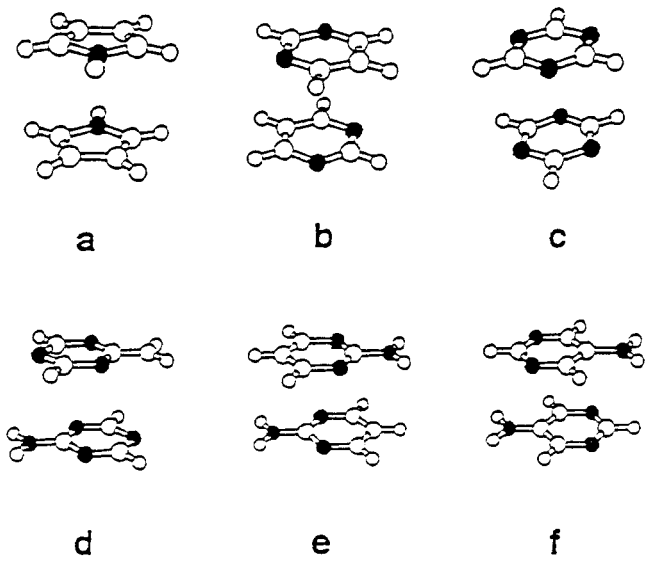


Figure 6. Stacked structures of six model dimers containing aromatic heterocycles: $(\text{pyrrole})_2$ (a), $(\text{pyrimidine})_2$ (b), $(\text{triazine})_2$ (c), $(\text{aminotriazine})_2$ (d), $(\text{4-aminopyrimidine})_2$ (e), $(\text{1-aminopyrimidine})_2$ (f).

Table 11. MP2 and CCSD(T) Interaction Energies (in kcal/mol) for Various Distances of Aromatic Stacked Complexes Evaluated with Modified cc-pVDZ Basis Set[a]

dimer[b]	orientation[c]	R (Å)	ΔE^{MP2}	$\Delta E^{\text{CCSD(T)}}$
$(\text{pyrrole})_2$	par	3.2	5.25	7.13
		3.6	0.97	2.03
	anti	3.2	2.88	4.73
		3.6	−0.63	0.45
$(\text{pyrimidine})_2$	anti	3.2	−1.99	0.00
		3.6	−3.87	−2.64
$(\text{triazine})_2$	anti	3.4	−3.77	−2.79
$(\text{4-aminopyrimidine})_2$	anti	3.4	−2.77	−1.27
$(\text{1-aminopyrimidine})_2$	par	3.2	1.75	4.50
		3.6	−1.57	−0.02
	anti	3.3	−7.04	−4.36
		3.6	−6.57	−5.02
$(\text{aminotriazine})_2$	anti	3.4	−3.87	−2.74

[a] Taken from Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263. [b] See Figure 6. [c] Par and anti mean parallel and antiparallel structures.

Analysis of the above-mentioned data leads us to expect that in the cases of H-bonded base pairs the MP2 stabilization energies evaluated with the medium basis set will be smaller than the actual values. This conclusion is based on the finding that the MP2 and CCSD(T) stabilization energies are almost identical for these clusters. Increasing the basis set size does not affect the HF interaction energy significantly, while the correlation interaction energy (in absolute value) is larger. The situation with the stacked pairs is different. Here we believe that the MP2 stabilization energies determined with medium diffuse basis sets having diffuse d-polarization functions are close to the actual values, owing to the compensation of errors. This conclusion stems from the fact that the MP2 stabilization energy calculated

with the medium (diffuse) basis set is close to the CCSD(T) one evaluated with the extended basis set.

IV.4. Application of the DFT Method

DFT methods with local and nonlocal approximations are known to work well for H-bonded systems in terms of structure, dipole moments, energetics, and vibrational frequencies.[244] Despite the fact that the DFT methods include correlation energy, it is not clear whether the intersystem correlation (dispersion) energy is properly taken into consideration. It was shown that both local and gradient-corrected DFT fail to describe the dispersion attraction in the case of noble gases.[245] Interestingly, the best performance was provided by the simplest method, the Dirac−Slater exchange. By investigating larger clusters such as the benzene dimer, it was shown[125] that dispersion energy is not included and the DFT stabilization energy differs considerably from the reference results obtained with inclusion of correlation energy. Besides London−van der Waals clusters (stabilized by dispersion attraction), the DFT also failed for another important class of molecular clusters, charge-transfer complexes, for which these methods predict unrealistically deep energy minima.[128]

Table 4 compares the DFT interaction energies for a set of H-bonded NA base pairs evaluated at HF/6-31G**-optimized geometries.[79] The DFT energies are in excellent agreement with *ab initio* MP2 values (cf. with ΔE^T in Table 4; for more details, see ref 79). In the case of standard DFT techniques, the best performance among various functionals was exhibited by the Becke3LYP one.[79] However, when applying a gradient optimization using the Becke3LYP/6-31G** method, we have evidenced overestimation of the deformation energy of bases while the H-bond lengths were too short.[79] The deviation between MP2 results and DFT results (single point calculations at HF/6-31G** geometries) is mostly below 1 kcal/mol. Interestingly, the same standard deviation was obtained for the same set of H-bonded NA base pairs with the SIESTA technique,[246] providing that the same geometry of H-bonded NA base pairs was used. Also in the case of the SCC-TB method, the interaction energies are close to the reference MP2 values.[247] It should be mentioned that results obtained with very economical SCC-TB and SIESTA DFT techniques are quite encouraging.

The situation with stacking complexes is much less optimistic. The DFT/Becke3LYP calculations[80] reproduced well the electrostatic part of the NA stacked pair interaction energy, but they do not include the dispersion energy. The dependence of stacking energy on twist agree well with the MP2 *ab initio* results, and this confirms that the electrostatic interaction is evaluated accurately. However, the stabilization energy is significantly underestimated,[80] because the dominating dispersion attraction seems to be neglected completely. Very similar results for 10 stacked NA base pairs were obtained with the SCC-TB method.[247] The stabilization energies are attractive, although significantly underestimated. The largest deviation (8.6 kcal/mol) was found for the GA stack, the smallest (2.8 kcal/mol) for the UU stack. It can be concluded that a reasonable inclusion of dispersion effects into DFT calculations is a key step which must be done in order to make these methods applicable for nucleic acids.

IV.5. Performance of Empirical Potentials

Several empirical potentials (force fields) are used for MD simulations of DNA and RNA in water environment. The quality of the results obtained depends on the performance of the simulation technique, as well as on the quality of the empirical potential used. The empirical potential should describe correctly all intramolecular as well as intermolecular degrees freedom. Among them the interactions of NA bases, both H-bonded and stacked play an important role. We decided to assess the quality of interaction energies of NA bases that were obtained by various empirical force fields.[98,99] The following methods were tested: AMBER, CHARMM, CVFF, CFF95, OPLS, and Poltev.[98,99] These potentials were used for calculation of stabilization energies of 26 H-bonded NA base pairs and 10 stacked NA base pairs. These energies were compared with nonempirical *ab initio* MP2/6-31G*(0.25)//HF-6-31G** values for planar base pairs taken from ref 79. The reason for testing these various potentials is obvious: simulation and modeling of DNA and RNA can be performed only by using some less demanding approach than the beyond-HF *ab initio* method. If, however, the potential of choice fails to give reliable stabilization energies for NA base pairs, then its use in DNA and RNA simulations is questionable. An advantage of the procedure described is that it is possible to test the quality of any other potential with benchmarks of beyond-HF *ab initio* data.

All the potentials tested were pairwise additive, include internal (bonding) and external (nonbonded) terms, and correspond to an all-atom model (i.e., all hydrogens are explicitly considered). None of the potentials used explicitly include a polarization term, and a constant dielectric equal to 1.0 was employed throughout this study. The following versions of empirical potentials were used: AMBER 4.1 with the force field of Cornell et al.;[139] CHARMM with empirical energy function parameter set 23;[248] CFF95 with the parameter set described in ref 249; OPLS with parameter sets of Pranata and colleagues;[250] and the Poltev potential described in ref 251. Table 12 contains stabilization energies for 26 studied H-bonded base pairs determined with five different empirical potentials and also with *ab initio* MP2 method. The results of linear regression ($Y = A + BX$) and absolute average errors are presented in Table 13. This table shows the largest correlation coefficient for AMBER 4.1 with the Cornell et al. force field,[139] followed by the OPLS and CFF95 potentials. The smallest standard deviation from the linear regression was again found for the AMBER 4.1 potential. Also, the smallest values of absolute average error were found for AMBER 4.1 and CHARMM potentials, and CFF95 still gave an acceptable value of 1.2 kcal/mol.

In the case of stacked NA base pairs, the twist dependence and vertical separation of bases were

Table 12. Stabilization Energies (in kcal/mol) of 26 H-bonded NA Base Pairs Obtained by *ab Initio* MP2 Method and Various Empirical Potentials[a]

base pair[b]	MP2[b]	A4.1[c]	CH23[d]	CFF95	CVFP[e]	OPLS	Polt[f]
GCWC	25.4	28.0	25.5	21.3	16.6	23.1	25.5
GG1	24.0	26.4	23.6	21.0	16.2	21.4	20.9
CC	18.8	18.7	18.1	14.2	10.5	15.1	17.5
GG3	17.1	19.4	24.3	16.8	13.6	17.7	18.6
GA1	15.7	14.7	15.1	14.9	9.6	12.3	13.4
GT1	14.7	16.1	14.0	13.8	10.0	12.9	11.0
GT2	14.3	15.6	13.6	12.5	11.6	12.3	13.4
AC1	14.3	13.5	12.9	13.0	8.1	9.9	13.7
GC1	13.9	15.3	15.1	12.2	9.4	12.3	16.8
AC2	14.1	13.7	11.4	14.7	11.7[g]	9.2	13.3
GA3	15.2	15.2	13.0	15.1	8.6	11.1	11.9
ATHO	13.3	14.5	13.3	12.6	7.9	10.8	11.9
ATRHO	13.2	14.5	13.1	12.1	8.3	10.6	11.7
ATWC	12.4	12.8	13.6	12.4	8.3	10.5	11.3
ATRWC	12.4	12.7	13.5	11.8	8.9	10.3	10.7
AA1	11.5	10.8	11.6	11.4	6.7	8.0	16.8
GA4	11.1	11.3	11.9	9.7	7.2	9.0	13.2
TC2	11.8	12.1	11.1	10.1	9.0	9.6	11.2
TC1	11.6	11.8	11.0	9.4	10.0	8.7	9.5
AA2	11.0	10.9	10.6	11.2	6.1	7.8	11.1
TT2	10.6	12.0	11.2	10.9	7.7	9.6	8.8
TT1	10.6	12.1	11.0	10.5	8.3	9.2	9.1
TT3	10.5	12.1	10.7	10.3	9.1	8.8	9.7
GA2	10.4	11.4	10.8	9.6	6.9	8.9	12.3
GG4	10.3	10.7	11.7	7.7	7.0	9.8	13.2
AA3	10.0	10.9	9.8	9.1	5.6	7.7	10.6

[a] Taken from Hobza et al. *J. Comput. Chem.* **1997**, *97*, 1136. [b] Cf. Figure 2. [c] AMBER 4.1. [d] CHARMM23. [e] CVFF force field with MEP/STO-3G atomic charges. [f] Poltev; uracil was considered instead of thymine. [g] Stacked structure.

Table 13. Linear Regression $Y = A + BX$ ($X \equiv \Delta E^{MP2}$) for Various Empirical Potentials[a]

method	A4.1[b]	CH23[c]	CVFF[d]	CFF95	OPLS	Polt[e]
R[f]	0.98	0.92	0.88	0.95	0.95	0.85
SD[g]	0.93	1.78	1.30	1.05	1.76	2.09
A	−0.59	0.14	0.79	1.64	−1.69	1.64
B	1.10	1.00	0.62	0.80	0.95	0.85
AAE[h]	0.9	1.0	4.4	1.2	2.4	1.7

[a] Taken from Hobza et al. *J. Comput. Chem.* **1997**, *97*, 1136. [b] AMBER 4.1. [c] CHARMM23. [d] CVFF with MEP/STO-3G atomic charges. [e] Poltev. [f] Correlation coefficient. [g] Standard deviation (in kcal/mol). [h] Average absolute error (in kcal/mol); AAE = $(1/26)\sum_1^{26}|\Delta E^{MP2} - \Delta E^X|$.

investigated. The standard potentials were compared with our MP2-fitted empirical potential, since it is very close to the original *ab initio* data.[84] In general, all potentials studied exaggerated the electrostatic interaction; i.e., the difference between maxima and minima on the twist curves was much larger than those found on the *ab initio* MP2 curve. The twist dependence is for some pairs strongly overestimated (compared to the MP2 values) by CHARMM, and all other potentials are closer to the MP2 potential. In general, the largest differences with respect to MP2 data were found for the CHARMM force field; the performances of AMBER 4.1, CFF95, and OPLS were mostly similar. A very similar conclusion was found also when investigating the energy dependence on vertical separation of bases in parallel and antiparallel arrangements; here, the accuracy is also influenced by the quality of the Coulombic term. It should be noted, however, that the van der Waals terms of all the force fields now appear to guarantee a proper value of vertical separation of the stacked systems and a proper value of absolute stacking energy. This is a marked improvement compared to some older parametrizations (cf. discussion in refs 22 and 54).

The AMBER 4.1 potential curves have a similar shape as the reference PES;, however, the maxima and minima of electrostatic energy are exaggerated by AMBER, especially for guanine and cytosine. This result can be easily explained. The atomic point charges of the AMBER 4.1 force field have been derived by a fitting to molecular electric potential around the monomers; thus, the PES should closely resemble the correct *ab initio* one. However, the fit has been made by assuming HF wave functions that exaggerate the dipole moments of the monomers, mainly of guanine and cytosine.

Among all tested methods, AMBER 4.1 with the force field of Cornell et al.[139] best reproduces the *ab initio* H-bonding and stacking stabilization energies. Further, all the studied force fields provided better descriptions of interactions of NA bases than the semiempirical quantum chemical methods (AM1, PM3, MNDO/M) and low-level (HF) *ab initio* treatment. Nevertheless, the HF *ab initio* method, even with a minimal basis set, yields rather reasonable geometries and stabilization energies of H-bonded pairs.

The recent force fields represent a considerable step forward in a proper description of base−base interactions. Further refinement of the force fields will require inclusion of the pyramidalization of amino groups of base to describe the out-of-plane H-bonds and amino−acceptor interactions and inclusion of a polarization term. A polarization term is important to describe, for example, protonated base pairs,[82] polarization of bases due to their interaction with metal cations,[183−187] and also nonadditivity of base-pair stacking along the double helix.[90]

IV.6. Thermodynamic Characteristics

Thermodynamic characteristics for the formation of H-bonded NA base pairs were evaluated using the *ab initio* data (geometries, harmonic vibrational frequencies, stabilization energies) utilizing the rigid rotor−harmonic oscillator−ideal gas approximation.[100] Theoretical interaction energies, interaction enthalpies, interaction entropies, and changes of the Gibbs energy for the formation of base pairs are collected in Table 14. The calculations were done for the most stable H-bonded base pairs but did not consider H-bonded structures where glycosidic nitrogens were involved in H-bonding. Further, no degeneracy factor has been applied to account for the fact that some dimers can have several isoenergetical structures. While the stabilization energies for various NA base pairs differ considerably (by more than 200%), the respective interaction entropy terms are more uniform and vary by less than 40%. Passing from ΔE to ΔH_T° (i.e. inclusion of ZPE and RT terms) slightly destabilizes pairs investigated. Following expectation, entropy is always important and partially or almost completely compensate the enthalpy stabilization. ΔG_T° is negative only for few the strongest NA base pairs. Calculated values are compared with experimental gas-phase interaction

Table 14. Thermodynamic Characteristics (in kcal/mol) for the Formation of NA Base Pairs Obtained by Rigid Rotor−Harmonic Oscillator−Ideal Gas Approximation with *ab Initio* Characteristics; Experimental Values from Field-Ionization Mass Spectrometry and Supersonic Beam Expansion are Given in Parentheses[a]

base pairs[b]	ΔE	ΔH_0°	T	ΔH_T°	$T\Delta S_T^\circ$	ΔG_T°	K_T
GCWC	−23.4	−21.3	298	−20.8	−12.2	−8.6	8.5×10^6 (17)[e]
			381	−20.5			
				(−21.0)[c]			
GG1[d]	−21.5	−20.7	298	−19.3	−9.2	−10.1	8.1×10^7 (8)[e]
CC	−17.5	−15.3	298	−14.9	−12.1	−2.8	1.1×10^2 (2)[e]
			381	−14.6			
				(−16.0)[c]			
GA1	−13.7	−11.9	298	−11.2	−11.3	0.1	1.6
GT1	−13.5	−12.9	298	−12.0	−10.9	−1.1	48
AC1	−13.5	−11.5	298	−11.0	−11.5	0.5	0.4
ATH	−12.7	−11.2	298	−10.4	−11.1	0.7	0.3 (1)[e]
			323	−10.3			
				(−13.0)[c]			
ATRH	−12.6	−11.1	298	−10.1	−11.1	1.0	0.185
ATWC	−11.8	−10.3	298	−9.5	−11.1	1.6	0.007
ATRWC	−11.7	−10.2	298	−9.4	−11.0	1.6	0.007
AA	−11.0	−9.2	298	−8.5	−10.9	2.4	0.002 (0.6)[e]
TC1	−10.7	−9.3	298	−8.5	−10.7	2.2	0.002
TT	−10.0	−9.1	298	−7.9	−9.9	2.0	0.003 (0.6)[e]
			323	−7.8			
				(−9.0)[c]			

[a] Taken from Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379. [b] See Figure 2. [c] See ref 34 (field-ionization mass spectrometry). [d] Due to the fact that the harmonic approach is not justified, the thermodynamic characteristics are not reliable and cannot be considered. [e] See ref 39 (supersonic beam expansion).

enthalpies[34] (field-ionization mass spectrometry results) and equilibrium constants[39] (supersonic beam experiments). The agreement in the former case was good and concerns not only the relative interaction enthalpies but even the absolute values; the average absolute error is less than 1.5 kcal/mol. On the other hand, experimental equilibrium constants for the NA base pair formation are far from the theoretical values and their validity was questioned. It is assumed that some experimental preconditions or assumptions given in ref 39 (reproducibility of the desorption process and/or reaching of the equilibrium in the beam expansion) were not fulfilled.

The comparison between the theoretical and experimental results described above is based on the assumption that only H-bonded structures of the NA base pairs exist in the gas phase. High-level *ab initio* calculations revealed that H-bonded structures are more stable than the stacked ones, but the question remains whether entropy is not changing the relation between H-bonded and stacked dimers. To answer this question, it is not enough to evaluate the entropy contribution for one particular structure of a dimer, but it is necessary to determine the entropy for all the structures of a pair. In other words, it is necessary to pass from the PES to the FES. The only way to fully describe the FES of NA base pairs is to evaluate the populations of various structures of the pair within the long-runs of MD simulations.

The first complete theoretical analysis of the gas-phase formation of a NA base pair has been performed recently, and the PES and FES of the uracil dimer were determined.[101] Uracil possesses a rather small dipole moment and it is thus expected that the H-bonded and stacked structures may be closer in energy. Furthermore, noncanonical uracil dimers occur in several RNAs.[252−254] Localization of the stationary point was performed by the MD/quenching technique, and 11 low-energy minima were located on the respective PES; among them seven are H-bonded, one is T-shaped, and three correspond to stacking arrangements (cf. Figure 7). Stabilization energies determined with the AMBER potential using the Cornell et al. force field are presented in Table 15. Evidently, H-bonded structures are more stable than the stacked ones, and the most stable

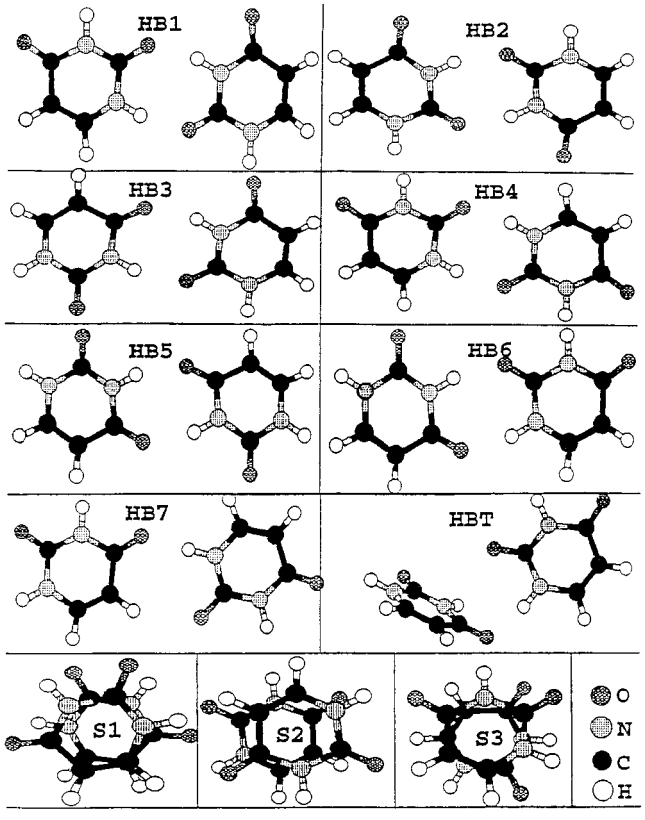


Figure 7. Structures of the uracil dimer; HB, T, and S mean H-bonded, T-shaped, and stacked structures.

Table 15. Interaction Energies and Changes of Gibbs Energies Determined by Rigid Rotor−Harmonic Oscillator−Ideal Gas Approximation (T = 298 K, p = 1 atm) Calculated by AMBER 4.1 and *ab Initio* Method for Various Structures of the Uracil Dimer (energies in kcal/mol[a])

		ΔE		$\Delta G°$	
structure[b]	σ[c]	AMBER 4.1	*ab initio*	AMBER 4.1	*ab initio*
HB1	2	−13.0	−12.7[d]	−1.3	−0.9
HB2	1	−11.1	−10.8[d]	0.9	1.5
HB3	2	−11.1	−10.4[d]	0.3	1.1
HB4	1	−15.9	−15.9[d]	−3.5	−3.0
HB5	1	−11.0	−10.5[d]	0.9	1.5
HB6	2	−13.4	−12.4[d]	−1.7	−1.0
HB7	2	−10.7	−10.5[d]	−0.7	0.2
T	4	−10.1	−7.8[d]	−0.5	1.5
S1	2	−9.4	−7.7[e]	1.5	
S2	2	−8.7	−7.4[e]	2.0	
S3	4	−8.1		1.7	

[a] Taken from Kratochvíl et al. *J. Phys. Chem. A* **1998**, *102*, 6921. [b] Cf. Figure 7; HB, T, and S mean H-bonded (planar), T-shaped, and stacked structure, respectively. [c] Symmetry number, i.e., the number of indistinguishable orientations of the particular structure. [d] MP2/6-31G*(0.25)//HF/6-31G**. [e] MP2/6-31G*(0.25)//MP2/6-31G*.

structure is a H-bonded dimer with two $N_1-H\cdots O_2$ H-bonds. To verify the empirical potential results, we performed the gradient optimization at the *ab initio* level for all the structures found by the MD/AMBER/quenching technique. From Table 15 it becomes evident that AMBER data agree excellently with the *ab initio* data. It must be mentioned that this is partly due to the fact that *ab initio* calculations were performed with the same basis set as used in the AMBER/RESP procedure. The best agreement between AMBER and *ab initio* was found for planar H-bonded structures where the largest absolute error is about 1 kcal/mol. In the case of stacked structures, the AMBER stabilization energies are overestimated by about 1.5 kcal/mol. This is, however, a remarkable success for the AMBER potential, since evaluation of geometries and stabilization energies requires the use of time-consuming correlated *ab initio* techniques. Relative populations of 11 minima found at the uracil dimer PES, determined by the MD simulations in the NVE microcanonical ensemble, are shown in Figure 8. The clearly dominant peak corresponds to H-bonded structure 4, which is also clearly dominant at the PES. Qualitatively, a new feature at the FES (with respect to the PES) is the population of stacked structures in general, and population of the stacked structure 3 in particular: The population of the latter structure is in fact the second highest. Evidently, entropy works for the energetically higher stacked structures compared to the H-bonded ones. Consequently, the order of stability of various dimer structures at the PES and FES differs, which indicates the need to compare the experimental data not with the results obtained from the PES (as it is mostly done) but from the FES. This is, of course, true only for experiments performed at temperatures distinctly different from 0 K. The present theoretical data, obtained from the MD/NVE simulations, are to be compared with the relevant experimental results, i.e., experiments studying the isolated dimer without its interaction with surroundings.

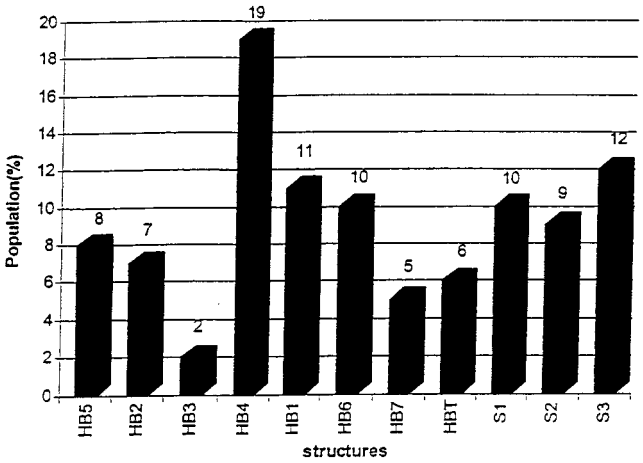


Figure 8. Relative population of various structures of the uracil dimer.

V. Conclusions

We have presented a review of the structure and properties of isolated NA bases and of structure, energetics, and properties of H-bonded and stacked NA base pairs as obtained from supermolecular nonempirical correlated *ab initio* calculations. The general findings of this review include the following:

(1) The amino group of NA bases is nonplanar and very flexible. The pyramidalization of the guanine amino group is significantly larger than that of the adenine and cytosine amino group. Direct experimental evidence about nonplanarity of isolated bases is missing. However, the reliability of the theoretical prediction is supported by calculating anharmonic vibrational spectra of an aniline molecule and the excellent agreement between its experimental and theoretical inversion frequencies. This gives confidence that the theoretical amino group nonplanarities of NA bases evaluated at the MP2 level with large basis sets are close to their actual values in the gas phase. Flexible amino groups of bases allow formation of out-of-plane H-bonds, mutual amino group contacts, and amino-acceptor interactions.

(2) Many H-bonded NA base pairs are nonplanar; nonplanar base pairs are significantly propeller twisted and buckled. The energy difference between nonplanar and planar pairs is, however, small (usually within 1 kcal/mol).

(3) Stabilization of H-bonded NA base pairs is of electrostatic origin and, therefore, rather reliable characteristics are obtained already at the Hartree−Fock level with medium-sized polarized basis sets of atomic orbitals. Nevertheless, inclusion of the electron correlation energy via the MP2 perturbational theory improves the results and further improvement can be achieved by applying large basis sets augmented by higher-angular momentum functions. On the other hand, the stability of stacked pairs is due to the dispersion energy, and thus, correlated calculations are strictly required for receiving reliable data on these pairs. Further, base-stacking calculations must be carried out using a basis set containing

diffuse d-polarization functions; failure to do so means a qualitative underestimation of the stabilization energies. However, it is not advised to apply extended basis sets of atomic orbitals for base stacking at the MP2 level, due to a moderate overestimation of base stacking by the MP2 method. The full counterpoise procedure must be used to eliminate the basis set superposition error in all interaction energy calculations of H-bonded and stacked base pairs, including the correlation part of the interaction energy. H-bonded pairs are systematically more stable than the corresponding stacked base pairs. An empirical potential consisting of a Lennard−Jones van der Waals term and a Coulombic term with atom-center point charges is sufficient to provide a reasonable description of base stacking. The charges must be fitted to reproduce the electrostatic potential around the monomers and the van der Waals term must be adjusted to reproduce the absolute values of stacking energies and optimal vertical separation of bases.

(4) Intermolecular vibrational frequencies of H-bonded pairs are almost constant, and for all pairs the buckle and propeller twist vibrations are the lowest ones. Harmonic and anharmonic intermolecular vibrational frequencies of the ATWC pair differ by less than 50%, and the harmonic approach is rather reliable. However, for base pairs with very low intermolecular vibrational modes, the harmonic approach is not valid. *Ab initio* evaluation of vibrational frequencies of stacked pairs is still impractical.

(5) Reliable structural and energetical results for H-bonded as well as stacked base pairs are only obtained at the correlated level, providing that the diffuse polarization functions are used. MP2 calculations with these basis sets yield underestimated H-bonded stabilization energies, while stacked stabilization energies will be close to the actual values.

(6) DFT methods fail to describe stacked base pairs, while they can provide reliable characteristics of H-bonded pairs.

(7) Use of current semiempirical methods of quantum chemistry for NA base pairs cannot be recommended.

(8) Entropy is always important in the formation of NA base pairs, and energetically less favorable stacked structures are preferred over the H-bonded ones. Reliable thermodynamic characteristics for H-bonded and stacked pairs are obtained only by using computer experiment simulations.

(9) Among various potentials tested, the AMBER potential with the Cornell et al. force field[139] best reproduces the *ab initio* H-bonding and stacking stabilization energies and is recommended for molecular dynamics simulations of DNA and RNA. Further improvement of the force field description requires inclusion of anisotropy of the interactions, amino group pyramidalization effects, and a polarization term. The polarization term would allow evaluation of many-body effects in stacked clusters, which modulate for example the sequence-dependence of base stacking energy along the double helix.

VI. Acknowledgments

We acknowledge with gratitude the contributions of our co-workers, as cited in the references. This work has been supported by grants from the Grant Agency of the Czech Republic (grants No. 203/97/0029 and 203/98/1166) and grants from Grant Agency of the Academy of Sciences of the Czech Republic (grants No. A4040903 and A4040904). Authors thank the referees for valuable and helpful comments that improved the quality of the review.

VII. References

(1) McCarty, M. *The Transforming Principle*; Norton: 1985.
(2) Avery, O. T.; MacLeod, C. M.; McCarty, M. *J. Exp. Med.* **1944**, *79*, 137.
(3) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
(4) Soyfer, V. N.; Potaman, V. N. *Triple-Helical Nucleic Acids*; Springer: New York, 1996.
(5) Gellert, G.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013.
(6) Zimmerman, S. B.; Cohen, G. H.; Davies, D. R. *J. Mol. Biol.* **1975**, *92*, 181.
(7) Tougard, P.; Chantot, J.-F.; Guschlbauer, W. *Biochim. Biophys. Acta* **1973**, *308*, 9.
(8) Wang, I.; Patel, D. J. *Biochemistry* **1992**, *31*, 8112.
(9) Laughlan, G.; Murchie, A. I. H.; Norman, D. G.; Moore, M. H.; Moody, P. C. E.; Lilley, D. M. J.; Luisi, B. *Science* **1994**, *265*, 520.
(10) Phillips, K.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171.
(11) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, H.-S.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751.
(12) Kennard, O.; Hunter, W. N. *Angew. Chem.; Int. Ed. Engl.* **1991**, *30*, 1254.
(13) Wahl, C.; Sundaralingam, M. *Biopolymers* **1997**, *44*, 45.
(14) Niedle, S.; Nunn, C. *Natur. Prod. Rep.* **1998**, 1.
(15) Calladine, C. R.; Drew, H. R. *J. Mol. Biol.* **1986**, *192*, 907.
(16) Šponer, J.; Kypr, J. *J. Mol. Biol.* **1991**, *221*, 761.
(17) Hunter, C. A. *J. Mol. Biol.* **1993**, *230*, 1025.
(18) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 27.
(19) Gorin, A. A.; Zhurkin, V. B.; Olson, V. K. *J. Mol. Biol.* **1995**, *247*, 34.
(20) Suzuki, M.; Yagi, N. *Nucl. Acid. Res.* **1995**, *23*, 2083.
(21) El Hassan, M. A.; Calladine, C. R. *J. Mol. Biol.* **1998**, *282*, 331.
(22) (a) Yanagi, K.; Prive, G. G.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 201. (b) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 277.
(23) Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341.
(24) Wang, A. H. J.; Quigley, G. J.; Kolpack, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G. A.; Rich, A. *Nature* **1979**, *282*, 680.
(25) Gessner, R. V.; Frederick; C. A.; Quigley, G. J.; Rich, A. H.; Wang, A. H.-J. *J. Biol. Chem.* **1989**, *264*, 7921.
(26) Smith, F. W.; Feigon, J. *Biochemistry* **1993**, *32*, 8682.
(27) Kettani, A.; Kumar, R. A.; Patel, D. J. *J. Mol. Biol.* **1995**, *254*, 638.
(28) Nonin S.; Leroy, J. L. *J. Mol. Biol.* **1996**, *261*, 399.
(29) Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998**, *75*, 968.
(30) Gehring, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561.
(31) Chen, L.; Cai, L.; Zhang, X.; Rich, A. *Biochemistry* **1994**, *33*, 13540.
(32) Berger, I.; Egli, M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12116.
(33) (a) Hagerman, P. J. *Annu. Rev. Biochem.* **1990**, *59*, 755. (b) Diekmann, S.; Mazzarelli, J. M.; McLaughlin, L. W.; von Kitzing, E.; Travers, A. A. *J. Mol. Biol.* **1992**, *225*, 729. (c) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature* **1986**, *320*, 501. (d) Bolshoy, A.; McNamara, P.; Harrington, R. E.; Trifonov, E. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2312. (e) Olson, W. K.; Zhurkin, V. B. In *Biological Structure and Dynamics, Proceedings of the Ninth Conversation*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Schenectady, *NY*, 1996; p 341, and references therein.
(34) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.
(35) Desfrancois, C.; Abdoul-Carime, H.; Carles, S.; Periquet, V.; Schermann, J. P.; Smith, D. M. A.; Adamovicz, L. *J. Chem. Phys.* **1999**, *110*, 11876.
(36) Desfrancois, C.; Abdoul-Carime, H.; Schermann, J. P. *J. Chem. Phys.* **1996**, *104*, 7792.

(37) Desfrancois, C.; Abdoul-Carime, H.; Schulz, C. P.; Schermann, J. P. *Science* **1995**, *269*, 1707.
(38) Schnier P. D.; Klassen, J. S.; Striimatter, E. F.; Williams, E. R. *J. Am. Chem. Soc.* **1998**, *120*, 9605.
(39) Dey, M.; Grotenmeyer, J.; Schlag, E. W. *Z. Naturforsch. A* **1994**, *49*, 776.
(40) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
(41) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001.
(42) Lee, H.; Darden, T.; Pedersen, L. *Chem. Phys. Lett.* **1995**, *243*, 229.
(43) Weerasinghe, S.; Smith, P. E.; Pettitt, B. M. *Biochemistry* **1995**, *34*, 16269.
(44) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.
(45) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134.
(46) (a) Špačková, N.; Berger, I.; Egli, M.; Šponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147. (b) Špačková, N.; Berger, I.; Šponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519.
(47) Soliva, R.; Laughton, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226.
(48) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227 and references therein.
(49) Shields, G. C.; Laughton, C. A. Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.
(50) Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805.
(51) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313.
(52) Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675.
(53) Cheatham, T. E., III; Brooks, B. R. *Theor. Chem. Acc.* **1998**, *99*, 279.
(54) Cieplak, P. In *The Encyclopedia of Comutational Chemistry*, Schleyer, P.v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; p 1922.
(55) Aida, M. *J. Theor. Biol.* **1988**, *130*, 327.
(56) Nagata, C.; Aida, M. *J. Mol. Struct. (THEOCHEM)* **1988**, *179*, 451.
(57) Aida, M.; Nagata, C. *Int. J. Quantum Chem.* **1986**, *29*, 1253.
(58) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
(59) Aida, M. *J. Comput. Chem.* **1988**, *9*, 362.
(60) Czerminski, R.; Kwiatkowski, J. S.; Person, W. B.; Szczepaniak, K. *J. Mol. Struct.* **1989**, *198*, 297.
(61) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. *Biopolymers* **1990**, *29*, 757.
(62) Colson, A. O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1991**, *96*, 661.
(63) Colson, A. O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1991**, *96*, 9767.
(64) Colson, A. O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1991**, *97*, 13852.
(65) Florián, J.; Hrouda, V.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 1457.
(66) Šponer, J.; Hobza, P. *Chem. Phys.* **1996**, *294*, 5590.
(67) Lewis, J. P.; Sankey, O. F. *Biohys. J.* **1994**, *69*, 1068.
(68) Šponer, J.; Hobza, P.; Leszczynski, J. In *Computational Chemistry. Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific Publisher: Singapore, 1996; p 185.
(69) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117.
(70) Šponer, J.; Hobza, P. In *The Encyclopedia of Computational Chemistry*, Schleyer, P. v R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; p 777.
(71) Leszczynski, J. *Int. J. Quantum Chem.; Quantum Biol. Symp.* **1992**, *19*, 43.
(72) Šponer, J.; Hobza, P. *J. Mol. Struct. (THEOCHEM)* **1994**, *304*, 35.
(73) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
(74) Bludský, O.; Šponer, J.; Leszczynski, J.; Špirko, V.; Hobza, P. *J. Chem. Phys.* **1996**, *105*, 11042.
(75) Šponer, J.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 706.
(76) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.
(77) Šponer, J.; Hobza, P.*J. Biomol. Struct. Dyn.* **1994**, *12*, 671.
(78) Hobza, P.; Šponer, J.; Polášek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792.
(79) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(80) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841.
(81) Florián, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *18*, 3010.
(82) Šponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695.
(83) Šponer, J.; Florián, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.
(84) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
(85) Gadre, S. R.; Pundlik, S. S. *J. Phys. Chem. B* **1997**, 101 3298.
(86) Brameld, K.; Dasgupta, S.; Goddard, W. A., III *J. Phys. Chem. B* **1997**, *101*, 4851.
(87) Meyer, M.; Sühnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619.
(88) Alhambra, C.; Luque, F. J.; Gago, F. Orozco, M. *J. Phys. Chem. B* **1997**, *101*, 3846.
(89) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489.
(90) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76.
(91) Florián, J.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1995**, *12*, 1055.
(92) Cysewski, P. *J. Chem. Soc.; Faraday Trans.* **1998**, *94*, 3117.
(93) Zhanpeisov, N.; Šponer, J.; Leszczynski, J. *J. Phys. Chem. A* **1998**, *102*, 2501.
(94) Gray, D. M. *Biopolymers* **1997**, *42*, 783 and references therein.
(95) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460.
(96) Sartorius, J.; Schneider, H.-J. *J. Chem. Soc. Perkin Trans.* 2 **1997**, 2319 and references therein.
(97) Newcomb, L. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 4993.
(98) Hobza, P.; Hubálek, F.; Kabeláč, M.; Mejzlík, P.; Šponer, J.; Vondrášek, *J. Chem. Phys. Lett.* **1996**, *257*, 31.
(99) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *97*, 1136.
(100) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379.
(101) Kratochvíl, M.; Engkvist, O.; Šponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921.
(102) Jansen, H. B.; Ross, P. *Chem. Phys. Lett.* **1969**, *3*, 140.
(103) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(104) van Duijneveldt, F. B. In *Molecular Interactions*; Scheiner, S., Ed.; Wiley: New York, 1997; p 81.
(105) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
(106) Urban, M.; Hobza, P. *Theor. Chim. Acta* **1975**, *36*, 215.
(107) Gutowski, M.; Chalasinski, G. *J. Chem. Phys.* **1993**, *98*, 5540.
(108) Cybulski, S. M.; Chalasinski, G. *Chem. Phys. Lett.* **1992**, *197*, 591.
(109) van Duineveldt, F. B.; van Duineveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.
(110) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.
(111) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372.
(112) Paizs, B.; Salvador, P.; Duran, M.; Suhai, S. J. *Comput. Chem.* Submitted.
(113) Famulari, A.; Raimondi, M.; Sironi, M.; Gianetti, E. *Chem. Phys.* **1998**, *232*, 275.
(114) Famulari, A.; Specchio, R.; Sironi, M.; Raimondi, M. *J. Chem. Phys.* **1998**, *108*, 3296.
(115) Hobza, P.; Bludský, O.; Suhai, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3073.
(116) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(117) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.
(118) Čížek, J. *Adv. Chem. Phys.* **1969**, *14*, 35. Sinanoglu, O. *J. Chem. Phys.* **1962**, *36*, 706.
(119) Nesbet, R. K. *Adv. Chem. Phys.* **1969**, *14*, 1.
(120) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359.
(121) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041.
(122) (a) Raghavachari, K.; Trucks, G. W.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1989**, *157*, 479. (b) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871.
(123) (a) Hohenberg, P.; Kohn, W. *Phys. Rev. B.* **1964**, *864*, 136. (b) Khon, W.; Sham, L. *J. Phys. Rev. A.* **1965**, *140*, 133.
(124) *Density Functional Methods in Chemistry*; Labanowski, J., Andzelm, J., Eds.; Springer: New York, 1991. Johnsson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612.
(125) Hobza, P.; Šponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.
(126) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
(127) Perez-Jordan, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.
(128) Ruiz, E.; Salahub, D. R.; Vela, A. *J. Am. Chem. Soc.* **1995**, *117*, 1141.
(129) Weselowski, T. A.; Parisel, O.; Ellinger, Y.; Weber, J. *J. Phys. Chem. A* **1997**, *101*, 7818.
(130) Weselowski, T. A.; Ellinger, Y.; Weber, J. *J. Chem. Phys.* **1998**, *108*, 6078.
(131) Carloni, P.; Andreoni, M. *J. Phys. Chem.* **1996**, *100*, 17797.
(132) Hutter, J.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **1996**, *118*, 8710.
(133) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haughk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(134) Sanchez-Portal, D.; Ordejon, P.; Artacho, E.; Soler, J. M. *Int. J. Quantum Chem.* **1997**, *65*, 453 and references therein.
(135) Gresh, N.; Garmer, D. R. *J. Comput. Chem.* **1996**, *17*, 1481. Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856. Åstrand, P. O.;

Wallqvist, A.; Karlström G. *J. Chem. Phys.* **1994**, *100*, 1262. Engkvist, O.; Åstrand, P. O.; Karlström, G. *J. Phys. Chem.* **1996**, *100*, 6950.
(136) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208. Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177.
(137) Wilson, E. B.; Decius, J. D.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
(138) Špirko, V.; Šponer, J.; Hobza, P. *J. Chem. Phys.* **1997**, *106*, 1472.
(139) Cornell, W.D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. E.; Kollamn, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(140) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199.
(141) Colominas, C.; Luque, F. J.; Orozco, M .*J. Am. Chem. Soc.* **1996**, *118*, 6811.
(142) Schuerman, G. S.; Van Meervelt, L.; Loakes, D.; Brown, D. M.; Lin, P. K. T.; Moore, M. H.; Salisbury, S. *J. Mol. Biol.* **1998**, *282*, 1005.
(143) Lippert, B.; Schollhorn, H.; Thewalt, U. *J. Am. Chem. Soc.* **1986**, *108*, 6616.
(144) Pichierri, F.; Holtheinrich, D.; Zangrando, E.; Lippert, B.; Randaccio, L. *J. Biol. Inorg. Chem.* **1996**, *1*, 319.
(145) Lippert, B. *J. Chem. Soc. Dalton Tr.* **1997**, 3971.
(146) (a) Metzger, S.; Lippert, B. *Angew. Chem.; Int. Ed. Engl.* **1996**, *35*, 1228. (b) Zamora, F.; Kunsman, M.; Sabat, M.; Lippert, B. *Inorg. Chem.* **1997**, *36*, 1583.
(147) (a) Cunąne L. M.; Taylor, M. R. *Acta Crystallogr. D* **1997**, *53*, 765. (b) Šponer, J.; Gorb, L.; Leszczynski, J.; Sabat, M.; Lippert, B.; Hobza, P. Manuscript in preparation.
(148) Estrin, D. A.; Paglieri, L.; Corongiu, G. *J. Phys. Chem.* **1994**, *98*, 5653.
(149) Kwiatkowski, J. S.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 941.
(150) Nowak, M. J.; Lapinski, L.; Kwiatkowski, J. S.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 3527.
(151) Ha, T. K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. *J. Mol. Struct.* **1996**, *376*, 375.
(152) Ha, T. K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1996**, *363*, 161.
(153) Broo, A.; Holmen, A. *Chem. Phys.* **1996**, *211*, 147.
(154) Fogarasi, G.*J. Mol. Struct.* **1997**, *413*, 271.
(155) Gould, I. R.; Burton, N. A.; Hall, H. J.; Hillier, I. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 147.
(156) Cysewski, P. *J. Chem. Soc.; Faraday Trans.* **1998**, *94*, 3117.
(157) Morpurgo, S.; Bossa, M.; Morpurgo, G. O. *Chem. Phys. Lett.* **1997**, *280*, 233.
(158) Boughton, J. W.; Pulay, P. *Int. J. Quantum. Chem.* **1993**, *47*, 49.
(159) Leszczynski, J. *J. Phys. Chem.* **1992**, *96*, 1649.
(160) Leszczynski, J. *J. Phys. Chem. A* **1998**, *102*, 2357.
(161) Kobayashi, R. *J. Phys. Chem. A* **1998**, *102*, 10813.
(162) Leszczynski, J. In *The Encyclopedia of Computational Chemisty*, Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; p 2951.
(163) Luque, F. J.; Lopez-Bes, J.-M.; Cemeli, J.; Aroztequi, M.; M. Orozco, M. *Theor. Chem. Acc.* **1997**, *96*, 105 and references therein.
(164) Gorb, L.; Leszczynski, J. *J. Am. Chem. Soc.* **1998**, *120*, 5024.
(165) Alhambra, C.; Luque, F. J..; Esterlich, J.; Orozco, M. *J. Org. Chem.* **1995**, *60*, 966.
(166) Orozco, M.; Hernandez, B.; Luque, F. J. *J. Phys. Chem. B* **1998**, *102*, 5228.
(167) Goodman, N. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493.
(168) Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L. *Acta Crystallogr.* **1962**, *15*, 310.
(169) Landridge, L.; Rich, A. *Nature* **1963**, *198*, 725.
(170) Inman, R. B. *J. Mol. Biol.* **1964**, *9*, 624.
(171) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059.
(172) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637.
(173) Doronina, S. O.; Behr, J.-P. *Chem. Soc. Rev.* **1997**, *26*, 63.
(174) Del Benne J. E. *J. Phys. Chem.* **1983**, *87*, 367.
(175) Florián, J.; Baumruk, V.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 5578.
(176) (a) Lavalle, L.; Fresco, J. R. *Nucleic Acid. Res.* **1995**, *23*, 2692. (b) Gallego, J.; Golden, E. B.; Stanley, D. E.; Reid, B. R. *J. Mol. Biol.* **1999**, *285*, 1039.
(177) Wang, C.; Gao, H.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **1991**, *113*, 5486.
(178) Gago, F. *Methods: Companion Methods. Enzymol.* **1998**, *14*, 277.
(179) Gadre; S. R.; Pundlik; S. S. *J. Phys. Chem. B, 1997*, 101, 3298.
(180) Soliva, R.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1997**, *18*, 980.
(181) Basch, H.; Krauss, M.; Stevens, W. J. *J. Am. Chem. Soc.* **1985**, *107*, 7267.
(182) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. *Biopolymers* **1990**, *29*, 757.
(183) Burda, J. V.; Šponer, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 7250.
(184) Šponer, J.; Burda, J. V.; Mejzlík, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613.
(185) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670.
(186) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1998**, *102*, 5951.
(187) Šponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1998**, *16*, 139.
(188) Gresh, N.; Stevens, W. J.; Krauss, M. *J. Comput. Chem.* **1995**, *16*, 843.
(189) Garmer, D. R.; Gresh, N. *J. Am. Chem. Soc.* **1994**, *116*, 3556.
(190) Basch, H.; Garmer, D. R.; Jasien, P. G.; Krauss, M.; Stevens, W. J. *Chem. Phys. Lett.* **1989**, *163*, 514.
(191) Jasien, P. G.; Fitzgerald, G. *J. Chem. Phys.* **1990**, *93*, 2554.
(192) Bakalarski, G.; Grochowski, P.; Kwiatkowski, J. S.; Lesyng, B.; Leszczynski, *J. Chem. Phys.* **1996**, *204*, 301.
(193) Johnson, R. C.; Power, T. D.; Holt, J. D.; Immaraporn, B.; Monat, J. E.; Sissoko, A. A.; Yanik, M. M.; Zagorodny, A. V.; Cybulski, S. M. *J. Phys. Chem.* **1996**, *100*, 18875.
(194) Shatzky-Schwartz, M.; Arbuckle, N.; Eisenstein, N. D.; Rabinovich, D.; Bareket-Samish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z. *J. Mol. Biol.* **1997**, *267*, 595 and references therein.
(195) Kasimi, N. E.-B.; Thakkar, A. T. *J. Mol. Struct. (THEOCHEM)* **1996**, *366*, 185.
(196) Hobza, P.; Šponer, J.; Leszczynski, J. *J. Phys. Chem. B* **1997**, *101*, 8038.
(197) Sevilla, M. D.; Besler, B.; Colson, A. O. *J. Phys. Chem.* **1995**, *99*, 106.
(198) Oyler, N.; Adamowicz, L. *J. Phys. Chem.* **1993**, *97*, 11122.
(199) Colson, A.-O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1992**, *96*, 9787.
(200) Colson A.-O.; Sevilla, M. D. *Int. J. Radiat. Biol.* **1995**, *67*, 627.
(201) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, *118*, 7574.
(202) Bertran, J.; Oliva, A.; Rodriquez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159.
(203) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.
(204) Prat, F.; Houk, K. N.; Foote, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 845.
(205) Nakatani, K. Okamoto, A.; Matsuno, T.; Saito, I. *J. Am. Chem. Soc.* **1998**, *120*, 11219.
(206) Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686.
(207) Gould, I. R.; Hillier, I. H. *Chem. Phys. Lett.* **1989**, *161*, 185.
(208) Riggs, N. V. *Chem. Phys. Lett.* **1991**, *177*, 447.
(209) Šponer, J.; Burcl, R.; Hobza, P. *J. Biomol. Struct. Dyn.* **1994**, *11*, 1357.
(210) Šponer, J.; Kypr, J. *Int. J. Biol. Macromol.* **1994**, *16*, 3.
(211) Luisi, B.; Orozco, M.; Šponer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123.
(212) Kung, H. C.; Wang, K. Y.; Parker, S. A.; Goljer, *I*, Bolton, P. H. *Magnet. Res. Chem.* **1996**, *34*, S47.
(213) Aamouche, A.; Ghomi, M.; Grajcar, L.; Baron, M. H.; Romain, F.; Baumruk, V.; Štěpánek, J.; Coulombeau, C.; Jobic, H.; Berthier, G. *J. Phys. Chem. A* **1997**, *101*, 10063.
(214) Larsen, N. W.; Hanses, E. L.; Nicolaisen, F. M. *Chem. Phys. Lett.* **1976**, *43*, 584.
(215) Kydd, R. A.; Krueger, P. J. *Chem. Phys. Lett.* **1977**, *49*, 539.
(216) Kydd, R. A.; Krueger, P. J. *J. Chem. Phys.* **1978**, *69*, 827.
(217) McCarthy, W. J.; Lapinski, L.; Nowak, M. J.; Adamowicz, L. *J. Chem. Phys.* **1995**, *103*, 656.
(218) Tatewaki, H.; Huzinaga, S. *J. Chem. Phys.* **1979**, *71*, 4339.
(219) Hobza, P.; Sauer, J. *Theor. Chim. Acta* **1982**, *61*, 41.
(220) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871.
(221) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767.
(222) Cambridge Convention *J. Biomol. Struct. Dyn.* **1989**, *4*, 627.
(223) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.
(224) Bouteiller, Y.; Behrouz, H. *J. Phys. Chem.* **1992**, *96*, 6033.
(225) Leclercq, J. M.; Allavena, M.; Bouteiller, Y. *J. Chem. Phys.* **1983**, *78*, 4606.
(226) Zhanpeisov, N.; Leszczynski, J. *Int. J. Quantum. Chem.* **1998**, *69*, 37.
(227) Zhanpeisov, N.; Leszczynski, J. *J. Phys. Chem. B* **1998**, *102*, 9109.
(228) Zhanpeisov, N.; Leszczynski, J. *J. Phys. Chem.* In press.
(229) Florián, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884.
(230) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500.
(231) Kroon-Batenburg, L. M. J.; van Duijneveldt, F. B. *J. Mol. Struct.* **1985**, *121*, 185.
(232) Hobza, P.; Mehlhorn, A.; Čársky, P.; Zahradník, R.; Hobza, P. *J. Mol. Struct. (THEOCHEM)* **1986**, *138*, 387.
(233) Subramanian, V.; Sivanesan, D.; Ramasami, T. **1998**, *290*, 189.
(234) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640. Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(235) Hobza, P.; Havlas, Z. *Chem. Phys. Lett.* **1999**, *303*, 447.
(236) Shiskhin, O. V. J. *Chem. Soc.; Chem. Commun.* **1995**, 1539.

(237) Shiskhin, O. V.; Šponer, J.; Hobza, P. *J. Mol. Struct.* **1999**, *477*, 15.
(238) Harrison, R. J.; Bartlett, R. J. *Int. J. Quantum Chem.: QCS* **1986**, *20*, 437.
(239) Hobza, P.; Šponer, J. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 115.
(240) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263.
(241) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.
(242) Yang, M.; Alexander, M. H.; Werner, H.-J.; Bemish, R. J. *J. Chem. Phys.* **1996**, *105*, 10462.
(243) Werner, H.-J. Unpublished data, 1994.
(244) (a) Latajka, Z.; Bouteiller, Y.*J. Chem. Phys.* **1994**, *101*, 9793. (b) Kieninger, M.; Suhai, S. *Int. J. Quantum Chem.* **1994**, *52*, 465.
(245) Dudzianowski, A. T. *J. Phys. Chem.* **1996**, *100*, 4781.
(246) Machado, M.; Ordejon, P.; Artacho, E.; Sanchez-Portal, D.; Soler, J. M. *J. Phys. Chem.* Submitted.
(247) Elstner, M.; Porezag, D.; Seifert, G.; Frauenheim, Th. *Th. Mater. Res. Soc. Symp. Proc.* **1999**, *538*, 541.
(248) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187. (b) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
(249) Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *115*, 162. Hagler, A. T.; Ewig, C. S. *Comput. Phys. Commun.* **1994**, *84*, 131.
(250) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810.
(251) Zhurkin, V. B.; Poltev, V. I.; Florentov, V. L. *Mol. Biol. (USSR)* **1980**, *14*, 1116.
(252) Cech, T. R.; Damberger, S. H.; Gutell, R. R. *Nat. Struct. Biol.* **1994**, *1*, 273.
(253) Grüne, M.; Fürste, J. P.; Klussmann, S.; Erdmann, V. A.; Brown, L. R. *Nucleic Acid Res.* **1996**, *24*, 2592.
(254) Lietzke, S. E.; Barnes, C. L.; Berglund, J. A.; Kundrot, C. E. *Structure* **1996**, *4*, 917.

CR9800255