

Jiří Šponer, Manoj K. Shukla, Jing Wang, and Jerzy Leszczynski

Contents

Introduction	1804
Hydrogen Bonding and Stacking Interactions in Nucleic Acids	1809
Level of Computations	1811
Gold Standard	1812
Other Approaches	1812
Geometries	1814
Interpreting the Computations	1817
Conclusion	1820
Bibliography	1820

Abstract

A comprehensive analysis of the benefits and pitfalls of quantum chemical methods used to determine the structures, properties, and functions of DNA and RNA fragments is presented. Main emphasis is given to the application of different ab initio quantum chemical methods. An overview of computations reveals that quantum chemical methods provide an important means to investigate structures and interactions in nucleic acids. However, judicious selection of

J. Šponer (✉)

Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic
e-mail: sponer@ncbr.chemi.muni.cz

M.K. Shukla

US Army Engineer Research and Development Center, Vicksburg, MS, USA
e-mail: Manoj.K.Shukla@usace.army.mil

J. Wang • J. Leszczynski

Department of Chemistry and Biochemistry, Interdisciplinary Center for Nanotoxicity, Jackson State University, Jackson, MS, USA
e-mail: jingw@icnanotox.org; jerzy@icnanotox.org

computational approach is necessary, depending upon the nature of the problem under investigation.

Introduction

Over the past several decades, investigations of the structure and property of nucleic acids have been an important subject of scientific research. Such investigations have been motivated by the fundamental roles played by ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) in biology.

DNA has been considered to be the central biology molecule, being the depository of genetic information, where hereditary information (of higher-level life forms) is encoded in the form of specific sequences of hydrogen bonds formed between the purine (adenine and guanine) and complementary pyrimidine (thymine and cytosine) bases. Obviously, any permanent variation in the hydrogen bonding pattern may change the function and can even be lethal (especially when specific mutations accumulate). On the other hand all kinds of beneficial mutations, ranging from small-scale mutations (such as point mutations, insertions and deletions) to large-scale mutations (such as chromosomal translocations), are the driving force of evolution. This demands that the structural integrity of DNA be maintained for the identity of each and every organism. It is, however, notable that the biology of DNA does not depend only on the “digital” information of the base sequence. In reality, key aspects of DNA storage in the chromatin and all major aspects of DNA-based control of gene expression are regulated by the subtle sequence-dependent variability of conformational and physicochemical properties of DNA double helix, which is definitely everything else than just a regular double helix.

RNA was, until the 1980s, considered a boring and unimportant cousin of DNA. However, since earth-shaking discovery of RNA catalysis in 1982 (Guerrier-Takada et al. 1983; Kruger et al. 1982) (1989 Nobel Prize for chemistry, Cech and Altman), major RNA discoveries keep coming one after another. We now assume that the RNA molecule is likely the primary molecule of life, the first modern replicator, which in later stages of primary evolution created chemically more stable DNA for better coding and proteins for more diverse catalysis. Since the early stages of evolution, RNA has kept control over many key processes in cellular life while also acquiring new functions. Thus, in the last two decades in biology and biochemistry much of the research focus shifted from DNA to RNA (not reflected by adequate efforts in the field of computational chemistry), as exemplified, e.g., by the 2006 Nobel Prize in Physiology and Medicine awarded to Fire and Mello for their 1998 discovery of RNA interference (Fire et al. 1998) and 2009 Chemistry Nobel Prize to Ramakrishnan, Steitz, and Yonath for solving (in 2000) the atomic resolution structures of the most formidable molecular machine, the ribosome (Ban et al. 2000; Wimberly et al. 2000). A ribosome is a large RNA assembly that has been in the course of evolution supplemented by dozens of ribosomal proteins. It is now well established that while less than 2 % of the genomic DNA codes protein sequences, over 80 % of the genome is actually transcribed into RNA during the cellular life,

most of it (obviously besides genes coding, e.g., ribosomal and transfer RNAs) as noncoding RNA molecules of yet unknown function. In other words, what was not a long time ago considered as “junk” DNA with no obvious role has emerged as DNA template coding for critically important regulatory RNAs that probably affect every corner of gene expression and regulation. One example is the discovery of small RNA molecules (microRNA) that regulate gene expression at multiple levels. The structural versatility and complexity of RNA molecules is incomparably larger than the variability of DNA.

Although the research related to nucleic acids has grown manifold, giving rise to various distinct research fields, the age-old question as to the origin of life on earth remains to be answered. We do not know if it was a spontaneous process which, over the period of time, evolved into the current form on earth or it was endowed from outer space. In fact, we still do not know all the species present on our planet. The existence of different simple molecules, which can be precursors of genetic monomers, such as water, carbon mono- and dioxides, formaldehyde, nitrogen, hydrogen cyanide, hydrogen sulfide, and methane, have been shown in cometary comas (Mix 2006). The purine base adenine has been observed in asteroids and comets. The existence of significant amounts of HCN and HNC molecules in the interstellar space is well known (Ishii et al. 2006). Tennekes et al. (2006) have measured the distribution of these isomers (HCN and HNC) in the protostellar dust core. Smith et al. (2001) have discussed the formation of small HCN-oligomers in the interstellar clouds. It has been demonstrated experimentally that under certain conditions adenine can be formed from the pentamerization of HCN in the solid, liquid, and gas phases (Miller and Urey 1959; Ponnampertuma et al. 1963). Based on the appearance of the brown-orange color as the consequence of impacts of comet P/Shoemaker-Levy 9 on the planet in 1994, the presence of HCN polymers has been speculated on Jupiter (Matthews 1997). The coloration of the Saturn has also been speculated to be due to the presence of HCN polymer. To understand prebiotic adenine synthesis, Glaser et al. (2007) recently performed theoretical calculations on the pyrimidine ring formation of monocyclic HCN-pentamers and found that the key steps proceed without any catalysts producing the purine ring under photolytic conditions and no activation barrier was involved. Roy et al. (2007) proposed a step-by-step pathway for formation of adenine from 2,3-diaminomaleonitrile (DAMN) and 4-amino-5-cyanoimidazole (4-aminoimidazole-5-carbonitrile, AICN) under prebiotic conditions. Experiments for abiotic adenine and purine synthesis from formamide were also reported (Hudson et al. 2012), and a related self-catalyzed mechanisms for the reactions were investigated theoretically (Wang et al. 2013a–c). Based on GC-MS, high-resolution FTIR spectroscopic results and theoretical calculations, an interdisciplinary effort by Ferus et al. (2015) simulated the high-energy synthesis of nucleobases from formamide during the extraterrestrial impact body of the Late Heavy Bombardment (LHB) period. This demonstrated that nucleobases may be synthesized in an impact plasma through reactions of the dissociation products of formamide, without any catalyst.

It is a valid assumption to speculate that life on earth probably evolved under acute harsh environments including the presence of different kinds of irradiations.

Therefore, it is expected that several structural transformations/refinements with respect to genetic code preservation must have taken place. Survival of the fittest prevailed over the period of time in evolution, thereby bringing the purine and pyrimidine bases as the genetic molecules. Further, since these molecules absorb ultraviolet (UV) irradiation, some sort of mechanism was needed to avoid the excited state photo reactions. This was probably achieved through the ultrashort excited state lifetimes of nucleic acid bases (de La Harpe and Kohler 2011; Middleton et al. 2009; Serrano-Andres and Merchan 2009; Shukla and Leszczynski 2007, 2008). Recent state-of-the-art investigations have suggested that such ultrafast excitation processes are achieved through internal conversion where excited and ground state potential energy surfaces conically intersect (Barbatti et al. 2010; Bisgaard et al. 2009; de Vries and Hobza 2007; de La Harpe and Kohler 2011; Kohler 2010; Middleton et al. 2009; Serrano-Andres and Merchan 2009; Shukla and Leszczynski 2007, 2008; Yamazaki et al. 2008). And thus the absorbed energy is efficiently dissipated in the form of heat. Therefore, it is not unexpected that accurate structural determination of nucleic acids and their fragments have been one of the fundamental areas of research. Another obvious precondition that has to be fulfilled in the initial selection of a nucleobase is its inability to tautomerize in aqueous solution, as tautomers would bias any nucleobase-based genetic code as well as RNA folding; thus none of the native nucleobases tautomerize under biochemical conditions to any appreciable amount.

Computational quantum chemical techniques are fast becoming an attractive alternative to the expensive and time-consuming experimental methods in determining the structures and activities of molecular systems. Although we would like to state that modeling of the exact experimental environment, in particular the large biological systems *in vivo*, is not yet possible, computational methods can still provide reliable predictions and thus can be useful to experimentalists. Theoretical methods are especially attractive in the area where experimental measurements are still not possible, e.g., the determination of excited state geometries of complex molecular systems. For the smaller molecular species, one can routinely apply a high level of electron correlated methods and large basis sets; however, for larger molecular systems one has to make a compromise between the level of theory and the basis set and thus with computational accuracy. Experimental determination of excited state geometries of complex molecules like nucleic acid bases is still not possible; only some limited information such as possibility of excited state nonplanarity can be deduced. On the other hand, quantitative information about the excited state geometries can be obtained using the reliable theoretical level, although one has to make a compromise between the theoretical method and the size of the system under investigation. Further, while there was some indication about the amino group nonplanarity in nucleic acid bases in the crystal environment (McMullan et al. 1980), quantitative prediction about amino group nonplanarity was obtained through the quantum mechanical calculations about a decade ago (Leszczynski 1992). It should be noted that such nonplanarity in the gas phase of molecules using experiment was obtained by Dong and Miller in 2002 (Dong and Miller 2002). However, we believe that theoretical and experimental methods are

complementary to each other, and a judicious decision is needed for their efficient application. One of the classical examples would be the tautomerism in guanine.

It is well known that due to the lowest ionization energy among nucleic acid bases, guanine is the primary target for nucleic acid damage by ionizing irradiation (Crespo-Hernandez et al. 2004; Lin et al. 1980). Further, low-energy electrons can also cause the strand break (Boudaffa et al. 2000; Gu et al. 2006, 2007a, b, 2010a–d, 2011a, b; Kumar and Sevilla 2007; Cauet et al. 2014). A combination of molecular dynamics and density functional theory approaches was applied to simulate the reactions that can damage DNA by the attachment of a low energy electron to the nucleobase (McAllister et al. 2015). A low-energy electron interaction with the phosphate group in DNA molecule demonstrates a single-strand break pathways (Bhaskaran and Sarma 2015). DNA labeled with electrophilic nucleobases was reported to be damaged by ionizing or UV radiation (Rak et al. 2015). A recent experimental study on single-strand DNA oligonucleotide suggests that there is a linear correlation between the low-energy electron-induced DNA damage and the presence of the guanine molecules in the sequence (Solomon et al. 2009). Further, guanine can potentially form the most diverse set of energetically accessible rare tautomers in nonpolar environments. Initially, based upon the infrared (IR) spectroscopic analysis of guanine in the argon matrix the presence of equal proportions of keto and enol forms was suggested (Sheina et al. 1987). But, the canonical form of guanine dominates in the polar solvent (Leszczynski 2000). Theoretical methods have generally predicted that the keto-N7H tautomer of guanine is the most stable in the gas phase, but the keto-N9H tautomer dominated in the water solution. At the MP2 and CCSD(T) levels along with several large basis sets, the four low-energy tautomers of guanine (keto-N9H, keto-N7H and cis and trans forms of enol-N9H) have been shown to be within 1 kcal/mol of energy (Gorb et al. 2005). The assignments of resonance-enhanced multiphoton ionization spectra of laser-desorbed, jet-cooled guanine have suggested the presence of four tautomers of guanine (Mons et al. 2002; Nir et al. 2001). Based upon the comparison of IR spectra of thermally vaporized guanine trapped in helium droplets with that of computed vibrational frequencies at the MP2 level with the 6-311++G(d,p) and aug-cc-pVDZ basis sets, Choi and Miller (2006) have assigned the presence of keto-N9H, keto-N7H, and cis and trans forms of enol-N9H tautomer of guanine. The results of guanine in helium droplets necessitated the reassignment of earlier R2PI data and accordingly, based upon the comparison of experimental and theoretical results, Mons et al. (2006) found the presence of enol-N9H-trans, enol-N7H, and two rotamers of the keto-N7H-imino tautomers of guanine in the supersonic jet-beam. Thus, in the new assigned R2PI spectra the stable keto-N9H and keto-N7H tautomers of guanine are not present. The high-level of theoretical calculations (Chen and Li 2006; Marian 2007) on guanine tautomers also supported the reassignment of the R2PI spectra of guanine tautomers in the supersonic jet-cooled beam and suggested the presence of efficient nonradiative deactivation channels as the reason for the missing of spectral origins of the stable tautomers in the R2PI experiments. Recently, Zhou et al. (2009) have performed a comprehensive investigation of guanine tautomers using the VUV photoionization technique, where

the gas phase of guanine was obtained at both the thermal vaporization and laser desorption methods. It was revealed that the method used to generate the gas phase sample of guanine has significant influence on the population of tautomers in the experiment. Consequently, it was found that in the thermal vaporization, a maximum of five most stable tautomers are populated and these results are in agreement with that obtained in the helium droplet experiment. On the other hand, when the laser desorption technique was used to make a gas phase sample of guanine, up to seven tautomers are populated. As noted above, however, guanine for obvious reasons does not tautomerize in biochemically relevant environments, as convincingly shown more than a decade ago also by advanced computational methods (Colominas et al. 1996). Eventual computations suggesting formation of tautomers of natural bases in water are to be dismissed and are at odds with all other experimental data.

Another unique feature of QM methods is their capability to reveal relation between molecular structures and molecular energies at the level of direct (contact, gas phase, electronic structures) interactions. This concerns mainly the two fundamental interactions, base stacking and base pairing. Stacking interactions play an important role in the biological structure, providing both thermodynamics stability and structuring of nucleic acids. For example, base stacking is assumed to be the primary determinant of sequence dependence of B-DNA structure and flexibility, which is the single most important feature of DNA that enters all of DNAs molecular interactions, genetic material storage, replication, and gene expression. The sequence dependence of DNA (and the role of base stacking) has been intensely studied. Despite this, all the experimental and theoretical research failed to provide clear rules allowing to predict the fine B-DNA structural variability from sequence (Calladine 1982; Dickerson and Drew 1981; Sponer and Kypr 1991; Suzuki et al. 1997; Wing et al. 1980; Yanagi et al. 1991). The research in this area has been stalled for some time and perhaps improved description of stacking with the help of modern QM methods could bring some new ideas.

Base pairing (extended beyond the base-base interactions) is especially interesting in large functional RNAs, where it determines their architectures and gives very strict constraints on RNA evolutionary patterns (Leontis et al. 2002; Sponer et al. 2010; Stombaugh et al. 2009; Zirbel et al. 2009). Large RNA molecules are organized as complex (and often very dynamical) jigsaw puzzles, where the exact shapes of the base pairs determine function and thus also allowable mutations (isostericity principle). Nevertheless, recent studies also indicate a nonnegligible role of energy of molecular interactions supplementing the basic isostericity principle (Zirbel et al. 2009).

Although stacking interactions are important and high-level *ab initio* methods are needed to account for such interactions, it is only possible to use these methods to model systems and very small fragments of large biomolecules, even using the large computational resources. Therefore, especially force field methods (and QM/MM method to certain extent) are mainly used to study larger biopolymers and other large biological systems. Nevertheless, QM calculations remain instrumental in reference studies on the nature and magnitude of molecular interactions in nucleic

acids and in verification of the other methods. Quantum chemical calculations provided the ultimate answer about the physicochemical nature of base stacking and characterized many other features of nucleobase interactions (Hobza and Sponer 1999; Morgado et al. 2009; Svozil et al. 2010).

Hydrogen Bonding and Stacking Interactions in Nucleic Acids

The most fundamental roles of nucleic acid bases (nucleobases) in biology and chemistry are their involvement in two qualitatively different mutual interactions: hydrogen bonding (base pairing) and aromatic base stacking.

The base pairing is utilizing the H-bond donor and acceptor capabilities of nucleobase exocyclic groups and ring nitrogen atoms. In RNA molecules, the base pairing also involves the sugar ribose which, in contrast to DNA deoxyribose, possesses hydroxyl group in the 2' position (Leontis et al. 2002; Sponer et al. 2005a, 2007, 2009, 2010; Stombaugh et al. 2009). The 2' -hydroxyl group is a powerful donor and acceptor of hydrogen bonds. In fact, key RNA base pair families utilize the 2' -OH group for base pairing, and these extended base pairs are known as sugar-edge (SE) base pairs or interactions. Many important "SE" base pairs include no direct base to base H-bonds, and yet they are crucially important for folding of complex functional noncoding RNA molecules and ribonucleoprotein particles. By noncoding RNAs we mean RNA molecules that are not translated to proteins via messenger RNA and perform different functions instead. Note that recent research highlighted that while less than 2 % of the human genome directly codes for proteins, at least 80 % of the genomic DNA is actually transcribed into RNA. Thus, majority of the genome encodes noncoding RNAs that play absolutely essential roles in life and evolution (many of the RNA functions have yet to be discovered, but they are assumed to be key players in fine regulation of gene expression), which is a finding that has truly revolutionized biology in recent years. The largest noncoding RNAs are, obviously, ribosomal RNAs. The most important RNA tertiary interactions (A-minor and P-interactions) are base pairs, triads, and quartets mediated by base-sugar and sugar-sugar interactions (Sponer et al. 2007). Recently, the RNA base pairing classification was extended to include base-phosphate (BPh) interactions, after recognizing that ~ 12 % of nucleotides in ribosome are involved with BPh interactions with other proximal or distal nucleotides while these interactions bring important evolutionary constraints (Zirbel et al. 2009).

Base stacking occurs between the aromatic faces of the nucleic acid bases and is at least equally as important as base pairing, for both thermodynamics stabilization and shaping of nucleic acids. Stacking is responsible for the local conformational variations and other sequence-dependent properties of B-DNA. With the help of eight micro-solvation water molecules, the dinucleoside phosphate deoxyguanylyl-3',5'-deoxycytidine dimer octahydrate, [dGpdC]₂ illustrative minimal unit of the DNA double helix has been constructed and its geometry fully optimized (B3LYP/6-31 + G(d,p) level) (Gu et al. 2011a). Similarly, four water molecules microhydrated

dApdT, dTpdA, dGpdC, and dCpdG species have been predicted to have reasonable stacking structure using both B3LYP and M05-2x functionals (Gu et al. 2009, 2010c). The Minnesota density functionals, M05-2X and M06-2X, level calculations were carried out to evaluate the stacking and H-bonding patterns for the structures of the nucleotide oligomers (dGpdC, dGpdG, and dGpdCpdG) for single-strand DNA. This level of calculations shows excellent agreement with the results from the MP2 theory (Gu et al. 2011a, 2012a, b). Unambiguous classification of base stacking is missing. One of the reasons is flexibility of base stacking, as the stacked bases can always slide and twist over a range of mutual stacked geometries, being not fixed by the individual H-bonds (note, however, that many complex RNA base pairs also possesses complex conformational space with multiple competing conformations). While classification of base pairing could have been done purely by considering structural data (i.e., geometries seen in X-ray structures), classification of base stacking will likely require appropriate energy analyses.

The *ab initio* QM technique can be used to determine optimal structures of molecular clusters and to calculate energies for any single geometry of the cluster. QM calculations provide molecular wave functions, which can be used to derive physicochemical properties, such as vibrational spectra, dipole and higher multipole moments, polarizabilities, proton affinities, NMR parameters, and others. Nevertheless, the main achievement of QM calculations was the description of the nature and energetics of nucleobase interactions. This is because the leading experimental approaches of structural biology, that is, mainly X-ray crystallography, provide purely structural data. Information about energetics of molecular interactions can be inferred only indirectly, while the interpretation of structural data ignoring energetics of molecular interactions is often misleading.

QM calculations can help to understand the role of molecular interactions in nucleic acids because of their capability to give a direct link between structures and energies. Nevertheless, QM calculations are always done on small systems and typically in the gas phase that is far from real environments and structural contexts. To make meaningful QM computations with biological relevance, we need to follow several rules (Svozil et al. 2010).

First, we need to make the strategic decision whether our computations aim to be indirectly or directly relevant to biology. By indirect relevance we mean, for example, the use of QM calculations for parameterization or validation of other methods (mainly the molecular mechanical force fields) or for basic understanding of the physical chemistry of interactions. By direct relevance we mean applications that range from calculations of some specific interaction patterns seen in structural studies (Sponer et al. 2003) through combined QM-bioinformatics studies aimed at classifying interactions (Sponer et al. 2003; Zirbel et al. 2009) up to QM/MM calculations of RNA catalysis (Banáš et al. 2009). Then, there are at least three tactical issues that need to be very carefully decided. We need to select the level of calculations, i.e., primarily the method and basis set. Equally important is the appropriate choice or preparation of geometries used in computations. Inappropriate geometries may easily blow up the whole effort and result in computations that are misleading (Svozil et al. 2010). Finally, the results should be properly interpreted. We also need to separately consider applications that deal with nucleic acid bases

but are directed to other areas of science (basic physicochemical experiments, origin of life studies, adsorption on surfaces, supramolecular assemblies, etc.)

Level of Computations

Let us assume that we have a dimer of two nucleic acid bases, A and B , with a given geometry. For that geometry, the interaction energy between A and B , $\Delta E^{A \cdots B}$, is the energy difference between the total electronic energy of the dimer $E^{A \cdots B}$ and the electronic energies E^A and E^B of isolated bases.

$$\Delta E^{AB} = E^{AB} - E^A - E^B \quad (1)$$

The interaction energy reflects a hypothetical dimerization process at 0 K and is not measurable. In order to be related to experimental dissociation energies D_0 and enthalpies of formation, the deformation energy of monomers and the zero-point vibration energy must be included. The zero-point energies and enthalpy and entropy contributions at nonzero temperature are usually calculated in the harmonic approximation. Since base pair complexes are weak, anharmonicity can play an important role, especially for stacked systems and particularly at higher temperatures. Nevertheless, except for direct comparison with gas phase experiment, interaction energy evaluation is the sufficient outcome of QM analysis.

Methods suitable for base stacking and base pairing calculations have been discussed many times (Morgado et al. 2009; Spomer et al. 2008; Gu et al. 2008; Svozil et al. 2010) and will thus be only briefly noted here. With modern computers, we have methods that are satisfactorily accurate.

Base stacking stabilization is dominated by the intermolecular electron correlation effects (i.e., the dispersion energy). Therefore, stacking calculations must be done with inclusion of electron correlation effects and with large basis sets of atomic orbitals. The dispersion energy is created in the space between the interacting monomers that are separated by $\sim 3.3\text{--}3.4$ Å. This space needs to be covered by atomic orbitals, dictating the use of diffuse-polarized basis sets. H-bonded complexes are not dominated by the dispersion energy, albeit it is still a very significant contribution. Thus, HF calculations or computations with “dispersion-neglecting” DFT methods, while not being accurate, are not entirely incorrect. When including electron correlation, higher angular momentum functions are important for base pairing, since the space between the interacting monomers is bridged by H-atoms so the requirement for the diffusivity of atomic orbitals is not as strict as for stacking.

The best accuracy is achieved by complete basis set (CBS) extrapolation methods, when two systematically improved basis sets are applied and the data is then extrapolated. The interaction energy computations, even with large basis sets, need to be corrected for basis set superposition error (BSSE). We oppose suggestions to ignore the BSSE correction or to attempt only its partial inclusion while assuming that the numbers can be correct due to error cancellation. This is a risky game. It is much better to provide BSSE-corrected numbers where a solid

estimate of the underestimation of the interaction is typically possible. Fortunately, the CBS calculations are intrinsically BSSE-free. Similarly, computations with modern parameterized DFT-D methods (see below) do not require BSSE correction, since it is indirectly (effectively) included via parameterization.

H-bonded complexes are rather well described by the MP2 method while for aromatic stacking this method typically significantly overshoots the stabilization. Thus, for stacking higher-order electron correlations are quite important.

Gold Standard

Within the variation (supramolecular) approach, definitely the method of choice for interaction energies would be the coupled cluster CCSD(T) method (in which the single and double excitations are evaluated iteratively while the triple excitations are included in a noniterative way). The CCSD(T) method yields a significant portion of the correlation energy. The MP2 method, including the double electron excitations at the second order of perturbation theory, overestimates the correlation interaction energy for stacking, as noted above.

The determination of a CBS limit of CCSD(T) calculations is still difficult. Until recently, the CCSD(T) calculations for larger complexes were performed only with medium basis sets (e.g., 6-31G*) and even these calculations were at the computer limits. Thus, the gold standard in base stacking and pairing calculations is the method sometimes abbreviated as CBS(T), which utilizes the similar basis set dependence of the CCSD(T) and MP2 energies. Thus, the difference between CCSD(T) and MP2 interaction energies ($\Delta E^{\text{CCSD(T)}} - \Delta E^{\text{MP2}}$) exhibits small basis set dependence, and the CCSD(T)/CBS interaction energy can be approximated as

$$\Delta E_{\text{CBS}}^{\text{CCSD(T)}} = \Delta E_{\text{CBS}}^{\text{MP2}} + \left(\Delta E^{\text{CCSD(T)}} - \Delta E^{\text{MP2}} \right) \Big|_{\text{medium basis set}},$$

which is abbreviated as CBS(T) to distinguish from full CCSD(T)/CBS computation (Sponer et al. 2006, 2008). Various extrapolation schemes have been suggested for the determination of the $\Delta E_{\text{CBS}}^{\text{MP2}}$ term; the one proposed by Halkier et al. (1999) is the most widely used.

For aromatic stacking interactions the Δ CCSD(T) correction term is systematically nonnegligible (repulsive) and should never be omitted. For H-bonding interactions, the Δ CCSD(T) correction term is typically very small (Sponer et al. 2004, 2006, 2008).

Other Approaches

Earlier calculations on base stacking were done with the MP2 method utilizing the modified 6-31G*(0.25) basis set (Hobza and Sponer 1999; Sponer et al. 1996b, 1997). The polarization d-functions of the standard 6-31G* basis with an exponent

of 0.8 were replaced by more diffuse ones with an exponent of 0.25, allowing inclusion of a major part of the dispersion energy. While the MP2/6-31G * (0.25) method is now outdated, the main conclusions reached by the MP2/6-31G * (0.25) studies remain valid.

DFT methods were for years not recommended for stacking calculations, because common DFT methods (based on the local density, its gradient, and the local kinetic-energy density) notoriously fail to capture the (nonlocal) dispersion energy (Hobza et al. 1995; Kristyán and Pulay 1994).

This is a common feature of all LDA and GGA functionals, not excluding even the most advanced meta-GGA functionals. Many recent DFT methods provide much better results for dispersion-controlled complexes (Zhao and Truhlar 2008).

Nevertheless, we still suggest use of caution (and testing) in their application to stacking complexes. An alternative (which can achieve, at least for now, better accuracy and major speed up) is based on augmenting the DFT energy by an empirical London dispersion energy term (Elstner et al. 2001; Grimme 2004; Jurecka et al. 2007).

To correct for the overlap effects, the dispersion energy is damped by distance-dependent damping function. The dispersion energy, represented by the C_6/R^6 formula, is calculated separately from the DFT calculation and is simply added to the DFT energy. The disadvantage of DFT-D methods is obviously the need to combine electronic structure calculations with classical “force field” correction term, which also affects the transferability of these methods. Thus these methods are expected to be surpassed in the future by “true” DFT-based dispersion-including methods; however, for the moment it seems to us that DFT-D is more satisfactory for routine calculations of nucleobase interactions. One present difficulty is that we have so many new methods in the literature that is difficult to choose. This issue is beyond the scope of this chapter and we refer the reader to specialized literature (Banaš et al. 2009; Spöner et al. 2008). We hope that standard (optimal) wide-spectrum dispersion-including DFT methods will soon be identified.

Interaction energies can also be obtained by perturbation methods, as a sum of perturbation contributions. The symmetry adapted perturbation theory (SAPT) provides the interaction energy as a sum of first-, second-, and higher-order perturbation terms (Heßelmann et al. 2005; Jeziorski et al. 1994). The first-order contribution contains the electrostatic and exchange energies, while the second-order term includes induction and dispersion energies. The charge transfer energy is included in the second-order induction energy and higher-order contributions. SAPT (with extended basis sets) yields accurate values of the energy components and also of the total interaction energies. The determination of the interaction energy is straightforward and is not biased by additional theoretical problems, such as the BSSE inherent to variational methods. The broad use of SAPT is, however, hampered by large computer requirements. A significant improvement was reached by the combination of SAPT and DFT theories (Jeziorski et al. 1994). The DFT-SAPT approach has been rather routinely used for base-base calculations. When making SAPT decompositions, it is extremely important to use well-defined geometries, as the SAPT components are exceptionally sensitive to inter-monomer

separation, much more than the total interaction energies. SAPT decompositions can be spoiled by inappropriate choice of geometries (Sponer et al. 2008). Note also that from the biological point of view, what primarily matters are the interaction energies. Thus the usefulness of decompositions should not be overinterpreted (Sponer et al. 2008).

Geometries

Quantum-chemical calculations can provide meaningful data only when the energies are derived at appropriate geometries.

The easiest systems to deal with are well-behaved base pairs where gradient optimization leads to relevant structures. Modern QM programs allow easy gradient optimizations of base pairs, where all coordinates (or parameters) are optimized. Standard optimizations are not corrected for BSSE. Some earlier studies where base pairs were optimized in a step-by-step manner are of historical interest only. Since the optimization itself is more computer-demanding than the subsequent interaction energy evaluation, very often a better level of theory (level X) is used for interaction energy calculation than for optimization (level Y). This is abbreviated as X/Y. For example, the abbreviation MP2/aug-cc-pVTZ/MP2/cc-pVDZ indicates that the optimized structure was obtained at the MP2/cc-pVDZ level, while the energies were derived for this optimized geometry at the MP2/aug-cc-pVTZ level.

The gradient optimization is good for systems with well-defined local minima, while the minima correspond to the biochemically relevant structures. Stacking patterns seen in nucleic acids do not correspond to minima on the potential energy surfaces of isolated stacked dimers and thus conformational scanning is preferred. Further, gradient optimizations of dispersion-controlled clusters are affected by enormous BSSE, unless a very large basis sets are used. With lower quality methods the structures are unstable and convert to H-bonded ones. In addition, gradient optimizations of stacked dimers lead to puckering of the aromatic ring (Hobza and Sponer 1998). This is usually not desirable since in real environments the nucleobases have some interactions at both their sides, probably preventing such puckering. Thus, stacking calculations are mostly carried out as a series of single points with fixed geometries and rigid monomers (Sponer et al. 1996b, 1997).

An attractive option is to take structures from experimental (X-ray) studies (Sponer et al. 1997). However, usual accuracy of these experiments does not guarantee their straightforward utilization in QM energy computations. First, it is not advisable to directly use monomer geometries from PDB files of NA X-ray structures. Due to limited resolution the monomer geometries carry limited experimental information about monomer geometries, while the bases are often deformed after the refinement. Such deformed monomer geometries frustrate the electronic structure. It is necessary to replace (for example via overlay) the monomers from the PDB files via QM-optimized monomers.

In addition, intermolecular X-ray geometries may cause substantial errors in calculations. Especially drastic distortions of the calculated energies can be introduced

by steric clashes in the refined crystal structures (Sponer and Kypr 1993). A real nightmare occurs when the X-ray base stacks are effectively vertically compressed or extended due to inaccurate determination of the interbase angles (propeller twist, base pair roll, etc). This requires a case-by-case judgment and some experience with crystallography. Note that a rather small error in the interbase distances (which may still be tolerable from the geometry point of view) can lead to a considerable energy artifact. This is when the geometry falls into a region of interatomic distances where the short-range repulsion starts to dominate. The calculated energy is a highly nonlinear function of the interatomic distance (Sponer et al. 2008).

Similarly, H-bonded base pairs are sensitive to experimental geometry errors due to the genuine close contact between H-bond partners. Besides data and refinement errors a bad geometry can result from the presence of two or more local substates. Substates cannot be distinguished except as having nominal resolution better than $\sim 1 \text{ \AA}$. The refined geometry reflects an averaged geometry which may have very poor energy. Fiber diffraction models cannot be recommended for direct calculations (Svozil et al. 2010).

We would like to caution against using averaged (3D-bioinformatics) geometries, as they can represent unrealistic single structures from the energy point of view. It is always advisable to generate a range of structures around such geometries and to analyze the properties of the potential energy surfaces. Actually, an open question remains whether the base stacking can be characterized by some single representative geometry. Most likely stacking states correspond to a range of populated geometries, as evidenced, for example, by significant coordinate fluctuations seen for stacked bases in explicit solvent molecular dynamics simulations (Svozil et al. 2010).

We are interested in analyses of specific interactions which are neither stacking nor H-bonding and are substantially affected by the overall topology of the studied systems. The best approach is to fix the intermolecular geometry of interest (typically using a set of three dihedrals and two valence angles plus one intermonomer distance per each dimer) and then relax the monomers intramolecularly (Sponer and Hobza 1994; Sponer et al. 2003; Vlieghe et al. 1999). This approach has been applied in studies of cross-strand close amino group contacts in B-DNA, DNA-drug interactions, bifurcated H-bonds, out-of-plane H-bonds, and some other interactions. If a steric clash is suspected, then the intermonomer distance can again be varied (Sponer and Hobza 1994; Sponer et al. 2003; Vlieghe et al. 1999).

A specific problem is represented by the complex RNA base pairing patterns involving the sugar edges (Sponer et al. 2005a, b, 2007). Many of these base pairs have multiple minima. For many of them, the functional (observed) structures do not correspond to any intrinsic gas phase minimum energy structures, since they are constrained by other interactions and the overall RNA topology. Some base pairs can be intrinsically water-mediated. Huge problems in computations can be created by the sugar hydroxyl group in position 3', which normally is involved in the covalent backbone chain. One option is 3'-methylation. In some cases, the phosphate groups participate in the interactions and need to be included in computations. This creates problems due to a strong ionic nature of the associated interactions which are, in

real systems, obviously attenuated by solvent screening. Close to insurmountable electrostatic problems arise when more than one phosphate directly participates in the interaction. Thus, QM studies of RNA base pairs often require applications of sophisticated geometrical constraints which need to be implemented case by case. For nonneutral systems, even optimizations upon inclusion of continuum solvent could be a viable option. The situation is further complicated by very limited resolution of the experimental structures of folded RNAs (not mentioning ribosome) and their dynamical nature. This can lead to large coordinate errors (including poorly refined *syn* vs. *anti* orientation of the bases or incorrect sugar puckers). Thus, studies of RNA base pairs are far from routine. Studies of geometries that are substantially rearranged compared to the experimental structures are of a little value, similar to studies neglecting the sugar rings for base pairs where the 2'-OH groups are directly involved in base pairing.

When gradient optimization is carried out, the monomer geometries are changed upon complexation. This is due to mutual adaptations of the monomers that improve the intermolecular interaction at the expense of the intramolecular energy terms. Some of the deformations can be directly related to the binding strength. However, for larger systems, some monomer rearrangements reflect rather long-range effects. For example, there could be a substantial reorientation of the flexible sugar-phosphate backbone upon complex formation. Thus, real deformations consist of two contributions. Direct deformations (always present) reflect the strength of the binding and may be complemented by various indirect larger-scale conformational rearrangements. Besides real deformations, the BSSE contributes to the deformation when standard gradient optimization is applied. The BSSE contribution is obviously a computational artifact.

As explained above, the interaction energies of the optimized complexes should be a posteriori corrected for the BSSE using the geometry of the complex and dimer-centered basis set. Then we separately calculate the deformation energy using the monomer basis sets, as difference of monomer energies in the deformed (complex) and optimized (isolated) monomer geometries.

$$E_{\text{Def}}^A = E^{A(\text{dimer geometry})} - E^{A(\text{monomer geometry})}$$

Thus, the interaction energy of a dimer is defined in the following way:

$$\Delta E^{A\dots B} = E^{A\dots B} - (E^A + E^B) + E_{\text{Def}}^A + E_{\text{Def}}^B.$$

The first three energies are calculated in a dimer-centered basis set. The intramolecular deformation energy actually cancels a large part of the intermolecular energy improvement caused by mutual monomer adaptations.

In some of the literature, the authors include deformation energy formally as part of the BSSE correction. We consider this a weird option which may substantially spoil the interpretation of the results. Although it might look more sophisticated mathematically, this approach is misleading, mixing apples and oranges, and is especially unsuitable for larger systems such as base pairs and other fragments

of biopolymers (Sponer et al. 2004). In fact, the integrated expression is, after formal rearrangements, *entirely identical* to the above definition, which in addition is older (Sponer et al. 1996a), i.e., the correction was commonly known before researchers started to include deformations into BSSE correction. Second, while BSSE is a mathematical artifact, monomer deformations are not. They correspond to fundamental properties of the studied clusters including their vibrational spectra and polarization/charge transfer effects. Thus, it is useful to evaluate the magnitude of the monomer deformations explicitly. For flexible systems with large indirect rearrangements (as explained above) any formal inclusion of the deformation term into the BSSE correction is meaningless. Thus, albeit widespread, for base pairs and larger systems of chemical and biological interest, this approach is not appropriate (Szalewicz and Jeziorski 1998).

An alternative approach is to use counterpoise-corrected gradient optimization where the BSSE is removed in each gradient iteration, although with a substantial increase of the computer requirements. It eliminates the BSSE part of the deformation energies while true deformations persist.

In base pairing studies the deformation energy can be calculated either with respect to the planar monomers, thus neglecting the amino group nonplanarity, or with respect to nonplanar bases. These two numbers differ simply by the difference between energies of planar and nonplanar monomers and can be easily compared when needed (Sponer et al. 2004).

Interpreting the Computations

QM calculations (on nucleobase dimers) reveal the binding energy between two bases in the gas phase, i.e., in *complete isolation*. They thus describe the *intrinsic* interactions of the systems with no perturbation by external effects such as solvent. The intrinsic intermolecular stabilities are directly linked to molecular structures and can be derived in any selected geometry. However, the gas phase interaction energies do *not* correspond to the stability of the interactions in nucleic acids, as measured by thermodynamics experiments. It is not possible to easily correlate the QM calculations with measured base pairing and stacking stabilities in nucleic acids. The apparent (measured) strength of the base-base interactions in nucleic acids in various experiments is determined by a complex interplay of many factors and the intrinsic base-base term is only one of them. Many researchers incorrectly believe that the experiments reflect the “true” stabilities of base-base interactions and vice versa.

A textbook example of complexity of molecular interactions is stacking of consecutive protonated cytosines. This is a highly repulsive interaction in the gas phase due to a charge-charge repulsion (Sponer et al. 1996c). Nevertheless, in intercalated i-DNA quadruplex, stacking of a set of consecutive closely spaced protonated cytosines occurs (Gehring et al. 1993). The i-DNA tetraplex is paired via cytosine – protonated (N3) cytosine base pairs, each possessing charge +1. Both cytosines are equivalent in X-ray and NMR experiments suggesting rather fast intra-base pair proton switches (Chen et al. 1994). Two duplexes intercalate to form

the tetraplex. Stability of i-motif is due to this massive accumulation of closely spaced protonated base pairs. In i-DNA, the vertical repulsion between consecutive protonated base pairs is counterbalanced by solvent screening effects and possibly specific interactions with the anionic backbone (Spackova et al. 1998). Thus i-DNA indeed has, in contrast to other DNAs, *repulsive intrinsic stacking energy terms*. This example clearly demonstrates the actual magnitude of mutual compensation of molecular interactions in nucleic acids. The i-DNA stability contradicts the gas phase stacking energy calculations and demonstrates why we cannot use these calculations to directly predict DNA stability. However, the relation can be also reversed. It is not possible to unambiguously evaluate the *intrinsic* stacking energetics based on thermodynamics studies of nucleic acids. There is no unambiguous way to decompose the measured free energies into separate terms that would correspond to stacking, base pairing, etc. In other words, we cannot make straightforward extrapolation from gas phase to nucleic acids, while, conversely, studies of nucleic acids bring no unambiguous information about the intrinsic base-base terms.

To show the full complexity of molecular interactions, let us underline that the screening is specific for i-DNA. Strikingly contrasting i-DNA is the behavior of consecutive protonated cytosines in C + -G.C triples of Pyr-Pur.Pyr triplexes. Consecutive protonated cytosines would be needed to recognize consecutive guanines in the second strand (Soliva et al. 1998). The vertical position of protonated cytosines in triplex would adopt arrangement closely resembling i-DNA, and also planar H-bonding of the third-strand protonated cytosines to N7 of second-strand guanines resembles the i-DNA base pairing. However, this sharply destabilizes the DNA triplex, and even two or three consecutive CH⁺ are not tolerated. This indicates that in triplex the screening of the vertical electrostatic repulsion by the backbone and solvent is less efficient than in i-DNA. Thus in this particular case of i-DNA and triplexes we cannot transfer experience concerning nucleobase interactions between two DNA forms. Each case should be studied separately. In other words, a given type of base pairing and base stacking may have entirely opposite roles in different nucleic acid forms. A given interaction may be a crucial stabilizing factor for one type of nucleic acid architecture (protonation of consecutive cytosines in the i-motif), while it may be even not tolerated in another architecture. This illustrates that there is no way to design some ultimate experiments to decide about the *common nature* of base stacking in nucleic acids. This simply is a wrong question. In order to understand the interactions in nucleic acids, we need to consider a wide range of systems, and the gas phase data represent an important part of the overall picture.

It is nevertheless clear that proper inclusion of solvent screening could help in interpretation of the QM data. Unfortunately, accurate inclusion of solvent effects into QM calculations is difficult. One option is to extend the studied system by a finite set of explicit water molecules. Such calculations still deal with a gas phase molecular cluster and do not correspond to bulk hydration. The cluster hydration patterns differ from those in water where the first shell waters interact with the second shell, etc., and the whole system is dynamical, as evidenced by large-scale explicit solvent simulations of nucleic acids. In a small cluster, individual

water molecules will form bridges and zippers between bases in order to maximize the number of H-bonds. The hydration picture in solution (MD data) and X-ray structures is different and reveals simple noncooperative in-plane hydration of the polar nucleobase sites (the nitrogens and oxygens). Water binding sites in common hydration sites around nucleic acids have water binding times ~ 50 – 500 ps. In complex RNA molecules or in molecular complexes, some hydration sites may be occupied by tightly bound water molecules (Rěblová et al. 2003). A substantial problem of cluster calculations is that the potential energy surface contains a large number of minima and, without an efficient sampling technique, it is virtually impossible to verify the true global minimum (Kabelác and Hobza 2001).

The other option is to include the solvent as a polarizable continuum. QM methods consider effects of the continuum on the electronic structure of the solute molecules, in contrast to classical continuum approaches. The outcomes are quite sensitive to the choice of parameters such as the atomic radii used to define the “solute” cavity; no universal accurate set of “true” radii can be established. The continuum calculations may be combined with cluster calculations, where the first hydration shell is treated explicitly. Even if QM continuum solvent calculation is properly performed, such calculations are not sufficient to achieve a direct correspondence with thermodynamics experiments. In practice we are neglecting, for example, the loss of degrees of freedom upon duplex formation, all effects associated with the presence of sugar-phosphate backbone including the only partial exposure of the bases to the solvent, etc. In addition, different base sequences are likely associated with different solute flexibilities (some base pair step sequences are stiff, others flexible) which will contribute to the free energy balance via solute entropy contributions. Reliable evaluation of these contributions is outside the applicability of available computational methods. Thus, the continuum solvent calculations should be considered just as rough estimates of the attenuation of the electrostatic contributions to the free energy upon solvation, which are still far from fully competent free energy computations.

All of the abovementioned considerations explain why the QM calculations on base stacking and base pairing do not (and should not) correlate with the measured thermodynamics properties of nucleic acids. Therefore, QM calculations and intrinsic interaction energies should never be interpreted as straightforward determinants of nucleic acid stabilities. This naive overinterpretation of otherwise very valuable computed data can discredit the computations. There is a plenty of evidence suggesting that the role of molecular interactions in thermodynamics stabilization of even the simplest duplex nucleic acids is more complex than usually assumed, and results from tiny, irregular, and case-specific interplays of all molecular forces, where literally a single specific hydration site or pocket can change the balance. Even simple base modifications and substitutions, as small as deletions of a single exocyclic group, may have complex and, at first sight, mutually contradicting context-dependent impacts on the measured thermodynamics stabilities that cannot be a priori predicted (Chen and Turner 2006; Siegfried et al. 2007).

We nevertheless suggest that this enormous complexity of molecular interactions defines new roles for modern computations, combing well-calibrated simulation

approaches and accurate QM calculations. The calculations can provide key insights into the tricky games of molecular interactions that are shaping up the molecules with their associated free energies and that are not fully understandable based on purely experimental approaches (Kopitz et al. 2008; Yildirim and Turner 2005; Yildirim et al. 2009).

Conclusion

QM calculations represent the leading tool to study intrinsic molecular interactions in nucleic acids, such as base stacking and base pairing. However, the QM data should not be overinterpreted, and any extrapolation to nucleic acids requires proper consideration of the gas phase nature of QM calculations. In addition, in order to obtain meaningful QM data, basic methodological requirements must be fulfilled. These include, in addition to the obvious selection of appropriate level of calculations, very careful selection or determination of geometries, which is discussed in detail in this chapter.

Acknowledgments JS acknowledges the support from the Grant Agency of the Academy of Sciences of the Czech Republic (grant IAA400040802), Grant Agency of the Czech Republic (grant 203/09/1476), Ministry of Education of the Czech Republic (LC06030), and Academy of Sciences of the Czech Republic (grants AV0Z50040507 and AV0Z50040702). MKS and JL are thankful for supports from NSF-CREST (grant HRD-0833178) and NSF EPSCoR (grant 362492-190200-01\NSFEPS-0903787) and computational support from Mississippi Center for Supercomputing Research (MCSR).

Bibliography

- Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Angstrom resolution. *Science*, 289, 905.
- Banáš, P., Jurecka, P., Walter, N. G., Šponer, J., & Otyepka, M. (2009). Theoretical studies of RNA catalysis: Hybrid QM/MM methods and their comparison with MD and QM. *Methods*, 49, 202.
- Bao, X., Wang, J., Gu, J., & Leszczynski, J. (2006). DNA strand breaks induced by near-zero-electronvolt electron attachment to pyrimidine nucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 5658.
- Barbatti, M., Aquino, A. J. A., Szymczak, J. J., Nachtigallova, D., Hobza, P., & Lischka, H. (2010). Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 21453.
- Bhaskaran, R., & Sarma, M. (2015). Low-energy electron interaction with the phosphate group in DNA molecule and the characteristics of single-strand break pathways. *The Journal of Physical Chemistry. A*, 119, 10130.
- Bisgaard, C. Z., Satzger, H., Ullrich, S., & Stolow, A. (2009). Excited-state dynamics of isolated DNA bases: A case study of adenine. *ChemPhysChem*, 10, 101.
- Boudaffa, B., Cloutier, P., Haunting, D., Huels, M. A., & Sanche, L. (2000). Resonant formation of DNA strand breaks by low-energy (3–20 eV) electrons. *Science*, 287, 1658.
- Calladine, C. R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. *Journal of Molecular Biology*, 25, 343.

- Cauet, E., Bogatko, S., Lievin, J., De Proft, F., & Geerlings, P. (2014). Electron-attachment-induced DNA damage: Instantaneous strand breaks. *The Journal of Physical Chemistry. B*, *117*, 9669.
- Chen, H., & Li, S. (2006). Theoretical study on the excitation energies of six tautomers of guanine: Evidence for the assignment of the rare tautomers. *The Journal of Physical Chemistry. A*, *110*, 12360.
- Chen, G., & Turner, D. H. (2006). Consecutive GA pairs stabilize medium-size RNA internal loops. *Biochemistry*, *45*, 4025.
- Chen, L., Cai, L., Zhang, X., & Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry*, *33*, 13540.
- Choi, M. Y., & Miller, R. E. (2006). Four tautomers of isolated guanine from infrared laser spectroscopy in helium nanodroplets. *Journal of the American Chemical Society*, *128*, 7320.
- Colominas, C., Luque, F. J., & Orozco, M. (1996). Tautomerism and protonation of guanine and cytosine. Implications in the formation of hydrogen-bonded complexes. *Journal of the American Chemical Society*, *118*, 6811.
- Crespo-Hernandez, C. E., Arce, R., Ishikawa, Y., Gorb, L., Leszczynski, J., & Close, D. M. (2004). Ab initio ionization energy thresholds of DNA and RNA bases in gas phase and in aqueous solution. *The Journal of Physical Chemistry. A*, *108*, 6373.
- de La Harpe, K., & Kohler, B. (2011). Observation of long-lived excited states in DNA oligonucleotides with significant base sequence disorder. *Journal of Physical Chemistry Letters*, *2*, 133.
- de Vries, M. S., & Hobza, P. (2007). Gas-phase spectroscopy of molecular building blocks. *Annual Review of Physical Chemistry*, *58*, 585.
- Dickerson, R. E., & Drew, H. R. (1981). Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *Journal of Molecular Biology*, *15*, 761.
- Dong, F., & Miller, R. E. (2002). Vibrational transition moment angles in isolated biomolecules: A structural tool. *Science*, *298*, 1227.
- Elstner, M., Hobza, P., Frauenheim, T., Suhai, S., & Eftimios, K. (2001). Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *Journal of Chemical Physics*, *114*, 5149.
- Ferus, M., Nesvomy, D., Sponer, J., Kubelik, P., Michalcíková, R., Shestivska, V., Sponer, J. E., & Civis, S. (2015). High-energy chemistry of formamide: A unified mechanism of nucleobase formation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 657.
- Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, *391*, 806.
- Gehring, K., Leroy, J.-L., & Gueron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, *363*, 561.
- Glaser, R., Hodgen, B., Farrelly, D., & Mckee, E. (2007). Adenine synthesis in interstellar space: Mechanisms of prebiotic pyrimidine-ring formation of monocyclic HCN-pentamers. *Astrobiology*, *7*, 455.
- Gorb, L., Kaczmarek, A., Gorb, A., Sadlej, A. J., & Leszczynski, J. (2005). Thermodynamics and kinetics of intramolecular proton transfer in guanine. Post Hartree-Fock study. *Journal of Physical Chemistry B*, *109*, 13770.
- Grimme, S. (2004). Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, *25*, 1463.
- Gu, J., Wang, J., & Leszczynski, J. (2006). Electron attachment-induced DNA single strand breaks: C₃' - O₃' σ -bond breaking of pyrimidine nucleotides predominates. *Journal of American Chemical Society*, *128*, 9322.
- Gu, J., Wang, J., Rak, J., & Leszczynski, J. (2007a). Findings on the electron-attachment-induced abasic site in a DNA double helix. *Angewandte Chemie*, *119*, 3549.
- Gu, J., Xie, Y., & Schaefer, H. F. (2007b). Electron attachment to DNA single strands: Gas phase and aqueous solution. *Nucleic Acid Research*, *2007*(35), 5165.

- Gu, J., Wang, J., Leszczynski, J., Xie, Y., & Schaefer, H., III. (2008). To stack or not to stack: Performance of a new density functional for the uracil and thymine dimers. *Chemical Physics Letters*, 2008(459), 164.
- Gu, J., Xie, Y., & Schaefer, H. F. (2009). Electron attachment to oligonucleotide dimmers in water: Microsolvation-assisted base-stacking forms. *Chemical Physics Letters*, 473, 213.
- Gu, J., Wang, J., & Leszczynski, J. (2010a). Electron attachment induced DNA single strand breaks at the pyrimidine sites. *Nucleic Acids Research*, 38, 5280.
- Gu, J., Wang, J., & Leszczynski, J. (2010b). Comprehensive analysis of the DNA strand breaks at the guanosine site induced by low energy electron attachment. *ChemPhysChem*, 11, 175.
- Gu, J., Wong, N.-B., Xie, Y., & Schaefer, H. F. (2010c). Electron attachment to a hydrated DNA duplex: The dinucleoside phosphate deoxyguanytyl-3′/5′-deoxycytidine. *Chemistry – A European Journal*, 16, 13155.
- Gu, J., Xie, Y., & Schaefer, H. F. (2010d). Electron attachment to hydrated oligonucleotide dimers: Guanylyl-3′/5′-cytidine and cytidyl-3′/5′-guanosine. *Chemistry – A European Journal*, 16, 5089.
- Gu, J., Wang, J., & Leszczynski, J. (2011a). Stacking and H-bonding patterns of dGpdC and dGpdCpdG: Performance of the M05-2X and M06-2X minnesota density functionals for the single strand DNA. *Chemical Physics Letters*, 512, 108.
- Gu, J., Wang, J., & Leszczynski, J. (2011b). Low energy electron attachment to the adenosine site of DNA. *Journal of Physical Chemistry. B*, 115, 14831.
- Gu, J., Wang, J., & Leszczynski, J. (2012a). Electron attachment to the cytosine-centered DNA single strands: Does base stacking matter? *The Journal of Physical Chemistry. B*, 116, 1458.
- Gu, J., Liang, G., Xie, Y., & Schaefer, H. F. (2012b). Electron attachment to solvated dGpdG: Effects of stacking on base-centered and phosphate-centered valence-bound radical anions. *Chemistry – A European Journal*, 18, 5232.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35, 849.
- Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W., & Olsen, J. (1999). Basis-set convergence of the energy in molecular Hartree-Fock calculations. *Chemical Physics Letters*, 302, 437.
- Heßelmann, A., Jansen, G., & Schütz, M. (2005). Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *Journal of Chemical Physics*, 122, 014103.
- Hobza, P., & Šponer, J. (1998). Significant structural deformation of nucleic acid bases in stacked base pairs: An ab initio study beyond Hartree-Fock. *Chemical Physics Letters*, 288, 7.
- Hobza, P., & Šponer, J. (1999). Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations. *Chemical Reviews*, 99, 3247.
- Hobza, P., Šponer, J., & Reschel, T. (1995). Density-functional theory and molecular clusters. *Journal of Computational Chemistry*, 16, 1315.
- Hudson, J. S., Eberle, J. F., Vachhani, R. H., Rogers, L. C., Wade, J. H., Krishnamurphy, R., & Springsteen, G. (2012). A Unified mechanism for abiotic adenine and purine synthesis in formamide. *Angewandte Chemie, International Edition*, 51, 5134.
- Ishii, K., Tajima, A., Taketsugu, T., & Yamashita, K. (2006). Theoretical elucidation of the unusually high [HNC]/[HCN] abundance ratio in interstellar space: Two-dimensional and two-state quantum wave packet dynamics study on the branching ratio of the dissociative recombination reaction $\text{HCNH}^+ + e^- \rightarrow \text{HNC/HCN} + \text{H}$. *The Astrophysical Journal*, 636, 927.
- Jeziorski, B., Moszynski, R., & Szalewicz, K. (1994). Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chemical Reviews*, 94, 1887.
- Jurecka, P., Cerny, J., Hobza, P., & Salahub, D. R. (2007). Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *Journal of Computational Chemistry*, 28, 555.
- Kabelác, M., & Hobza, P. (2001). At nonzero temperatures, stacked structures of methylated nucleic acid base pairs and microhydrated nonmethylated nucleic acid base pairs are favored

- over planar hydrogen-bonded structures: A molecular dynamics simulations study. *Chemistry – A European Journal*, 7, 2067.
- Kohler, B. (2010). Nonradiative decay mechanisms in DNA model systems. *Journal of Physical Chemistry Letters*, 1, 2047.
- Kopitz, H., Zivkovic, A., Engels, J. W., & Gohlke, H. (2008). Determinants of the unexpected stability of RNA fluorobenzene self pairs. *ChemBioChem*, 9, 2619.
- Kristyán, S., & Pulay, P. (1994). Can (semi)local density functional theory account for the London dispersion forces? *Chemical Physics Letters*, 229, 175.
- Kruger, K., Grabowski, P. J., Zaugg, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31, 147.
- Kumar, A., & Sevilla, M. D. (2007). Low-energy electron attachment to 5'-thymidine monophosphate: Modeling single strand breaks through dissociative electron attachment. *Journal of Physical Chemistry B*, 111, 5464.
- Leontis, N. B., Stombaugh, J., & Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research*, 30, 3497.
- Leszczynski, J. (1992). Are the amino groups in the nucleic acid bases coplanar with the molecular rings? Ab Initio HF/631G and MP2/631G Studies. *International Journal of Quantum Chemistry: Quantum Biology Symposium*, 19, 43.
- Leszczynski, J. (2000). Isolated, solvated, and complexed nucleic acid bases: Structures and properties. In M. Hargittai & I. Hargittai (Eds.), *Advances in molecular structure research* (Vol. 6, p. 209). Stamford: JAI Press.
- Lin, J., Yu, C., Peng, S., Akiyama, I., Li, K., Lee, L. K., & LeBreton, P. R. (1980). Ultraviolet photoelectron studies of the ground-state electronic structure and gas-phase tautomerism of purine and adenine. *Journal of the American Chemical Society*, 102, 4627.
- Marian, C. M. (2007). The guanine tautomer puzzle: Quantum chemical investigation of ground and excited states. *The Journal of Physical Chemistry. A*, 111, 1545.
- Matthews, C. N. (1997). Hydrogen cyanide polymers from the impact of Comet P/Shoemaker/Levy 9 on Jupiter. *Advances in Space Research*, 19, 1087.
- McAllister, M., Smyth, M., Gu, B., Tribello, G. A., & Kohanoff, J. (2015). Understanding the interaction between low-energy electrons and DNA nucleotides in aqueous solution. *Journal of Physical Chemistry Letters*, 6, 3091.
- McMullan, R. K., Benci, P., & Craven, B. M. (1980). The neutron crystal structure of 9-methyladenine at 126 K. *Acta Crystallographica. Section B*, 36, 1424.
- Middleton, C. T., de La Harpe, K., Su, C., Law, Y. K., Crespo-Hernandez, C. E., & Kohler, B. (2009). DNA excited-state dynamics: From single bases to the double helix. *Annual Review of Physical Chemistry*, 60, 217.
- Miller, S. L., & Urey, H. C. (1959). Organic compound synthesis on the primitive earth. *Science*, 130, 245.
- Mix, L. J. (2006). The astrobiology primer: An outline of general knowledge – Version 1, 2006. *Astrobiology*, 6, 735.
- Mons, M., Dimicoli, I., Piuze, F., Tardivel, B., & Elhanine, M. (2002). Tautomerism of the DNA base guanine and its methylated derivatives as studied by gas-phase infrared and ultraviolet spectroscopy. *The Journal of Physical Chemistry. A*, 106, 5088.
- Mons, M., Piuze, F., Dimicoli, I., Gorb, L., & Leszczynski, J. (2006). Near-UV resonant two-photon ionization spectroscopy of gas phase guanine: Evidence for the observation of three rare tautomers. *The Journal of Physical Chemistry. A*, 110, 10921.
- Morgado, C. A., Jurecka, P., Svozil, D., Hobza, P., & Šponer, J. (2009). Balance of attraction and repulsion in nucleic-acid base stacking: CCSD(T)/complete-basis-set-limit calculations on uracil dimer and a comparison with the force-field description. *Journal of Chemical Theory and Computation*, 5, 1524.
- Nir, E., Janzen, C., Imhof, P., Kleinermanns, K., & de Vries, M. S. (2001). Guanine tautomerism revealed by UV–UV and IR–UV hole burning spectroscopy. *Journal of Chemical Physics*, 115, 4604.

- Ponnamperuma, C., Lemmon, R. M., Mariner, R., & Calvin, M. (1963). Formation of adenine by electron irradiation of methane, ammonia, and water. *Proceedings of the National Academy of Sciences of the United States of America*, *49*, 737.
- Rak, J., Chomicz, L., Wiczak, J., Westphal, K., Zdrowowicz, M., Wityk, P., Zyndul, M., Makurat, S., & Golon, L. (2015). Mechanisms of damage to DNA labeled with electrophilic nucleobases induced by ionizing or UV radiation. *The Journal of Physical Chemistry. B*, *119*, 8227.
- Réblová, K., Spackova, N., Štefl, R., Csaszar, K., Koa, J., Leontis, N. B., & Šponer, J. (2003). Non-Watson-Crick base pairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA Loop E. *Biophysical Journal*, *84*, 3564.
- Roy, D., Najafian, K., & von Ragué Schleyer, P. (2007). Chemical evolution: The mechanism of the formation of adenine under prebiotic conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 17272.
- Serrano-Andres, L., & Merchán, M. (2009). Are the five natural DNA/RNA base monomers a good choice from natural selection? *Journal of Photochemistry and Photobiology, C: Photochemistry Reviews*, *10*, 21.
- Sheina, G. G., Stepanian, S. G., Radchenko, E. D., & Blagoi, Y. P. (1987). IR spectra of guanine and hypoxanthine isolated molecules. *Journal of Molecular Structure*, *158*, 275.
- Shukla, M. K., & Leszczynski, J. (2007). Electronic spectra, excited state structures and interactions of nucleic acid bases and base assemblies. *Journal of Biomolecular Structure & Dynamics*, *25*, 93.
- Shukla, M. K., & Leszczynski, J. (2008). Computational study of UV-induced excitations of DNA fragments. In *Radiation induced molecular phenomena in nucleic acids* (Challenges and advances in computational chemistry and physics, Vol. 5). Dordrecht: Springer.
- Siegfried, N. A., Metzger, S. L., & Bevilacqua, P. C. (2007). Folding cooperativity in RNA and DNA is dependent on position in the helix. *Biochemistry*, *46*, 172.
- Smith, I. W. M., Talbi, D., & Herbst, E. (2001). The production of HCN dimer and more complex oligomers in dense interstellar clouds. *Astronomy & Astrophysics*, *369*, 611.
- Soliva, R., Laughon, C. A., Luque, F. J., & Orozco, M. (1998). Molecular dynamics simulations in aqueous solution of triple helices containing d(G.C.C) trios. *Journal of the American Chemical Society*, *120*, 11226.
- Solomon, T., Seitz, H., & Sturm, H. (2009). DNA damage by low-energy electron impact: Dependence on guanine content. *Journal of Physical Chemistry B*, *113*, 11557.
- Spackova, N., Berger, I., Egli, M., & Šponer, J. (1998). Molecular dynamics of hemiprotonated intercalated four-stranded i-DNA: Stable trajectories on a nanosecond scale. *Journal of the American Chemical Society*, *120*, 6147.
- Šponer, J., & Hobza, P. (1994). Bifurcated hydrogen bonds in DNA crystal structures. An ab initio quantum chemical study. *Journal of the American Chemical Society*, *116*, 709.
- Šponer, J., & Kypr, J. (1991). Different intrastrand and interstrand contributions to stacking account for roll variations at the alternating purine-pyrimidine sequences in A-DNA and A-RNA. *Journal of Molecular Biology*, *221*, 761.
- Šponer, J., & Kypr, J. (1993). Theoretical analysis of the base stacking in DNA: Choice of the force field and a comparison with the oligonucleotide crystal structures. *Journal of Biomolecular Structure & Dynamics*, *11*, 277.
- Šponer, J., Leszczynski, J., & Hobza, P. (1996a). Structures and energies of hydrogen-bonded DNA base pairs. A nonempirical study with inclusion of electron correlation. *The Journal of Physical Chemistry*, *100*, 1965.
- Šponer, J., Leszczynski, J., & Hobza, P. (1996b). On the nature of nucleic acid base stacking. Nonempirical ab initio and empirical potential characterization of 10 stacked base pairs. Comparison of stacked and H-bonded base pairs. *The Journal of Physical Chemistry*, *100*, 5590.
- Šponer, J., Leszczynski, J., Vetterl, V., & Hobza, P. (1996c). Base stacking and hydrogen bonding in protonated cytosine dimer: The role of molecular ion-dipole and induction interactions. *Journal of Biomolecular Structure & Dynamics*, *13*, 695.
- Šponer, J., Gabb, H. A., Leszczynski, J., & Hobza, P. (1997). Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA. A quantum-chemical study. *Biophysical Journal*, *73*, 76.

- Sponer, J., Mokdad, A., Sponer, J. E., Spackova, N., Leszczynski, J., & Leontis, N. B. (2003). Unique tertiary and neighbor interactions determine conservation patterns of Cis Watson-Crick A/G base-pairs. *Journal of Molecular Biology*, *330*, 967.
- Sponer, J., Jurecka, P., & Hobza, P. (2004). Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *Journal of the American Chemical Society*, *126*, 10142.
- Sponer, J. E., Leszczynski, J., Sychrovský, V., & Šponer, J. (2005a). The sugar edge/sugar edge base pairs in RNA. Stabilities and structures from quantum chemical calculations. *Journal of Physical Chemistry B*, *109*, 18680.
- Sponer, J. E., Spackova, N., Leszczynski, J., & Šponer, J. (2005b). Principles of RNA base pairing: Structures and energies of the trans Watson-Crick/sugar edge base pairs. *Journal of Physical Chemistry B*, *109*, 11399.
- Sponer, J., Jurecka, P., Marchan, I., Luque, F. J., Orozco, M., & Hobza, P. (2006). Nature of base stacking. Reference quantum chemical stacking energies in ten unique B-DNA base pair steps. *Chemistry – A European Journal*, *12*, 2854.
- Sponer, J. E., Réblová, K., Mokdad, A., Sychrovský, V., Leszczynski, J., & Šponer, J. (2007). Leading RNA tertiary interactions: Structures, energies, and water insertion of A-minor and P-interactions. A quantum chemical view. *Journal of Physical Chemistry B*, *111*, 9153. and references cited therein.
- Sponer, J., Riley, K. E., & Hobza, P. (2008). Nature and magnitude of aromatic stacking of nucleic acid bases. *Physical Chemistry Chemical Physics*, *10*, 2595.
- Sponer, J., Zgarbova, M., Jurecka, P., Riley, K. E., Sponer, J. E., & Hobza, P. (2009). Reference quantum chemical calculations on RNA base pairs directly involving the 2-OH group of ribose. *Journal of Chemical Theory and Computation*, *5*, 1166.
- Sponer, J., Sponer, J. E., Petrov, A. I., & Leontis, N. B. (2010). Quantum chemical studies of nucleic acids can we construct a bridge to the RNA structural biology and bioinformatics communities? *Journal of Physical Chemistry B*, *114*, 15723.
- Stombaugh, J., Zirbel, C. L., Westhof, E., & Leontis, N. B. (2009). Frequency and isostericity of RNA base pairs. *Nucleic Acids Research*, *37*, 2294.
- Suzuki, M., Amano, N., Kakinuma, J., & Tateno, M. (1997). Use of a 3D structure data base for understanding sequence-dependent conformational aspects of DNA. *Journal of Molecular Biology*, *274*, 421.
- Svozil, D., Hobza, P., & Sponer, J. (2010). Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations? *The Journal of Physical Chemistry. B*, *114*, 1191.
- Szalewicz, K., & Jeziorski, B. (1998). Comment on “on the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy”. *Journal of Chemical Physics*, *109*, 1198 [Journal of Chemical Physics, *104*, 8821 (1996)].
- Tennekes, P. P., Harju, J., Juvela, M., & Toth, L. V. (2006). HCN and HNC mapping of the protostellar core Chamaeleon-MMS1. *Astronomy & Astrophysics*, *456*, 1037.
- Vlieghe, D., Sponer, J., & Van Meervelt, L. (1999). Crystal structure of d(GGCCAATTGG) complexed with DAPI reveals novel binding mode. *Biochemistry*, *38*, 16443.
- Wang, J., Gu, J., Nguyen, M. T., Springsteen, G., & Leszczynski, J. (2013a). From formamide to purine: An energetically viable mechanistic reaction pathway. *The Journal of Physical Chemistry. B*, *117*, 2314.
- Wang, J., Gu, J., Nguyen, M. T., Springsteen, G., & Leszczynski, J. (2013b). From formamide to purine: A self-catalytic reaction pathway provides a feasible mechanism for the entire process. *The Journal of Physical Chemistry. B*, *117*, 9333.
- Wang, J., Gu, J., Nguyen, M. T., Springsteen, G. & Leszczynski, J. (2013c). From Formamide to Adenine: A Self-Catalytic Mechanism for an Abiotic Approach. *J. Phys. Chem. B*, *117*, 14039.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T., & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, *407*, 327.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature*, *287*, 755.

- Yamazaki, S., Domcke, W., & Sobolewski, A. (2008). Nonradiative decay mechanisms of the biologically relevant tautomer of guanine. *The Journal of Physical Chemistry. A*, *112*, 11965.
- Yanagi, K., Prive, G. G., & Dickerson, R. E. (1991). Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *Journal of Molecular Biology*, *217*, 201.
- Yildirim, I., & Turner, D. H. (2005). RNA challenges for computational chemists. *Biochemistry*, *44*, 13225.
- Yildirim, I., Stern, H. A., Sponer, J., Spackova, N., & Turner, D. H. (2009). Effects of restrained sampling space and nonplanar amino groups on free-energy predictions for RNA with imino and sheared tandem GA base Pairs. Flanked by GC, CG, iGiC or iCIG base pairs. *Journal of Chemical Theory and Computation*, *5*, 2088.
- Zhao, Y., & Truhlar, D. G. (2008). Density functionals with broad applicability in chemistry. *Accounts of Chemical Research*, *41*, 157.
- Zhou, J., Kostko, O., Nicolas, C., Tang, X., Belau, L., de Vries, M. S., & Ahmed, M. (2009). Experimental observation of guanine tautomers with VUV photoionization. *The Journal of Physical Chemistry. A*, *113*, 4829.
- Zirbel, C. L., Sponer, J. E., Sponer, J., Stombaugh, J., & Leontis, N. B. (2009). Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Research*, *37*, 4898.