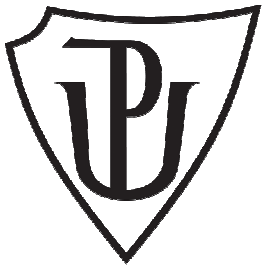
Palacký University Olomouc
Faculty of Science
2011



Ph.D. Thesis

Physical Chemistry

EMPIRICAL AND QUANTUM MECHANICAL MODELING OF BIOMOLECULES

MARIE ZGARBOVÁ

Department of Physical Chemistry

Supervisor: Doc. RNDr. Petr Jurečka, Ph.D.

Declaration of the author

I declare that I have worked out this thesis by myself using the cited references. Neither the thesis nor any of its part was previously used for obtaining any academic degree.

In Olomouc, 7th July 2011 ….. …..…….

Marie Zgarbová

Acknowledgement

It is a plesure to thank my supervisor Assoc. Prof. Petr Jurečka for his guidance and strong support. I am very grateful to Assoc. Prof. Michal Otyepka and Prof. Jiří Šponer for giving me the opportunity to collaborate with them.

My special thanks go to Prof. Javier Luque for his cooperation and valuable discussions during my stay at the University of Barcelona.

I would like to express my gratitude to all my colleagues at the Department of Physical Chemistry for friendly and inspiring working environment.

This thesis would not have been possible without the support of my friends and family.

TABLE OF CONTENTS

1 INTRODUCTION

Intermolecular (noncovalent) interactions are ubiquitous in the fields of physics, chemistry and biology. They determine the properties of gases, liquids, solids, chemical complexes, and biological compounds. Particularly in biology, weak noncovalent interactions play a leading role. They are essential for understanding of many biological processes in all living organisms, such as protein folding, drug binding, and supramolecular assembly.

In past decades, there has been an increasing interest in the field of molecular modeling of noncovalent interactions. We have witnessed an expansion of theoretical works that without any doubt provide a valuable complement to experimental studies. Study of these interactions by means of computational chemistry has already helped to answer some important questions, such as those regarding their nature and origin.

Nowadays, two major theoretical approaches – quantum mechanical and molecular mechanical methods – are widely used to study these interactions in biomolecules. Especially molecular mechanics is becoming very popular due to its ability to describe hundreds of thousands of atoms. In the past years, both techniques have seen rapid development and increasing applicability. Despite their success, evaluation and improvement of currently used methods is still very desirable.

Theoretical study of intermolecular interactions in nucleic acids is the main aim of this thesis. High-level quantum mechanical methods are used to investigate noncanonical interactions in RNA and assess a reliability of currently used force fields. The performance of the force fields is considered from two different points of view. First, an analysis of the individual interaction energy components is carried out with the aim to obtain deeper insight into the function of force fields. Second, the main factors that can influence final parameters in the force fields are evaluated and the acquired knowledge is used in construction of a new force field modification. Newly derived torsion parameters are tested and verified by extensive molecular dynamics simulations of both canonical and noncanonical RNA structures.

2 CLASSIFICATION OF INTERMOLECULAR INTERACTIONS

A number of physical phenomena contribute to attraction and repulsion between molecules or their parts. The most important intermolecular forces of attraction and repulsion can be classified into several types according to their origin.

There are four major types of interactions contributing to the overall stability: electrostatic, induction, dispersion, and repulsion. The electrostatic interactions are due to classical electrostatic interaction between the static charge distributions of the molecules or their parts. Electrostatic interaction may be either attractive or repulsive, depending on the orientation of interacting molecules. The induction forces are always attractive interactions that originate from the distortion of electron density of a particular molecule. The distortion is induced by the electric field produced by all its neighbors. The dispersion effects arise due to the fluctuation of the charge distributions of the molecules. In the region of short distances, the overlap of molecular electronic is significant and the repulsion contributions dominate [1-2].

3 NUCLEIC ACIDS

Deoxyribonucleic (DNA) and Ribonucleic (RNA) acids are essential biomolecules found in all living organisms. A basic building block of each nucleic acid is a monomeric unit called nucleotide. The nucleotides are linked covalently by phosphodiester bonds and form a polymeric structure of nucleic acids. Each nucleotide consists of three components: a heterocyclic nucleobase (purine or pyrimidine), a pentose sugar (ribose for RNA, deoxyribose for DNA), and a phosphate group. Four nuclobases can be found in DNA structure: the purines adenine (A) and guanine (G), and the pyrimidines cytosine (C) and thymine (T). In RNA the thymine is replaced by uracil (U). Nucleic acids differ in the structure of a 5-membered furanose ring – DNA contains 2’-deoxyribose, while RNA contains ribose (difference is only in the presence of the hydroxyl group). For overview, see, e.g., refs. [3-5].

3.1 Base Pairing

Interactions involving nucleic acids bases are of particular importance. RNA differs from DNA in several factors, such as presence of 2’-OH group, the different sugar pucker and the resulting different type of helices. Due to these factors, principles of RNA base pairing are strikingly different from those of DNA. Two qualitatively different interaction types involving nucleic acid bases are common in both DNA and RNA nucleic acids: hydrogen bonding and aromatic base stacking. Regarding the hydrogen bonding, while the DNA is dominated by standard Watson-Crick (WC) base pairs, RNA base pairing is more complex.

In the classic WC base-pairing scheme, cytosine pairs with guanine by forming three hydrogen bonds and thymine pairs with adenine by forming two hydrogen bonds. In RNAs structures, about 60% of bases participate in this canonical pattern (see ref. [6] and reference therein). In fact, each nucleobase may interact through any of the three edges [7-8]. These are Watson-Crick edge, the Hoogsteen (for purines) or "C-H" edge (for pyrimidines), and the sugar edge (SE) (Figure 1). Systematic and descriptive classification and nomenclature of the RNA base pairing patterns was done by Leontis and Westhof in 2002 [8]. In addition to interaction with any of three edges, orientation of nucleobases with respect to its sugar (glycosidic bond orientation) is considered. The classification thus distinguishes 12 families (Table 1) with 168 possible base pairings.

Non-WC (namely sugar edge) base pairs can interact with the 2'-OH group of ribose either as donors or as acceptors of the H-bond [8]. These sugar edge pairs are very important in building of the three-dimensional RNA architecture. They are, for instance involved in crucial RNA tertiary interactions, such as the A-minor motif [9-11], packing interactions [12-13], and in many interactions in the internal RNA loops [14-18].

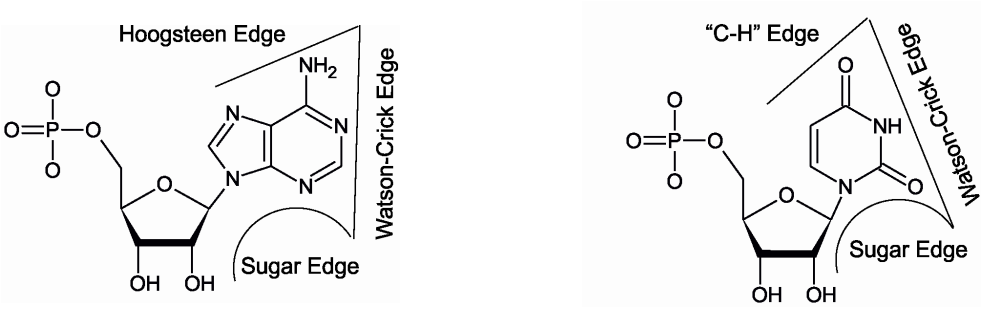


Figure 1: *Three edges for purines (left) and pyrimidines (right) involved in base pairing [6], adapted from ref. [19].*

Table 1: *The 12 families of edge-to-edge base pairs [6].*

No.	Glycosidic bond orientation	Interacting edge
1	*Cis*	Watson-Crick/Watson-Crick
2	*Trans*	Watson-Crick/Watson-Crick
3	*Cis*	Watson-Crick/Hoogsteen
4	*Trans*	Watson-Crick/Hoogsteen
5	*Cis*	Watson-Crick/Sugar Edge
6	*Trans*	Watson-Crick/Sugar Edge
7	*Cis*	Hoogsteen/Hoogsteen
8	*Trans*	Hoogsteen/Hoogsteen
9	*Cis*	Hoogsteen/Sugar Edge
10	*Trans*	Hoogsteen/Sugar Edge
11	*Cis*	Sugar Edge/Sugar Edge
12	*Trans*	Sugar Edge/Sugar Edge

3.2 Base – Phosphate Interactions

Another group of noncanonical interactions occurring mainly in RNA are those involving the phosphate group. Importance of the base – phosphate (BPh) interactions was clearly pointed out only recently by Zirbel *et al.* [20], who also introduced a systematic nomenclature of this relatively broad family of interactions. It was shown [20] that about 12% of nucleotides in the ribosomal crystal structures are involved in BPh interactions. They contribute to stabilization of various structural motifs in RNA, such as UUCG and GNRA hairpin tetraloops (see below) [21-23], or the sarcin-ricin internal loop [20, 24-25].

According to Zirbel classification, 10 different binding labels (labeled 0BPh-9BPh) are distinguished. Together with 4 nucleobases, 17 different BPh binding patterns have been suggested. Possible interactions for each nucleobase are shown in Figure 2.

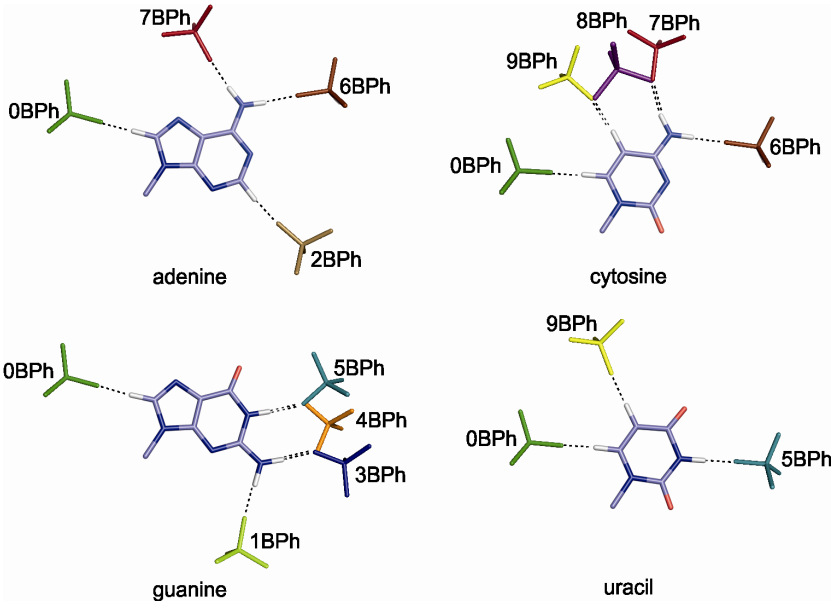


Figure 2: *Classification of the base-phoshate interactions [20].*

3.3 Nucleic Acids Conformation

Double helix is a dominant conformation of DNA [26] It commonly adopts one of the three biologically relevant forms A-DNA, B-DNA, and Z-DNA (Figure 3).Two main conformations of B-DNA form (known as BI and BII) are occured. However, DNA is also known in other forms, such as DNA triple and quadruple helices, junction structures, and parallel helices.

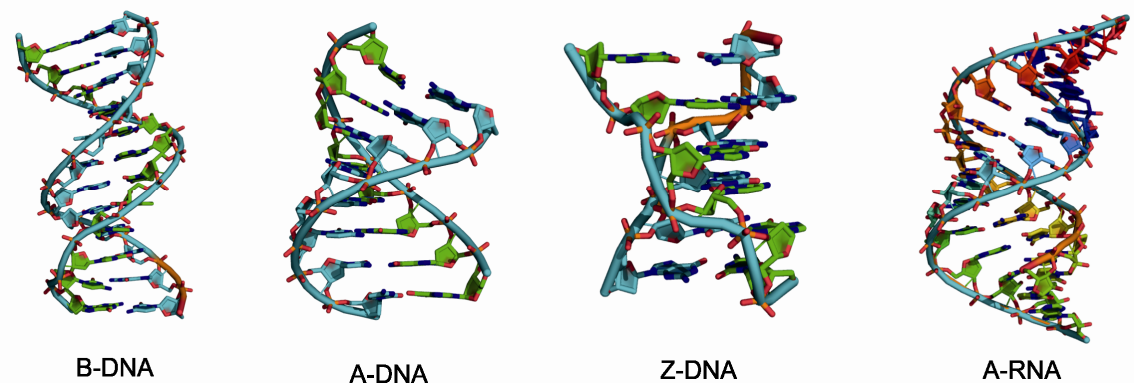


Figure 3: *Different types of nucleic acid double helix; B-DNA (dodecamer, pdb code 1BNA [27]), A-DNA (octamer, pdb code 1ANA [28]), Z-DNA (hexamer, pdb code 1DCG [29]), and A-RNA (tetradekamer, pdb code 1RNA [30]).*

When RNA forms a double stranded helix, it occurs in the double helices with structures similar to the A form of DNA (Figure 3). However, the higher level RNA structure is much more complex compared to DNA. The secondary structure of RNA involves canonically base-paired regions (helices) and non-paired regions (loops). Thus, common secondary structural elements are abovementioned duplexes, single-stranded regions, hairpins, bulges, internal loops, and junctions. Interactions between distinct secondary structural elements determine tertiary structure. Plenty of tertiary interactions in RNA are distinguished. Among these are coaxial helices, A-minor motif, tetraloop receptor, ribose zipper, kissing hairpins, pseudoknots, and tRNA D-loop:T:loop motif (see, e.g., refs. [31-32]). The A-minor motif (37%), coaxial helices (32%), and ribose zippers (20%) are the most frequent tertiary motifs (non-redundant set of 54 high-resolution crystal structures, see, ref. [32] for more details).

In one study (Appendix III) of this work, tetraloops were extensively studied by molecular dynamics simulations. Tetraloops are the most common and best-studied hairpin loops. Generally, they consist of a base paired stem region and a single-stranded loop region with independent sequence and structure. Loops of the GNRA and YNMG type (N stands for any nucleotide, R for purine, Y for pyrimidine, and M for adenine or cytosine) are the most frequent [33] and unusually stable thermodynamically compared to the others [34]. Tetraloops are known to participate in various biochemical processes, including nucleation in RNA folding [22, 35-36] and formation of tertiary contacts [37-39].

The conformational space of nucleic acids (and each nucleotide) is determined by six backbone torsion angles (α, β, χ, δ, ε, ζ) and a glycosidic angle χ (Table 2 and Figure 4).

Table 2: *Nucleic acid torsion angles.*

Angle	Sequence
α	O3'-P-O5'-C5'
β	P-O5'-C5'-C4'
γ	O5'-C5'-C4'-C3'
δ	C5'-C4'-C3'-O3'
ε	C4'-C3'-O3'-P
ζ	C3'-O3'-P-O5'
χ	O4'-C1'-N1-C2 (pyrimidine) O4'-C1'-N9-C4 (purine)

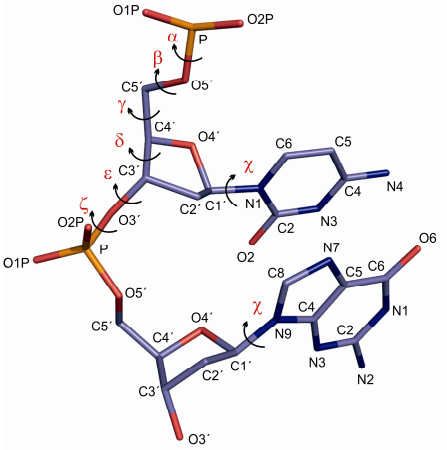


Figure 4: *Nucleic acid torsion angles.*

Each torsion angle can be assigned to a particular conformational region characteristic for different type of nucleic acids. Table 3 shows average value for each torsion angle in various nucleic acid types (B-DNA, A-DNA, Z-DNA and A-RNA).

Table 3: *The average values (°) of the backbone angles in A-DNA[40] (canonical A-form), B-DNA [40] (canonical B-form), Z-DNA [40], and A-RNA [41] in high-resolution crystals. (R stands for purines, Y stands for pyrimidines).*

	α	β	γ	δ	ε	ζ	χ
A-DNA	293	174	56	81	203	289	199
B-DNA	298	176	48	128	184	265	258^{R} 241^{Y}
Z-DNA	71^{R} 201^{Y}	183^{R} 225^{Y}	179^{R} 54^{Y}	$95^{R,C3'\text{-endo}}$ $148^{R,C2'\text{-endo}}$ 141^{Y}	240^{R} 267^{Y}	301^{R} 75^{Y}	$63^{R,C3'\text{-endo}}$ $76^{R,C2'\text{-endo}}$ 204^{Y}
A-RNA	295	173	54	80	210	287	199

The glysosidic bond determines the orientation of a nucleobase with respect to the sugar. Two major orientations, *syn* and *anti*, are found. They range in the intervals (0°, 90°) and (180°, 240°), respectively. Conformation with $\chi > 240°$ is often called high-*anti*. Sugar pucker (see below) correlates with χ values. RNA is typically found in the A form with the χ population peaking at around 200° (*anti*). The most frequent DNA form, B form, adopts χ values around 250° (high-*anti*).

A sugar ring in nucleic acids is non-planar and adopts various conformations characterized by two parameters: pseudorotation phase angle P, and puckering amplitude τ_m. Two major types of pucker are commonly observed in the crystal structures of nucleic acids. The South C2'-*endo* puckers have P values in the range 140° to 185°, and the North C3'-*endo* puckers in the range -10° to +40° [5]. The puckering amplitude averages around 40° ± 5° [4].

Conformation of nucleic acids can be further characterized by a various rotational and translational, both global and local, parameters. The global parameters depict the overall arrangement of the base pairs in double-stranded helices, while the local parameters describe the orientation between successive base pairs. Parameters are further classified according to isolated base pair or two pairs (step). [4-5] A summary of parameters is given in Table 4 [42] and in Figure 5 and 6 [43].

Table 4: *A summary of rotational and translational parameters [42].*

		local		global
		step parameters	base-pair parameters	
motion	axis	name	name	name
rotation	z	twist	opening	$twist_g$
	y	roll	propeller	tip
	x	tilt	buckle	inclination
translation	z	rise	stagger	$rise_g$
	y	slide	stretch	y-displacement
	x	shift	shear	x-displacement

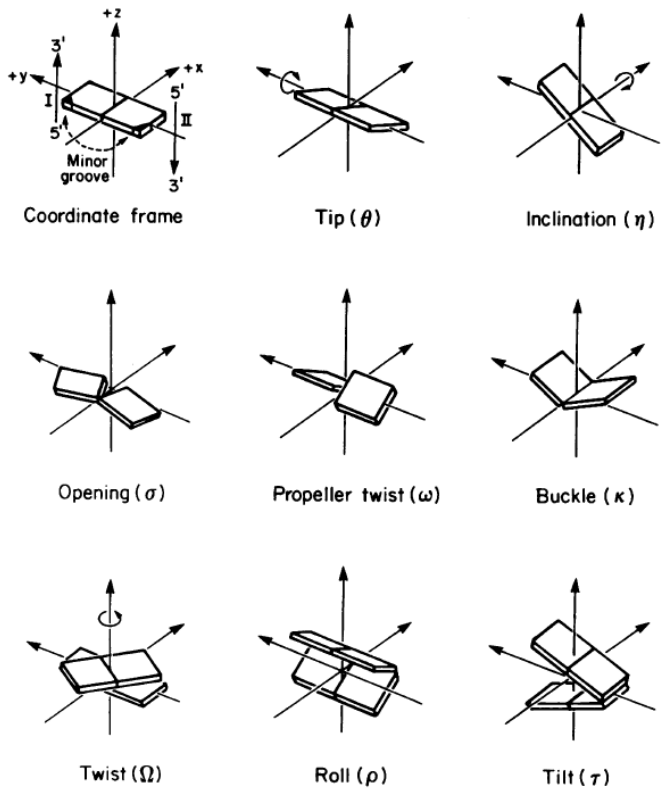


Figure 5: *Definitions of various rotations involving two bases of a pair (upper two rows) or two successive base pairs (bottom row) - adapted from ref. [43].*

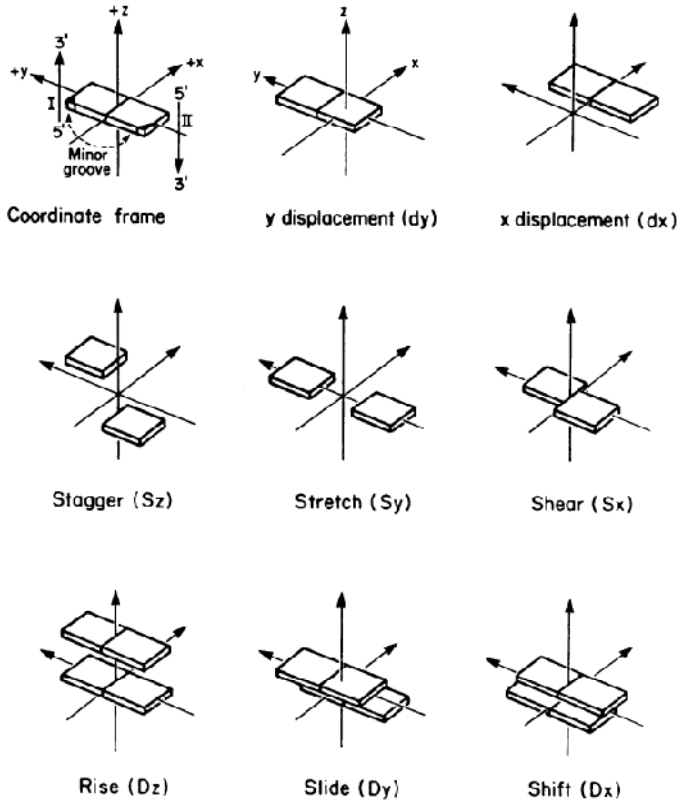


Figure 6: *Definitions of various translations involving two bases of a pair (upper two rows) or two successive base pairs (bottom row) – adapted from ref. [43].*

4 THEORETICAL METHODS

The next several paragraphs deal with theoretical methods used in this work. The special attention is paid to the quantum mechanical and molecular mechanical methods and approaches used to model intermolecular interactions.

4.1 Quantum Mechanical Methods

Quantum mechanical (QM) methods have become useful and important tools for investigation of chemical properties of atoms and molecules. Over the last several years, computer development and various approximations allowed use of QM methods for larger systems, including biomolecules. *Ab initio*[1] QM methods do not use any empirical parameters and are based on theory from the first principles. In general, the accuracy of *ab initio* calculations depends on the size of the basis set of atomic orbitals, which are applied to construct the molecular orbitals and on the inclusion of electron correlation effects. Two major groups – wave function theory (WFT) and density functional theory (DFT) – of QM methods are widely used.

4.1.1 Wave Function Based Methods

In the wave function based methods, the wave function, which is found by solving the time-independent Schrödinger equation (Equation 1), is used to describe the electronic system.

$$\hat{H}\psi = E\psi \qquad 1$$

In Equation 1, $\hat{H}$ is the Hamiltonian describing energy of the system, ψ is the wave function, and E is energy of the system.

However, the Schrödinger equation can be solved analytically only for one-electron systems. For any other system, implementation of some mathematical approximation and numerical solution of the equation is required.

The Hartree-Fock (HF) method (for overview, see, e.g., ref. [44]) is a basic method for approximate solution of the Schrödinger equation. It assumes that the N-body wave function of the system is approximated by only a single Slater determinant. The variational principle implies that the acquired energy is greater than or equal to the exact energy. Nevertheless, the HF approximation assumes that electron moves in the averaged field of the other electrons. Consequently, the HF method does not cover important part of electronic energy, the correlation energy (defined as a difference between the exact non-relativistic energy and HF energy calculated at the HF limit).

A number of methods, called post-HF, have been designed to estimate correlation energy, e.g., perturbation Møller-Plesset (MP), coupled clusters (CC), and configuration interaction (CI). For overview, see, e.g., refs. [44-45].

[1] The term "*ab inito*" comes from Latin meaning from the beginning.

Møller-Plesset Perturbation Theory

Møller-Plesset perturbation (MP) [46] theory is due to its affordable computational cost and often satisfactory performance one of the most frequently used WFT methods for calculations of noncovalent interactions. Estimate of the correlation energy is based on the perturbation theory. Second (MP2), third (MP3) and fourth (MP4) order Møller-Plesset perturbation theory are often used in quantum chemical calculations.

In one of our studies (Appendix I), the SCS(MI)-MP2 was used to estimate the total interaction energies of various RNA base pairs. SCS(MI)-MP2 is the molecular interaction optimized variant of the spin component scaled second order Møller-Plesset perturbation method (SCS-MP2) [47]. The SCS-MP2 simply modifies MP2 approach by separate scaling of the correlation energy contributions from the singlet and triplet states of electrons. Its molecular interaction variant was parameterized against a set of hydrogen bonded, dispersion, and mixed complexes (S22 training set) [48-49].

Coupled Cluster Theory

The coupled cluster theory (CC) is one of the most accurate approaches for estimation of the electron correlation. The idea of this method [50-51] stems from the exponential formulation of the wave operator and its expansion into clusters of excitation operators. The exponential form allows effective inclusion of effects of higher electronic excitations.

The first model of CC theory, known as CCD (couple cluster double) includes all double excitations. CCSD method includes all single and double excitations. The coupled cluster method with full inclusion of single, double and triple excitations is denoted as CCSDT method. However, more applicable is the CCSD(T) [52], which incorporates single and double excitations iteratively and triple excitations are included via additional perturbative treatment of triple excitations. Among all CC approximations, CCSD(T) is one of the most popular for intermolecular interactions due to its very good accuracy/computational cost ratio [53].

4.1.2 Density Functional Theory

In the Density Functional Theory (DFT), the energy of the system is determined from the electron density instead of the wave function. It is originally based on the theorem of Hohenberg and Kohn [54], and later developed by Kohn and Sham [55]. Currently it is widely applied due to its relatively low computational cost and good accuracy comparable with other *ab inito* QM methods (see, e.g., ref. [56]) In comparison with the WFT methods, the DFT depends less on the basis set size.

The DFT energy is determined with the use of a density functional. Unfortunately, the exact form of the density functional is not known. Up until now, various classes of density functionals have been proposed. The simplest one - Local Density Approximation (LDA) - depends only on the electron density and is no longer used in treatment of intermolecular interactions. A Generalized Gradient Approximation (GGA) functionals (such as BLYP [57-58]) consider also the electron density gradient. The

next step in the improvement of DFT performance was introduction of the hybrid functionals. These functionals (e.g., B3LYP [59-60], PBE0 [61]) contain some portion of the explicit HF exchange and are nowadays the most popular among of all functionals. Another group of DFT functionals – meta-GGA (e.g., TPSS [62]) functionals - utilize also the (non-interacting) kinetic energy density [63].

Due to the approximations made in the density functionals none of the GGA, hybrid GGA, or meta-GGA functionals are able to describe dispersion energy. Numerous efforts were taken to include dispersion interaction into DFT. For instance, Zhao and Truhlar attempted to reparameterize the existing density functionals. This resulted in functionals performing very well for various types of complexes [64-66]. Regarding the noncovalent complexes, the M06-2X functional was recommended.

In addition, standard DFT can be augmented with an empirical dispersion term. Thus, the total energy is a sum of DFT energy and the empirical dispersion energy. The dispersion energy is represented by the C_6/r^6 formula and is calculated separately, see, e.g., refs. [67] or [68]. The added dispersion term is damped by an empirical damping function that reduces the dispersion attraction at intermediate and short distances and corrects for the overlap effects. The DFT-D (DFT augmented with empirical dispersion) method is capable of providing very accurate results for noncovalent interactions (see, e.g., [67] and [68]). The great advantage of these calculations is their computational economy. Thus, this method is applicable for medium and even large noncovalent complexes, with reliable results also for dispersion bound complexes [67].

Furthermore, one of the GGA functionals (B97) [69] has been reparameterized with inclusion of dispersion correction in the parameterization process. This Grimme's effort led to improved accuracy mainly for hydrogen-bonded complexes [70].

4.1.3 Interaction Energy

In general, two basic approaches are used to calculate interaction energy, supermolecular method and perturbation method. According to the supermolecular approach, the interaction energy (ΔE) is determined as a difference between energy of a complex (E) and energies of the isolated subsystems (e.g., for two subsystems A and B, $\Delta E=E_{AB}-E_A-E_B$). A major drawback of the supermolecular technique is the basis set inconsistency, which leads to the basis set superposition error (BSSE). The BSSE is a mathematical artifact arising due to different description of the supersystem and the subsystems. Elimination of BSSE is possible through the counterpoise (CP) method where the subsystems are treated in the basis set of the whole system. This method was independently proposed by Jansen and Ross [71], and Boys and Bernardi [72].

In contrast, the perturbation method calculates the intermolecular interactions directly and is free from BSSE. The symmetry-adapted perturbation theory (SAPT) is the most important of the perturbation techniques nowadays.

Symmetry Adapted Perturbation Theory

The symmetry-adapted perturbation theory (SAPT) [73] allows for deeper understanding of the intermolecular interactions. This approach is designed to decompose the intermolecular energy into its physically meaningful components. Originally, SAPT was formulated as a double-perturbative series, in which the intermolecular and intramolecular perturbations were treated separately. However, the large computational demand was a great disadvantage of this first approach.

In 2002, a DFT-based SAPT method was developed [74-81]. This new approach allowed for investigation of the extended molecular complexes. DFT-SAPT treats the considered monomers at the DFT level. The DFT-SAPT total interaction energy is usually determined (up to the second order) as a sum of the polarization (electrostatics) (E_{pol}^{1}), exchange repulsion (E_{ex}^{1}), induction (E_{ind}^{2}), dispersion (E_{disp}^{2}), and δ(HF) terms; the induction and dispersion components also have their exchange counterparts (E_{ex-ind}^{2} and $E_{ex-disp}^{2}$) (see Equation 2)

$$E_{int} = E_{pol}^{1} + E_{ex}^{1} + E_{ind}^{2} + E_{ex-ind}^{2} + E_{disp}^{2} + E_{ex-disp}^{2} + \delta(\mathrm{HF}) \qquad 2$$

The δ(HF) term represents the estimated higher-order Hartree-Fock (HF) contributions and is determined as a difference between the variational HF interaction energy and the sum of electrostatic, exchange-repulsion, induction, and exchange-induction energies (up to the second perturbation order). For more details, see, e.g., refs. [75, 81].

The original SAPT provides accurate results comparable with the reference QM methods, such as CCSD(T) [73]. Quality of the DFT-SAPT partly depends on quality of the density functional used. However, it provides reliable total interaction energies, which are usually comparable with the pure SAPT [75, 81].

Highly Accurate Interaction Energies

It has been realized that results close to chemical accuracy can be achieved by performing the coupled-cluster calculations covering the single and double electron excitations iteratively and the triple electron excitations perturbatively (CCSD(T)) in combination with the complete basis set (CBS) extrapolation (see, e.g., ref. [53] and references therein). These calculations provide results that can be used for testing, validation or parameterization of the lower-level *ab initio* QM wave function and density functional methods as well as semiempirical and empirical computational methods.

Estimate of the MP2/CBS energy is possible via extrapolation towards the basis set limit. Approach of Helgaker and co-workers [82-83] (Equation 3 and Equation 4) is usually used for the extrapolation of HF and MP2 calculations to the CBS limit.

$$E_X^{HF} = E_{CBS}^{HF} + Ae^{-\alpha X} \qquad 3$$

$$E_X^{MP2} = E_{CBS}^{MP2} + BX^{-3} \qquad 4$$

E_X and E_{CBS} are energies for the basis set with the largest angular momentum X and for the complete basis set, respectively, and α is a parameter fitted by the author [82-83].

Furthermore, density fitting/resolution-of-the-identity approximation [84] for two-electron integrals is frequently used for MP2 calculations. It allows to use of larger basis sets for MP2 thanks to the lower computational cost of this approximation [85]. The ΔCCSD(T) correction term is determined as a difference between the CCSD(T) and MP2 interaction energies calculated with a small basis set (e.g., 6-31G*). The complete CCSD(T)/CBS energy is constructed as described in Equation 5 and Equation 6.

$$\Delta E^{CCSD(T)} \approx \Delta E^{HF}_{CBS} + \Delta E^{MP2}_{CBS} + \Delta CCSD(T) \quad (5)$$

$$\Delta CCSD(T) = \left(\Delta E^{CCSD(T)} - \Delta E^{MP2}\right)|_{small\ basis} \quad (6)$$

4.2 Molecular Mechanics

Molecular mechanics (MM) methods, also force field based or empirical methods represent a powerful tool for investigation of large biological molecules. They offer unique atomic view that may be difficult to obtain from experiment, thus they can complement the picture based on experimental data. For overview, see, e.g., refs. [86-88].

The force field consists of the potential energy function and parameter set necessary for this function. The potential energy function is used to describe the relationship of the structure to the energy of the system. Potential energy function of most of the current force fields is a function of pairs of atoms; this characteristic is called as two-body addidivity. Actually, many body effects can be implicitly included in the parameterization.

One of the most widely used forms of potential energy function is described by Equation 7. It is a sum of bonded (bond stretching, angle bending, dihedral term) and nonbonded contributions (long-range electrostatics and van der Waals forces) to the total energy. For demonstration of the individual contributions, see Figure 7.

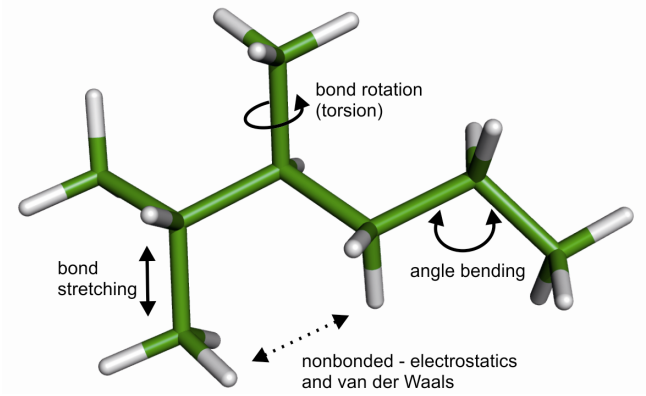


Figure 7: *Individual terms of the MM total energy.*

$$E = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2) + \sum_{dihedrals} \frac{V_n}{2}(1 + cos\,(n\varphi - \delta) + \sum_{nonbonded} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_0 r_{ij}} \right\} \quad (7)$$

In Equation 7, b is the bond length; θ is the valence angle; φ is the dihedral angle; r_{ij} is the distance between atoms i and j. Parameters include: the bond force constant and equilibrium distance, k_b and b_0, respectively; the valence angle force constant and equilibrium angle, k_θ and θ_0, respectively; the dihedral force constant, multiplicity and phase angle, V_n, n, and δ, respectively. In addition, nonbonded parameters are included: partial atomic charge, q_i; Lennard-Jones (LJ) well-depth, ε_{ij}; distance at which the potential reaches its minimum, R_{min}. ε_0 is the dielectric constant of vacuum.

Equation 7 is common for currently used biomolecular force fields, such as CHARMM [89], Cornell *et al.* (AMBER) [90], GROMOS [91], or OPLS [92]. Additional or alternate terms for bonded and nonbonded interactions are used in some force fields in order to increase their accuracy and reliability. This includes, for instance, higher order terms to treat the bond and valence angle terms, cross terms [93-98], Morse function for bonds [99-100]. Other terms for nonbonded interactions comprise the alternatives to the LJ 6-12 potentials. For instance, the Buckingham potential [101] uses an exponential term to treat repulsion, and, 14-7 term is used in the MMFF (Merck Molecular Force Field) force field [102].

Regarding the electrostatic term, majority of the biomolecular force fields treat this interactions using Coulombic interaction of point charges. The polarization is included implicitly by choosing the partial atomic charges that overestimate molecular dipoles [103]. Force fields that include the polarization explicitly are referred to as polarizable or nonadditive force fields [104]. Polarization term is, for instance, determined via Equation 8.

$$U_{pol} = \frac{1}{2}\sum_{\mathrm{i}} \mu_i E_i \qquad 8$$

where μ_i is the induced dipole moment of atom i and E_i is the electrostatic field at the position of atom i. The dipole moment is $\mu_i = \alpha_i E_i$, where α_i is polarizability of atom i.

Force fields are usually parameterized to reproduce experimental data. The empirical parameters are obtained directly from experimental results (NMR, X-RAY, spectroscopic techniques) or, alternatively, from quantum chemical calculations of suitable model compounds (backbone angles, dihedral parameters).

In this work, the AMBER package (Cornell *et al.* force field) was preferentially used. In this force field, parameters are fit to reproduce geometries, vibrational spectra and conformational energies of small molecules. The parameters are then applied directly to larger model compounds. Partial atomic charges are also acquired using *ab initio* methods and fitted using restrained electrostatic potential (RESP) [105] fitting procedure.

4.2.1 Force Field for Nucleic Acids

The Cornell *et al.* and CHARMM force fields are probably the most prevalent and extensively tested force fields for nucleic acids. The AMBER force field family (established for nucleic acids simulations) is based on Cornell *et al.* ff94 [90] and its subsequent modifications ff98 [106], ff99 [105], and parmbsc0 [107]. The CHARMM force field was designed by Mackerell *et al.* and includes two main versions - CHARMM22 [108-109], and CHARMM27 [110-111].

The improvement and development of the empirical force fields resulted in the past several years in more reliable description of the structure, energetics, and dynamics of nucleic acids. In the course of force field development, several shortcomings were described and reported. For instance, in 1997 and 1998, Feig and Pettitt pointed to the improper treatment of the equilibrium between the A and B forms of DNA [112-113].

The deviations of helicoidal parameters from canonical B values have been shown by Cheatham and Kollman in 1996 [114]. Problems with sugar puckering and helical repeat were observed in the ff94 of AMBER force field [106].

One of the crucial components of any empirical force field is the torsion space parameterization. Importance of the torsion parameters stems from the fact that they contribute to relative stability of the various biomolecular conformers (or folds), and the kinetics of the transitions between those conformers. Thus, in the realm of nucleic acids, torsion parameters influence for instance relative stabilities of the A/BI/BII/Z forms of DNA, and stability of all of the numerous motifs found in RNA structures, such as loops, bulges, etc. Many of these conformational equilibria are functionally relevant and are involved in many biological processes. For instance, the BI/BII equilibrium was suggested to play a key role in the recognition of DNA by proteins [115].

Current empirical torsion parameters provide generally reasonably good description of the most common nucleic acid forms, such as the frequently studied double helical DNA (canonical form). Good performance for these common structures is partially because they were considered in the process of force field development. Unfortunately, when less common (noncanonical) structures are studied, current torsion parameters may turn out to be inadequate. Then, the torsion parameters should be revised. Most force fields show long history of the torsion space readjustments. For instance, in the force field family based on Cornell *et al.* parameterization [116], the torsions connected with sugar pucker [106], *c* [106, 117–118] (see next paragraph for more detailed description) or α/γ [107] were revised in the past few years, some of them several times.

The glycosidic torsion χ experienced probably more refinement than any other torsion. In the AMBER ff98 [106] force field, the *c* torsion was modified together with sugar pucker, followed by minor readjustments of the sugar pucker parameters in the AMBER ff99 [105] force field. In 2008 Ode *et al.* [117] focused on the χ reparameterization with the help of QM calculations *in vacuo*. The energy profile was obtained at the MP2/aug-cc-pVTZ//HF/6-31+G(d,p) level on small model compounds. Recently, Yildirim *et al.* [118] performed another *c* reparameterization based on QM profiles (MP2/6-31G(d)//HF/6-31G(d)) *in vacuo*. In ref. [118] it was shown that the latest version of χ parameters improve the *syn* vs *anti* balance in simulation of isolated RNA nucleosides.

Frequent readjustments seen in the torsion parameterizations indicate that the torsion parameter development is accompanied by intrinsic difficulties. Refinement of force fields (especially for nucleic acids) or theirs parts is still very desirable due to the continuing problems and inaccuracies in simulations of nucleic acids.

4.3 Solvation

Environment is a powerful factor influencing behavior of molecules. The necessity of including the environment effects on molecules has led to development of various techniques and theoretical models to treat solvation. Nowadays, two basic approaches – explicit and implicit – are widely used.

4.3.1 Explicit Solvation

Explicit solvation models are represented by individual water molecules. Every water model is described by its geometry and molecular mechanics parameters, such as atomic charges and Lennard-Jones parameters. Explicit water models are characterized by the number of points used to define the model (atoms plus dummy sites). Furthermore, they differ in their flexibility and in the inclusion of polarization effects. Up to now, various explicit water models have been proposed. Among them, the Jorgensen's TIP*n*P [119-121] and Berendsen's SPC/E [122] are the most prominent in biomolecular simulations.

Explicit solvation involves specific effects that in principle cannot be modeled by an implicit models. These effects arise, for instance, due to the specific geometry of solvent molecules and their spatial charge distribution.

4.3.2 Implicit Solvation

Implicit solvent models (for overview, see, e.g., refs. [123-126]) represent the solvent as a continuum. Interactions of the continuum solute can be classified in the terms of several contributions, such as electrostatic, cavitation, exchange repulsion, and dispersion. Among all these contributions, the electrostatic term is highly dominant. The electrostatic interaction of a solute with the solvent depends on the charge distribution and polarizability of the solute and on the ability of the solvent to be polarized, which is characterized by the dielectric constant (ε) of the solvent. In the process of continuum solvation, solute molecules are placed into a cavity in solvent. The charge distribution of the solute polarizes the dielectric medium and the medium generates screening charge on the cavity surface (or on the grid points in the Poisson-Boltzmann (PB) model). Obtaining screening charges involves solution of the Poisson's equation.

Many solvent models for treatment of electrostatic solvation energy have been suggested over the time. Most of them predict the electrostatic interaction using continuum theory based on the Poisson's equation. They can be utilized in both QM (e.g., self-consistent reaction field (SCRF) model) and MM (e.g., PB, Generalized Born (GB) model) calculations.

In continuum solvation, the nonpolar terms are determined separately. One way to determine them is through the solvent accessible surface area (SASA). This model involves determination of the surface of each atom that is in contact with solvent.

Self-Consistent Reaction Field (SCRF) Method

The SCRF method is an implicit solvent model used only in QM calculations. It involves the quantum mechanical self-consistent reaction field (SCRF) equations into a boundary element problem. A number of models adopting this approach have been proposed. Frequently applied are, e.g., polarizable continuum model (PCM) [127] and dielectric PCM in which the surrounding medium is modeled as a conductor (conductor-like screening model - COSMO) [128]. In the SCRF model, the charge distribution of molecule is allowed to relax (polarizable solute) and molecule is additionally polarized by solvent.

Poisson-Boltzmann Model (PB)

The PB model (for overview, see, e.g., ref. [129] and references therein) is one of the most common models to estimate electrostatic energy. In this approach, the Poisson-Boltzmann equation (PBE) is solved as a function of the solute's charge density, a spatially dependent dielectric coefficient, and ion concentration. The finite difference method in which the solute is mapped onto a three-dimensional grid (instead of using a cavity) is a widely used numerical method for solution of PBE. When PB is used in the force field calculations, the molecule is described by fixed-point charges (it is not polarized) and only a solvent is polarized.

Generalized Born (GB) Approximation

The GB [130] model is based on the approximate solution of the exact Poisson-Boltzmann equation for a single atom. It is often used in combination with the SASA method, which estimates the nonpolar solvation energy (GBSA model) [131]. Despite of its lower accuracy, the GBSA model is very popular in molecular dynamics simulation due to its low computational demands.

5 RESULTS AND DISCUSSION

This chapter gives a brief overview of the obtained results. It is divided into three main parts. In the first part - Reference Quantum Chemical Calculations of Interaction Energies in RNA – the intermolecular interactions by various QM methods are presented. All calculations were carried out on the models of noncanonical RNA interactions in order to estimate relative stability *in vacuo* as well as in solvent. Furthermore, this part focuses on the performance and comparability of various QM methods.

The second part - The Intermolecular Interaction Energy Components in QM and MM – focuses on the analysis of the individual empirical energy components with the goal to improve understanding of the empirical force fields.

The third part - Accurate Reference Torsion Profiles for use in Force Field Parameterization - is aimed at development of a new parameterization methodology based on the high-level QM data, on its applications and testing of its reliability. This part takes advantage of the results obtained in the previous two parts, namely the improved understanding of some shortcomings of the studied empirical force fields.

5.1 Reference Quantum Chemical Calculations of Interaction Energies in RNA

RNA structures are stabilized by a plethora of noncanonical interactions. Such interactions involve the nucleotide bases, sugar and phosphates groups. Many of the noncanonical interactions have already been systematically characterized with experimental, bioinformatics and theoretical approaches [7-12, 14-18, 20, 132-139]. All these interactions are absolutely essential for RNA structure and function (see sections 3.1 and 3.2 for some examples).

Molecular modeling of noncanonical intermolecular interactions in RNA is able to provide a useful supplement to experimental techniques and bioinformatics. Many physical properties and characteristics of these interactions could hardly be obtained through an experiment.

Several studies aimed at characterization of RNA base pair patterns and using advanced quantum chemical calculations have been already published. (see, e.g., refs. [140-148]). A range of methods can inform us on individual components of interaction, such as its polarity or hydrophobicity. Very accurate interaction energies (CCSD(T)/CBS) are useful as a benchmark for force field calibration. A fast and accurate DFT-D method, an empirically corrected DFT variant, is capable of describing of also the stacking contacts in RNA. Decomposition of the interaction energies into separate components can be achieved by the DFT-SAPT method.

Quantum chemical calculations were carried out for two different sets of geometries. The first set consists of a representative selection of 25 RNA base pairs utilizing the ribose moiety for base pairing (Appendix I) (here denoted as sugar-edge (SE) base pair interactions). This set also contains various substates in which some of the nucleobase

amino groups are in amino-acceptor positions with respect to the 2'OH group of ribose. The second set comprises 32 individual BPh (Appendix IV) structures selected from the crystal structures of ribosomal subunits 1j5e and 1s72 [134, 149]. For the model of BPh interaction see Figure 8 (right). Additionally, larger models of BPh interactions taken from molecular dynamics (MD) simulations of the sarcin-ricin loop domain, hairpin ribozyme and *glmS* riboswitch have been considered.

Various QM methods were used in these two studies. Most accurate data were obtained with the MP2/CBS method corrected for the higher electron correlation effects using the CCSD(T) method with a smaller basis set of atomic orbitals (MP2/CBS(T)). The DFT-SAPT method with the aug-cc-pVDZ basis set was used for further evaluation of the intermolecular interactions. The DFT-D [67-68] approach that includes the dispersion forces via a damped empirical correction was also used. In addition to gas phase energies, the solvent effects were calculated. For the sugar-edge base pair interactions, the energy was further estimated with the SCS(MI)-MP2 [47] method. For the BPh interactions, additional computations were carried out with an empirical force fields.

The following abbreviations for the base pairs are used: "t" and "c" stand for trans and cis, and "W". "H", and "S" stand for Watson Crick, Hoogsteen, and Sugar edges, respectively. When ribose is included in the calculations, it is denoted by a preceding "r". Thus tWS G/rC means trans-Watson Crick/Sugar Edge base pair where guanine interacts its Watson Crick edge with sugar edge of cytosine.

Total Interaction Energy

The total interaction energy is governed mainly by the number and type of common hydrogen bonds, their strength and the overall complementarity of the electrostatic potentials. The gas phase interaction energies of SE base pair interactions range from −10 to −31 kcal/mol at the MP2/CBS level while energies of individual BPh structures calculated at the same level range from −1 to −36 kcal/mol. In comparison with these results, interaction energies for the canonical base pairs G=C and A-U are −32 and −17 kcal/mol, respectively. It was shown that from the chosen set of the SE base pair interactions, the strongest one is tWS G/rC (Figure 8) (−30.6 kcal/mol), which contains two NH…O and one OH…O hydrogen bonds.

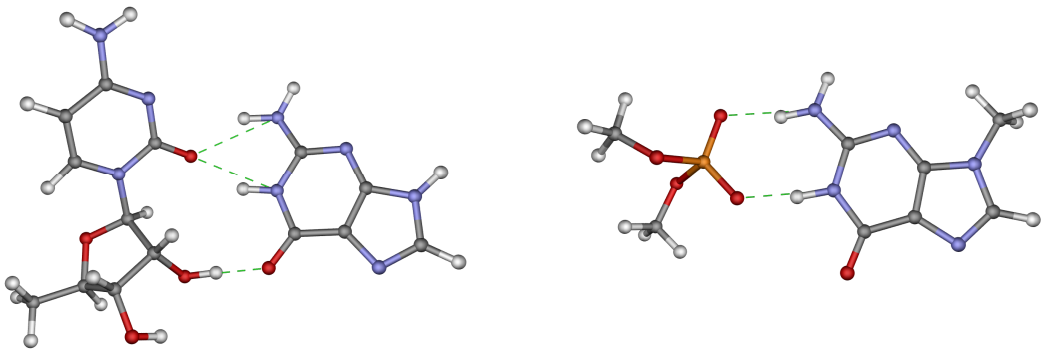


Figure 8: *Structures of the tWS G/rC (left) and G4BPh interactions (right).*

The BPh interactions represent a set of diverse H-bonds spanning from very weak contacts up to very strong hydrogen bond interactions. The strongest BPh interaction is G4BPh (noted already by Zirbel *et al.* [20]) (Figure 8) that includes two strong H-bonds between the nitrogen donor groups of the base and two distinct oxygen acceptors of the phosphate group.

DFT-SAPT Decomposition

In both studies, the DFT-SAPT method was used to decompose the total interaction energy into physically meaningful components. Noncanonical SE base pair interactions exhibit a weaker relative contribution of the electrostatics compared to the canonical base pairs (disregarding the alternative amino acceptor geometries). The induction component contributes between 16% for the tHS A/rA and 25% of the total attraction for the tWS U/rC. Many RNA base pairs exhibit a modestly increased role of dispersion attraction compared to canonical base pairs. The dispersion contribution to the sum of all attractive terms varies between 16% in the G=C base pair and 29% in the tHS A/rA. The damped empirical dispersion term (dispersion term calculated by DFT-D) agree fairly well with the DFT-SAPT here.

Regarding the BPh interactions, there are two types of interactions in the BPh families. Interactions involving polar hydrogens are very similar to the regular H-bond. They are mainly of electrostatic origin (60% of the total attractive energy). Induction and dispersion contribute significantly, by about 25 and 15%, respectively. The larger induction energy compared to the regular H-bonds is due to stronger electric field created by the phosphate group. The second type of interaction (weak CH…phosphate interaction), is much less electrostatic with dominant dispersion and induction contribution.

Comparison with Lower-Level Methods

Study of the SE base pair interactions shows good comparability of the SCS(MI)-MP2 method with the MP2/CBS results with an average error of 0.51 kcal/mol. Both studies demonstrate very good agreement of the DFT-D calculations with the TPSS [62] functional and 6-311++G(3df,3pd) basis set.

Comparison with Force Field Calculations

Force field calculations were performed only for the BPh interactions. Force field energies for BPh interactions are significantly underestimated. The difference compared to CBS(T) data is in the range of 2 – 10 kcal/mol.

Several reasons for this underestimation have been proposed. One reason is the lack of polarization, which is only partially compensated for by the HF-derived MM charges overestimating the static polarity of the interacting monomers. This affects basically all geometries. Another reason is overestimation of the short-range repulsion effects by the Lennard-Jones term of the force field. The r^{-12} repulsive term of the force field is too steep and this contributes to the QM vs. MM difference in cases where there is a really close separation of the monomers (see below). Finally, we used fixed-point charges taken from the Cornell *et al.* AMBER force field in our calculations as we primarily focused on assessment of the performance of the currently used force fields. The fixed-point charges might poorly represent variable structures taken from X-ray or MD simulations. Thus, the agreement between the QM and MM calculations could be probably improved if the charges were fit for each conformation separately.

Inclusion of Solvent Effects

Inclusion of the solvent effects into calculations is necessary to obtain estimates of the relative contributions of the SE base pairs and BPh interactions to RNA stability in water environment. The calculations indicate that the most stable complexes in the water are tWS G/rC and cWS C/rC from the SE base pair interactions. Then, calculations confirm G4BPh family as the most stable BPh interaction pattern in the water.

Importance of the Noncanonical Interactions for RNA Structure Stabilization

The representatives of the BPh interactions from three structures (hairpin ribozyme, *glmS* riboswitch, and sarcin-ricin domain) show that BPh interactions play a crucial role in the stabilization of the noncanonical parts of RNA structures. The best example is the sarcin-ricin motif, in which structure of the intricate S-turn submotif is most likely primarily determined by the BPh interactions.

5.2 Intermolecular Interaction Energy Components in QM and MM

Empirical force fields proved to be a powerful tool for studying of biomolecular interactions. They are usually used in connection with molecular dynamics simulations (MD) that allow for observation of the system's behavior in time. However, the reliability of the MD simulation strongly depends on the applied force field function and parameter set.

Current force fields are inspired by fundamental physical laws that determine the functional form of force field. Due to simplifications made, force fields are generally not universally applicable. Nevertheless, many successes have been already achieved thanks to the force fields and it is very desirable to improve the current methods. Such improvement cannot be achieved without more accurate reference data. With growing computational power, it is possible to use the higher level of QM methods and extended model compounds in the parameterization process. Also, deeper insight into the force field may be very helpful for construction or refinement of new force fields. Reference QM calculations can be used to reveal potential errors that are present in the force field.

Obtaining more detailed insight into the empirical methods was the main motivation for the study presented in the Appendix II. The intermolecular interaction energy components for several molecular complexes were calculated using empirical force fields and compared with the reference QM values. Published paper (Appendix II) provides results for seven model molecular complexes, namely acetate…methanol, acetate…methylammonium, methylammonium…methanol, methylamine…methanol, methanol dimer, hydrogen bonded (HB) uracil dimer, and stacked uracil dimer. Results obtained on the first five compounds have been already briefly discussed in my diploma thesis. Work was subsequently augmented by uracil dimers and published. Therefore, the following results and comments concern mainly the uracil dimers or results that are valid for all molecular complexes.

The work focuses on the intermolecular interactions (nonbonded terms) and the Cornell *et al.* force field implemented in the AMBER [150] package. Force field interaction energies were calculated using the Cornell *et al.* parameters with three different charge sets: the RESP HF/6-31G* vacuum charges used in many Cornell *et al.* [90] force fields variants (ff94, ff98, ff99, parmbsc0, etc) (ii) Duan *et al.* [151] RESP B3LYP/cc-pVTZ charges with continuum solvent (ε=4.0) as in ff03, and (iii) B3LYP/vac, RESP B3LYP/cc-pVTZ vacuum charges. Pure B3LYP/cc-pVTZ/vac charges used in this work are not corrected for self-polarization, so they are not same as those from the Cieplak *et al.* [152] polarizable force field ff02. The self-polarization is applied in order to adjust the charges for use with polarization term but the real vacuum electrostatics is better represented by the pure charges.

All QM calculations were performed with the DFT-SAPT method implemented in the Molpro [153] program package. The LPBE0AC [81] (a localized version of the PBE0 functional with asymptotic correction) functional in combination with the aug-cc-pVTZ basis set was used. The dispersion components were extrapolated using aug-cc-pVDZ and aug-cc-pVTZ basis sets to the complete basis set. Obtained QM results are of

very good quality and serve as the reference values for the force field comparisons [154-155].

Comparison of the force field interaction energy components with the corresponding DFT-SAPT values is not straightforward mainly due to an effective character of the force fields. The effectiveness of the potential originates mainly from the inter-component error cancellation and non-additivity of the force fields. Non-additivity stems from the contributions of many-body terms, which are not included in the most of the currently used pair-additive force fields. Another serious problem in such comparison arises at intermolecular distances where the electron clouds overlap. Since the electronic overlap effects are not taken into account in the currently used force fields, the distance-dependence of the individual interaction energy components was examined. The electron overlap effects are most important at the typical van der Waals (vdW) distances and at the very short distances. The positions of the monomers were varied with respect to the equilibrium distance (0.0 in the graphs below) by –0.5, –0.4, –0.3, –0.2, –0.1 (shortening), 0.0, 0.1, 0.3, 0.6, 1.0, 2.0, 5.0 and 10.0 Å (prolongation). Furthermore, ratios of the MM and QM energies as a function of the intermolecular distance are compared in addition to actual values.

Table 5 describes the DFT-SAPT and force field interaction energy components that are considered in this study.

Table 5: *DFT-SAPT quantum mechanical interaction energy components and force field terms.*

	Electrostatics	Induction	Dispersion[a)]	Repulsion[a)]
DFT-SAPT	E_{elst}^{SAPT} ($E_{elst}^{(1)}$)	E_{ind}^{SAPT} ($E_{ind}^{(2)} + E_{exch\text{-}ind}^{(2)} + \delta(HF)$)	E_{disp}^{SAPT} ($E_{disp}^{(2)} + E_{exch\text{-}disp}^{(2)}$)	E_{exch}^{SAPT} ($E_{exch}^{(1)}$)
Cornell *et al.*[b)]	E_{elst}^{MM} [b)]		E_{disp}^{MM} (B/r^6)	E_{rep}^{MM} (A/r^{12})
Duan *et al.*[c)]	E_{elst}^{MM} [c)]		E_{disp}^{MM} (B/r^6)	E_{rep}^{MM} (A/r^{12})
Cieplak *et al.*[d)]	E_{elst}^{MM} [d)]	E_{pol}^{MM}	E_{disp}^{MM} (B/r^6)	E_{rep}^{MM} (A/r^{12})

[a] All examined force fields use the same dispersion and repulsion parameters

[b] In Cornell *et al.* force field RESP HF/6-31G* charges are used

[c] In Duan *et al.* force field RESP B3LYP/cc-pVTZ/ε = 4.0 charges are used

[d] In Cieplak *et al.* force field RESP B3LYP/cc-pVTZ charges are used

Electrostatics

The electrostatic contribution is usually the most important part of the interaction energy, at least for the polar molecules. Thus, the accuracy requirement for electrostatic term is higher compared to other energy contributions. Furthermore, its accuracy cannot be compromised even at large distances because of its long-range character. The electrostatic energy calculated by the DFT-SAPT method is expected to be very accurate at both long and short ranges [78]. Empirical electrostatic term is calculated as a sum of the Coulomb interactions of the atomic charges. Its accuracy strongly depends on the quality of the point charges. The electrons overlap effect at typical vdW distances

is not covered by the point-charge model. Therefore, underestimation of the MM electrostatics attraction at short separations can be expected.

In general, at large separations the MM/QM ratio converges to a value not far from 1, while at a short range it drops far below 1 due to the missing overlap electrostatic stabilization. In some cases, the MM/QM ratio is significantly different from 1, e.g., the ratio for HB uracil dimer is 0.8 for Cieplak *et al.* force field (see Figure 9 that shows the results for both uracil dimers and Cieplak *et al.* force field). This is probably caused by inaccuracies of the RESP charges at long range.

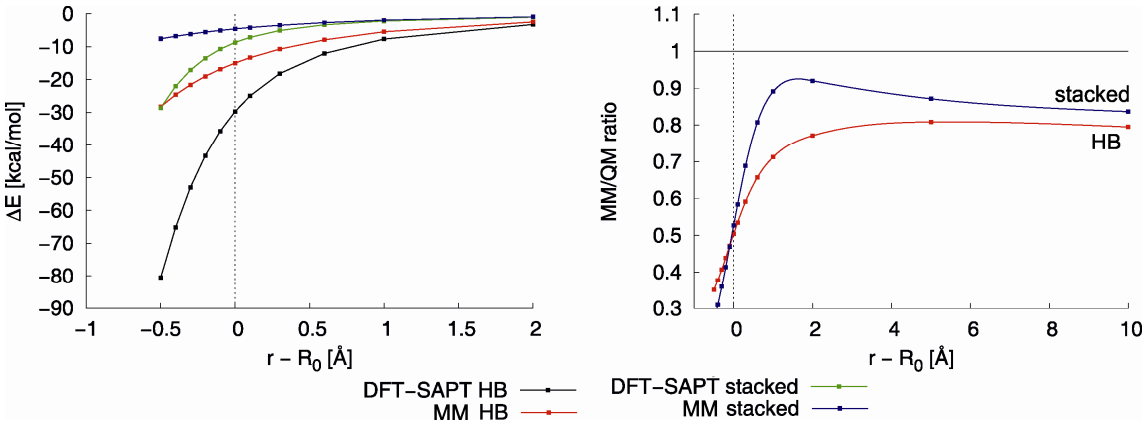


Figure 9: *Electrostatic component of the interaction energy for Cieplak et al. force field compared with DFT-SAPT reference energies (left) and the MM/QM ratio (right). Zero on the x axis corresponds to the equilibrium distance R_0. Data only for hydrogen bonded (HB) uracil dimer (R_0=6.075Å) and stacked structure (R_0=3.166Å) of uracil dimer are shown.*

For HB uracil dimer, the Cornell *et al.* and Duan *et al.* electrostatic is slightly overestimated at large intermolecular separations due to the overestimation of the molecule dipoles (results not shown, see Appendix II). This happens for all complexes where the dipole moments are oriented favorably. This result is consistent with already reported [151] distinction between the Cornell *et al.* electrostatics and the solvent-polarized electrostatics.

Taken together, the electrostatic component of current force fields represents the real electrostatic interactions reasonably well. The short range underestimation is not critical, because it can be compensated by the remaining force field parameters (see below).

Dispersion

The dispersion energy can be described by the asymptotic expansion (Equation 9). Coefficients $C_{6,8}$… are molecular dispersion coefficients representing the polarizability of the whole molecules.

$$E_{disp} = -C_6/r^6 - C_8/r^8 - C_{10}/r^{10}\ldots \qquad 9$$

If the polarizabilities vary with the directions, dispersion coefficients are anisotropic too. At short intermolecular distances, the dispersion becomes damped due to the electron overlap [73, 156].

Our results showed that for the small complexes MM/QM ratio is somewhat larger than 1 at large intermolecular distances. This is because the force field C_6 coefficient is overestimated on average. In general, the long-range force field dispersion agrees quite well with DFT-SAPT reference values, mainly for isotropic molecules. The force field dispersion however notably underestimated at typical vdW distances. This can be partly explained by the dispersion series truncation (the higher order contributions – C_8/r^8, C_{10}/r^{10}… – are omitted in force fields).

In case of uracil dimers, very large deviations from DFT-SAPT values have been observed (see Figure 10). Furthermore, deviations strikingly differ for the stacked and HB uracil dimer, despite they use the same C_6 coefficients. This can be explained only by the effect of anisotropy of the uracil polarizability and consequent anisotropy of its dispersion coefficient. Unfortunately, the effect of anisotropy introduces nonsystematic errors in force fields. Note that these errors cannot be corrected within the classical pair-additive force fields used today.

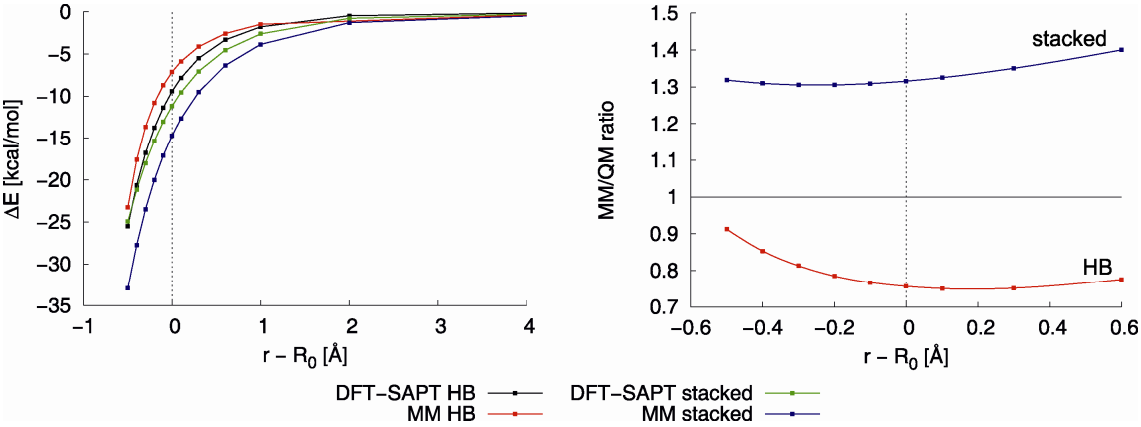


Figure 10: *Dispersion component of force field interaction energies (MM) compared with DFT-SAPT (QM) reference values (left) and the MM/QM ratio (right). Zero on the x axis corresponds to the equilibrium distance R_0. Interaction energies in kcal/mol. Data only for hydrogen bonded (HB) uracil dimer and stacked structure of uracil dimer are shown.*

Anisotropy of the C_6 coefficient of the nucleic acids is known for a long time. For instance, in cytosine dimers the C_6 component in the direction perpendicular to the ring plane is about three-fold smaller than in other directions [157]. Other nucleic acid basis exhibit similar anisotropic behavior. This means that in the force fields the in-plane dispersion interactions must be strongly underestimated while the interactions perpendicular to the base plane must be overestimated. Consequence for the description of the stacking interactions may deserve further study.

Apart from the observed systematic deviations, the overall force field dispersion agrees quite well with the DFT-SAPT reference values. It is proposed that force field dispersion parameters could be obtained directly from theoretical calculations in future parameterization efforts.

Repulsion

While the quantum mechanical repulsion could be best approximated by a simple exponential function, the force fields use the well known $1/r^{12}$ term, originating from the Lennard-Jones formula. Due to these different mathematical forms, the MM term overestimates repulsion at both long and very short distances. However, overestimation of the MM repulsion at large intermolecular distances is unimportant because the magnitude of repulsion is negligible there. At slightly more than the equilibrium distance and close to the vdW, the MM repulsion becomes underestimated (see Figure 11). Magnitude of this underestimation depends on the character of the molecular complex, and in some cases reaches as much as 80%. Just this underestimation plays an important role in the total error cancellation. However, note that the distance dependence of the repulsion error is a quite complicated function, which makes error cancellation in current force fields problematic and system-dependent. Replacing the current form with a simple exponential function might bring significant improvement in this respect.

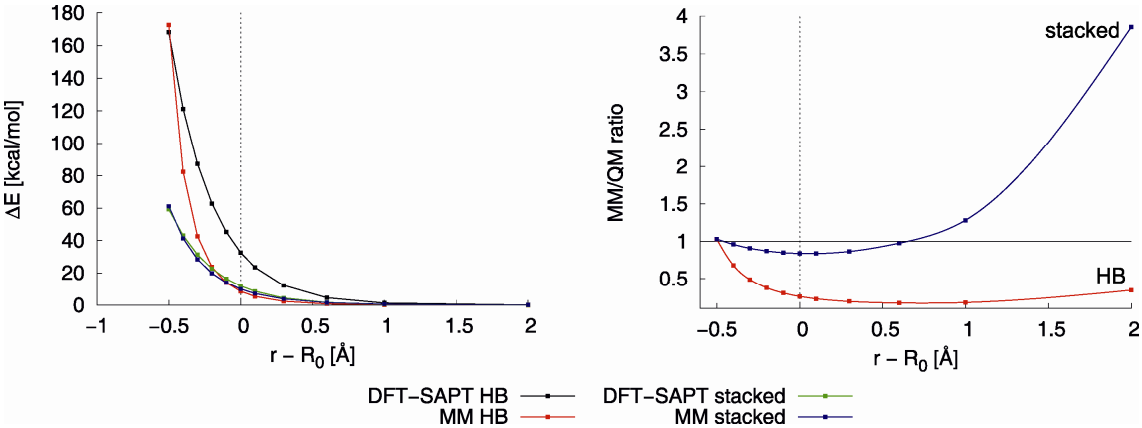


Figure 11: *Repulsion component of force field interaction energies (MM) compared with DFT-SAPT (QM) reference values (left) and the MM/QM ratio (right). Zero on the x axis corresponds to the equilibrium distance R_0. Interaction energies in kcal/mol. Data only for hydrogen bonded (HB) uracil dimer and stacked structure of uracil dimer are shown.*

Polarization

As in the previous cases, the DFT-SAPT induction (polarization) energies are fairly accurate and can serve as a reference value here [75-76, 79]. There are several ways how to include polarization explicitly in the force fields. In this work, the induced-dipole model as implemented in AMBER 9 [150] was considered. The polarization energy is calculated iteratively within the simple point dipole interaction model [158-159]. Each atom is assigned isotropic atomic polarizabity and an induced dipole can be generated on this atom by the surrounding permanent charges and by the surrounding induced dipoles.

Our results show that the polarization term in the Cieplak *et al.* force field is underestimated at both at large intermolecular distances and close to the vdW minimum

(see Figure 12). What is more, this underestimation is much larger for the uracil dimers than for the smaller complexes studied previously. One possible explanation is connected with the way the 1-4 electrostatic interactions are treated in the Cornell *et al.* force field (for details see Appendix II). We suggest that this fairly large inaccuracy may seriously limit quality of the results obtained with the Cieplak *et al.* polarizable force field. Note however, that the results obtained for ions are much more accurate (see Appendix II).

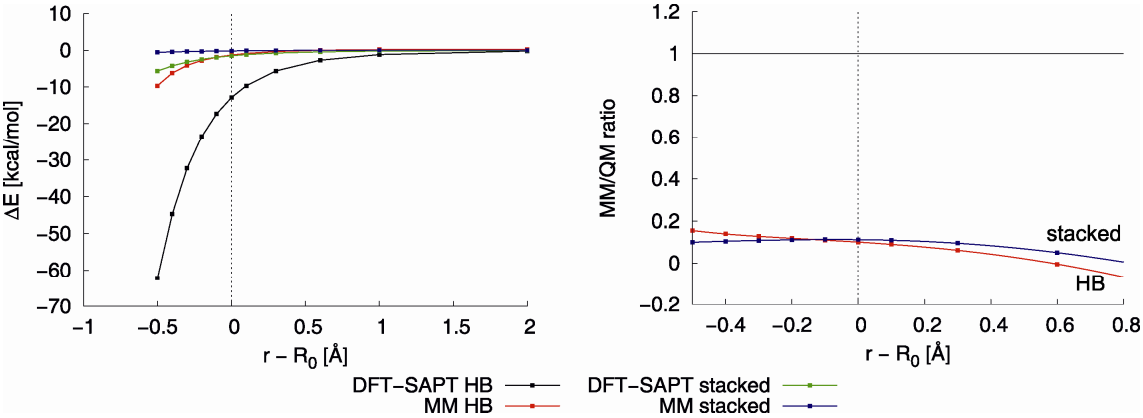


Figure 12: *Induction component of force field interaction energies (MM) compared with DFT-SAPT (QM) reference values (left) and the MM/QM ratio (right). Zero on the x axis corresponds to the equilibrium distance R_0. Interaction energies in kcal/mol. Data only for hydrogen bonded (HB) uracil dimer and stacked structure of uracil dimer are shown.*

Total Energy

In spite of fact that individual MM energy components exhibit rather large errors, total interaction energy shows reasonable agreement with the QM reference around equilibrium distance. Errors of individual interaction energy components tend to cancel out in a systematic manner. Our results showed a regular pattern of cancellation of the underestimated exchange repulsion and underestimated attractive terms (see Figure 13 that shows the absolute errors of the individual energy components for the stacked uracil dimer and the resulting total error as a function of the intermolecular distance). A systematic way of the error cancellation comes from the fact that all the errors exhibit close to exponential behavior.

Such systematic way of the error cancellation breaks down at very short distances. The empirical repulsion term $1/r^{12}$ becomes a bad approximation for the exponential growth of the repulsion in this region. The repulsion term thus becomes the main source of errors here.

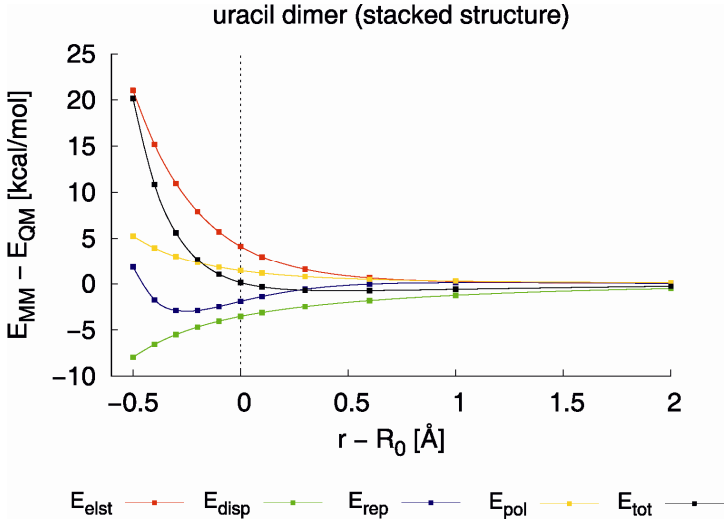


Figure 13: *Absolute error of the individual interaction energy components (kcal/mol) as a function of distance for Cieplak et al. force field and uracil dimer stacked complex.*

In general, the extent of cancellation is very large at vdW distances and the performace of current force fields depend on it to a high degree. The functional form of the AMBER force field is very similar to other widely used force fields. Thus, presented results are relevant for other commonly used force fields, such as GROMACS, CHARMM, AMOEBA, OPLS, *etc.*

5.3 Accurate Reference Torsion Profiles for use in Force Field Parameterization

Molecular dynamics (MD) simulations are currently a widely used approach for studies of structure and behavior of nucleic acids. The past 15 years witnessed significant advances in the development and availability of the empirical force fields for nucleic acids. Considering only the Cornell *et al.* force fields family, several adjustments have been proposed in order to improve the reliability of the MD simulations.

Parameterization of the nucleic acids is particularly complicated due to the flexible anionic sugar-phosphate backbone. The force field must simultaneously reproduce properties of canonical nucleic acid forms and numerous noncanonical topologies [87, 160-165]. An appropriate description of the torsion space seems to be particularly important.

Several papers have been published recently [166-168] showing serious problems with description of the noncanonical RNA structures. One of these force field artifacts was discovered and described during extensive MD simulations carried out in our laboratory and reported recently by Mlynsky *et al.* [168]. It was manifested by large structural distorsions in MD simulations of Hairpin Ribozyme. The A-type helix formed a distorted structure, which was named “ladder-like” conformation. The transition of helix to “ladder-like” conformation was irreversible at the time scale from tens to hundreds of nanoseconds and was characterized by a shift of the glycosidic angle χ from ~200° to ~270°, loss of helical twist, a change of the sugar pucker from C3’-endo to C2’-exo, and increases in slide and P-P distances in their radial distribution function. Previous MD study (Appendix III) has also shown that deformations of this type are actually common also in simulations of smaller RNA fragments and tetraloops (see Figure 14 for “ladder-like” structure of the GAGA tetraloop). It should be noted that appearance of the “ladder-like” artifact depends critically on the simulation timescale and it usually takes at least tens of ns to emerge (this is also the reason why the artifact would not have appeared in most previous RNA simulation studies).

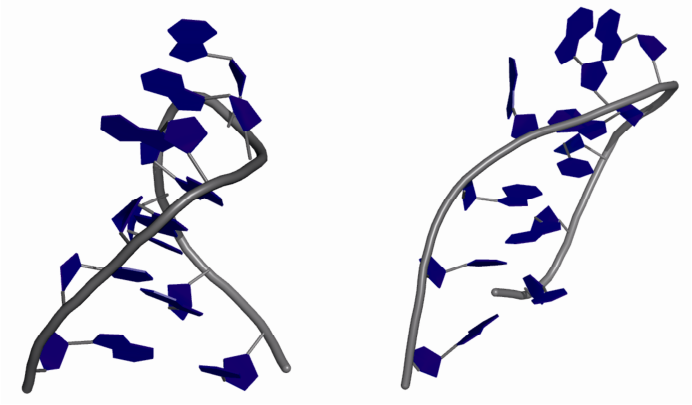


Figure14: *X-RAY (left) and “ladder-like”(right) conformer of the GAGA tetraloop.*

To correct the “ladder-like” artifact we decided to modify the current force field. In order to improve upon the currently used parameters we began developing a new parameterization methodology (Appendix V). The parameterization procedure is based on high-level quantum mechanical (QM) reference data obtained for relatively large model systems (in comparison with those used in the original force field derivation). A new protocol takes into account some previously neglected solvation-related effects and geometry relaxation effects. This approach was applied to the dihedral angle χ (which determines relative orientation of the sugar moiety and nucleobase), because we hypothesized that the χ torsion angle could be responsible for the formation of the “ladder-like” structures. In addition, we have evaluated errors arising from other commonly applied methodological assumptions. Parameterization methodology consists of several parts (see below). Accuracy of each of them influences to a great extent final parameters.

Choice of the Method for Geometry Optimization

DFT-based methods were considered for geometry optimizations, due to their advantageous balance between quality and speed. Several levels of theory were tested in order to determine the level of computations required. Among them, the combination of PBE functional with 6-311++G(3df,3pd) (LP) basis set (in COSMO implicit solvent) has been chosen as the most reliable for optimization in the force field parameterization procedure. PBE/LP level is known to provide the best results for polar contacts in molecules [67] which contribute to the shape of the potential energy surface for model compound used in this study (model of the whole nucleosides, see Figure 15 for cytosine).

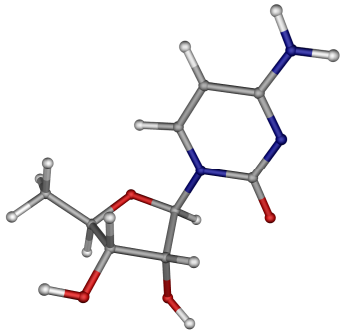


Figure 15: *Model of the ribonucleosides, used in the derivation of c torsion parameters for cytosine.*

Choice of the Method for Single Point Calculations

Determination of the sensitivity of the torsion profile to the level theory is very important because even very small changes in the torsion profile can cause substantial discrepancies in MD simulations. In this study, the MP2/CBS level with CCSD(T)/cc-pVDZ correction has been used as a reference (here denoted CBS(T)). Our tests have shown that MP2/CBS method is sufficiently accurate to serve as QM method for deriving new parameters. Differences between CBS(T) and MP2/CBS are only about

0.1 kcal/mol around the high-*anti* minima. In addition, MP2/CBS yields correct balance of the *anti* and high-*anti* regions (torsion angles 210° and 250°, respectively). Other tested methods were not able to reach these demands on accuracy (for the comparison of tested methods, see Figure 16). For instance, the MP2/6-31G* method predicts the modeled structure to be significantly less stable that does the reference CBS(T) method in the high-*anti* region. The MP2/6-31G* provides an incorrect balance of the *anti* and high-*anti* regions. The PBE-D-1.06-23/LP profile exhibits error around the energy barriers. The M06 and M06-2X DFT functionals overestimate the lower transition barrier and yield an incorrect balance of the *syn* and *anti* region.

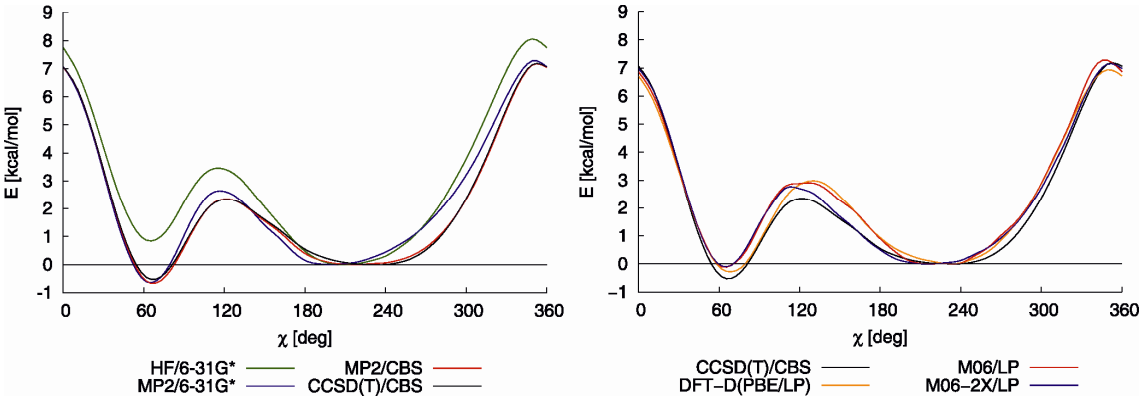


Figure 16: *Torsion profiles for guanine nucleoside calculated at various levels of theory in vacuo.*

Effects of Geometry Relaxation

Currently used torsion parameters are usually obtained using only the QM optimized geometries for both the QM and MM levels of calculations. However, it was found that using the same geometry for deriving parameters as difference between QM and MM energies may introduce errors in the resulting profiles (see Figure 17). On the other hand, when both the QM and MM geometries are relaxed independently, the MM potential energy surface derived is more similar to the QM potential energy surface (PES) in terms of the relative energies of key PES regions (minima, barrier heights). Therefore, in this study, geometry relaxation was carried out separately at both the QM and MM levels.

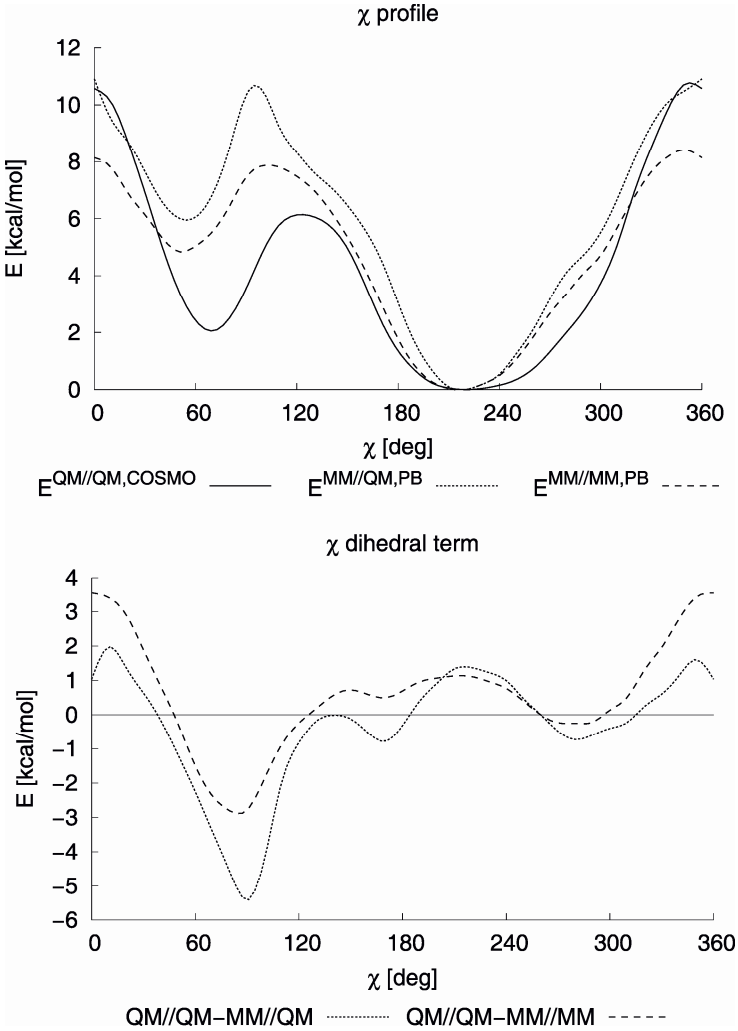


Figure 17: *Top, torsion profiles for cytosine nucleoside calculated as QM energy based on QM-optimized geometry ($E^{QM//QM,COSMO}$, full line), MM_{-c} energy based on MM-optimized geometry ($E_{-c}^{MM//MM,PB}$, dashed line) and MM_{-c} energy based on QM-optimized geometry ($E_{-c}^{MM//QM,PB}$, dotted line). Bottom, c dihedral terms $E_{dih,c}^{solv}$ derived from $E^{QM//QM,COSMO}$ - $E_{-c}^{MM//MM,PB}$ (QM//QM-MM//MM, dashed line) and $E^{QM//QM,COSMO}$ - $E_{-c}^{MM//QM,PB}$ (QM//QM-MM//QM, dotted line) normalized to c = 250°. The parmbsc0 force field was used in all cases, and energies are in kcal/mol.*

Including Solvent Effects

An alternate way of deriving torsion parameters intended for MD simulations in solvent was examined. The key idea is inclusion of certain nonspecific solvation effects that are neglected when torsion parameters are obtained in vacuum and that are not covered by the solvent models used in the MD simulations. Two perhaps most important effects that contribute to variation of the solvation energy with conformation are: (i) solvent response to conformation-induced variation in solute charge distribution, and, (ii) conformation-dependent solute-solvent polarization. Magnitude of these two effects was investigated for the *c* torsion angle using a modified cytosine nucleoside as a model (results not shown). The both mentioned effects are rather large in magnitude

and significantly modify the shape of the *c* torsion profile. If they are neglected, incorrect barrier heights, relative stability of the *syn* and *anti* regions, and relative stability of the *anti* and high-*anti* regions is obtained.

5.3.1 Derivation of c Parameters

Considering all the above mentioned points, a viable procedure including the missing solvation effects in the torsion angle parameterization based on the difference between the self-consistent reaction field (QM/COSMO) and Poisson-Boltzmann (MM/PB) continuum solvation energies was suggested (Equation 10).

$$E_{\mathrm{dih},\chi}^{\mathrm{solv}} = E^{\mathrm{QM//QM,COSMO}} - E_{-\chi}^{\mathrm{MM//MM,PB}} \qquad 10$$

In this way only the part of the solvation effects that cannot be described by the solvent model used in the MD simulation is included in the torsion parameters and double counting of solvation contributions is avoided. However, some portion of the solvation energy is included in the internal energy of the solute (in the torsion parameters), and, therefore, the torsion parameters can be used only for the MD simulations in solvent (water is considered here) and they are not valid for simulations carried out in vacuum. The derived torsion parameters can be used for simulations with any standard non-polarizable explicit solvent model and even with implicit solvent models. Using our parameterization approach, new parameters for the glycosidic torsion angle (χ_{OL}, where OL stands for Olomouc), intended for use in RNA simulations with Cornell *et al.* force field, were derived.

5.3.2 Verification of Derived Parameters for A-RNA Duplexes

The newly derived parameters have been extensively tested on 1RNA [30], 1QC0 [169], and 2R20' [170] A-RNA duplexes. Modified parameters χ_{OL} have been compared to Yildirim *et al.* modification (χ_{YIL}) [118] and original ff99 [105] parameters set, both with and without parmbsc0 [107] correction. The Ode *et al.* (χ_{ODE}) [117] parameterization was not considered in the simulations as it accelerates formation of "ladder-like" structures (Appendix III). The χ torsion profiles and χ dihedral terms for all available χ modifications (including χ_{ODE}) are for illustration shown in Figure 18.

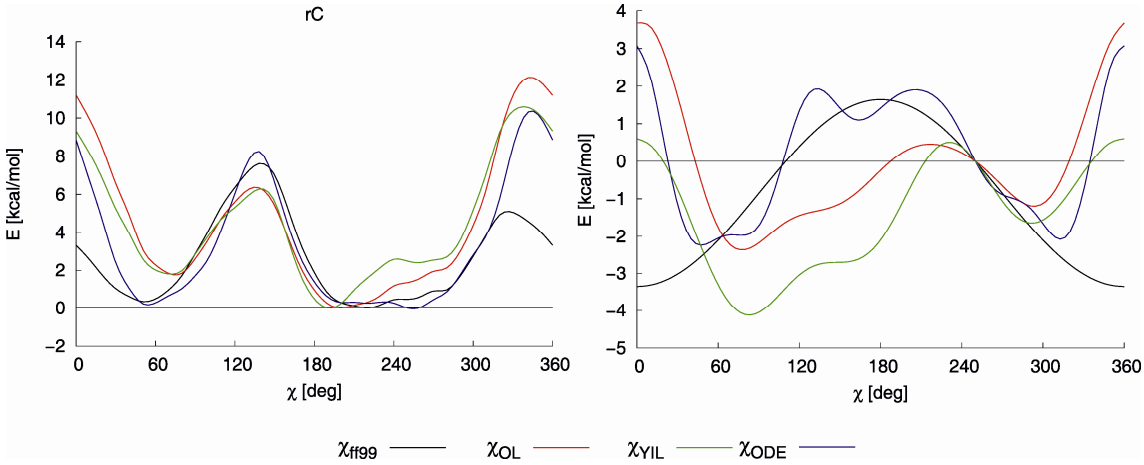


Figure 18: *Torsion profiles for the c angle (left) and the c dihedral terms (right) for model of cytidine; ff99 (black), Ode et al. (blue), Yildirim et al. (green) and parameters derived herein (c_{OL}, red).*

Among the most sensitive parameters for A-RNA are inclination, roll, major groove width and propeller twist. Table 6 shows the average values of the parameters for one of the A-RNA duplexes only because it well illustrate the general trends in all tested A-RNA duplexes.

Table 6: *Average structural parameters (last 20 ns of 100 ns simulations) for the A-RNA duplex 1RNA obtained using the ff99bsc0 force field with various χ corrections (values with ff99 force field in italics). Standard deviations are shown for the unmodified force fields for orientation and they are very similar for the other force fields. RMSD is mass-weighted, for all atoms.*

Parameter	X-RAY	no χ correction	χ_{YIL}	χ_{OL}
χ/°	199.3±9.0	213.3±14.0	199.3	205.6
		211.2±12.0	*197.2*	*207.5*
P/°	18.4 ±1.6	24.1 ±17.2	15.8	19.7
		25.0±19.3	*14.4*	*21.9*
Minor groove width/Å	16.1±0.4	16.2±0.8	15.5	15.9
		15.9±0.8	*15.2*	*15.8*
Major groove width/Å	12.3±1.5	15.3±2.4	15.9	14.8
		17.1±2.5	*18.1*	*16.9*
Slide/Å	−1.29±0.1	−1.06±0.59	−1.70	−1.40
		-1.29±0.51	*-1.92*	*-1.40*
Roll/°	9.96±5.1	14.10±8.12	7.47	11.05
		12.07±6.81	*5.60*	*10.40*
Propeller/°	−18.8±2.3	−11.5±10.6	−13.4	−15.1
		-14.5±9.4	*-11.5*	*-9.8*
Inclination/°	18.8±10.8	25.8±14.1	13.2	19.4
		21.4±12.0	*10.3*	*19.1*
Helical twist/°	32.7±4.1	32.3±6.0	32.0	32.8
		32.7±4.7	*31.3*	*31.6*
RMSD/Å		2.03	1.96	1.70
		1.96	*2.44*	*2.12*

For inclination, the experimental value is 18.8°, while ff99bsc0, ff99, ff99bsc0χ_{OL} and ff99χ_{YIL} give values of 25.8°, 21.4°, 19.4° and 10.3°, respectively. In this case, the ff99bsc0χ_{OL} value is closest to the experimental data. The ff99χ_{YIL} inclination is likely too low. The inclination trend is mirrored by roll values: experimental value is 9.96°, while ff99bsc0, ff99, ff99bsc0χ_{OL} and ff99χ_{YIL} values are 14.1°, 12.1°, 11.1° and 5.7°, respectively. The experimental value for major groove width is 12.3 Å, while ff99bsc0, ff99, ff99bsc0χ_{OL} and ff99χ_{YIL} give 15.3 Å, 17.1 Å, 14.8 Å and 18.1 Å, respectively. The *anti* stabilization and preventing a formation of ladder-like structures influences the absolute value of propeller twist; the experimental value for this variable of 1RNA is −18.8°, and we obtained values of −11.5°, −14.5°, −15.1° and −11.5° using ff99bsc0, ff99, ff99bsc0χ_{OL} and ff99χ_{YIL}, respectively.

Our results show that ff99bsc0χ_{OL} is the best combination of parameters currently available for A-RNA. Its use eliminates emergence of the "ladder-like" structures but still allows A-RNA to adopt significant inclination, roll and propeller twist.

5.3.3 Verification of Derived Parameters for Tetraloops

The main goal of the refinement of the χ angle was to correct the undesirable destabilization of the *anti* (200°) region with respect to the high-*anti* (250°) region observed with the ff99 and ff99parmbsc0 force fields, which led to the formation of the "ladder-like" structures in MD simulations of RNA molecules. An extensive set of MD simulations (Appendix III) on UUCG, GAGA, and GAAA tetraloops has shown that new parameters for χ torsion help to prevent the formation of "ladder-like" formation and subsequent degradation of the structure.

Furthermore, using the χ_{OL} modification in connection with the ff99bsc0 for UUCG tetraloop has led to stabilization of some signature interactions. This improvement is probably due to reliable description of the *syn* region of G in all studied modification of χ torsion.

In conclusion, the correlation between occurrence of the high-*anti* substates and formation of the "ladder-like" structures, which were first revealed in our laboratory by Mlynsky ***et al.*** [168] was confirmed. The newly derived parameters successfully remove this artifact mainly thanks to improvements in the parameterization procedure suggested in this work. χ_{OL} parameters are now available as a default parameter set for RNA simulations in the ff10 variant of the Cornell ***et al.*** force field that collects a variety of modifications and updates to the ff99 force field.

6 SUMMARY

This thesis was aimed at studying intermolecular interactions in biomolecules with the most modern methods of computational chemistry. Both quantum mechanical and molecular mechanical methods have been used to investigate binding and conformational properties of biologically relevant molecules.

This work consists of three main parts. The first part deals with reference quantum chemical calculations of noncanonical interactions in RNA. Attention was paid to interactions of non-Watson-Crick base pairs directly involving ribose 2'-OH group and interactions between bases and phosphate groups. Several quantum mechanical methods have been used in order to estimate stabilization *in vacuo* and in water. Obtaining the reference quantum chemical results is particularly important for assessment of the performance and reliability of lower-level quantum mechanical methods. In addition, these studies can be useful as a reference for eventual force fields refinements.

Performing a series of quantum chemical calculations for noncanonical binding patterns in RNA demonstrates a quantitative evaluation of stabilization energy. The reference CBS(T) calculations reveal a wide range of interaction energies of the individual patterns of the base phosphate interactions as well as non-WC base pairing interactions directly involving the ribose 2'-OH group interactions. Base phosphate interactions range from −1 to −36 kcal/mol, while the gas phase energies for the 25 base pairs in which the sugar interaction is involved range from −10 to −31 kcal/mol. This means that the strongest noncanonical interactions are comparable in stability to the WC base pairs (C=G, −32 kcal/mol) and thus may play a substantial role in stabilization of all types of RNA molecules.

Decomposition of energy contribution into the physically meaningful components by the DFT-SAPT methods pointed out larger contribution of induction energy in the set of base – phosphate interactions compared to regular H-bonds caused by stronger electric field created by the phosphate group. Comparison of the lower-level QM methods with the reference showed very good performance of the DFT-D method for both sets of molecules.

In the second part of the work, I focused on gaining insight into individual force field terms. The individual energy components calculated with reference DFT-SAPT method have been compared with the related terms obtained by the Cornell *et al.* force field family. This kind of analysis may be important for potential future force field development and readjustment.

Although comparison of DFT-SAPT and force fields values is not straightforward, several conclusions stem from the distance dependence of the interaction energy components. At large distances both the coulombic and dispersion empirical terms correspond fairly well with the DFT-SAPT electrostatics and dispersion terms, respectively. At the onset of electronic overlap the electrostatic term is strongly underestimated due to missing electron overlap effects in the force fields. Apart from the observed systematic deviations (introduced mainly by the anisotropy of the polarizabilities) the overall force field dispersion agrees quite well with the DFT-SAPT

reference values. It is proposed that force field dispersion parameters could be obtained directly from theoretical calculations in future parameterization efforts. The polarization term in the Cieplak *et al.* force field seems to be underestimated with respect to the DFT-SAPT results at both large and short intermolecular distances. Among all terms, the repulsion energy is a source of significant part of errors in the force fields. However, this term also plays an essential role in the error cancellation. Its underestimation at vdW distances compensates the underestimation of all attractive components. At very short distances, the repulsion component is strongly overestimated due to use of $1/r^{12}$ term in the Lennard-Jones formula in the current force fields. We assume that replacing this approximation by a more physical exponential form might bring a significant improvement. In spite of errors of individual energy components, global performance of the force field is in reasonable agreement with reference QM results.

The nucleic acid force field development is a subject of the last part of my thesis. An advanced parameterization procedure for derivation of new torsion parameters has been suggested, considering several factors (such as choice of the model system, level of QM calculations and inclusion of solvent) that affect final parameters. New parameterization methodology has been used to derive new parameters for the glycosidic χ torsion in nucleic acids.

Modified parameters for the glycosidic torsion angle (χ_{OL}, where the abbreviation OL stands for Olomouc) are based on the accurate high-level QM reference data obtained for relatively large model molecule. Inclusion of previously neglected solvation effects led to the correctness of undesirable destabilization of the *anti* region with respect to the high-*anti* region observed in MD simulations. This inappropriate description of the *anti* vs high-*anti* balance resulted in the recently identified artifacts in MD simulations of RNA hairpin loops - formation of “ladder-like” structures. Suggested parameters have successfully prevented formation of the ladders and represent a significant improvement in MD simulations of noncanonical structures of nucleic acids. χ_{OL} parameters are now available as a default parameter set for RNA simulations in the newest variant of the Cornell *et al.* force field.

The present study shows that current quantum mechanical methods may provide valuable insight into the function of the empirical force fields. It is worth noting that careful consideration of the abovementioned effects (i.e., behavior of the force field components and solvation and relaxation effects) is crucial when QM methods are used in force field refinement and development.

7 SHRNUTÍ

Předložená dizertační práce se zabývá popisem mezimolekulových interakcí v biomolekulách prostředky výpočetní chemie. Ke studiu zejména vazebných a konformačních vlastností v biologicky relevantních molekulách či jejich modelech byly použity jak kvantově mechanické, tak molekulově mechanické metody.

Předmětem práce jsou zejména tři hlavní témata. Prvním je popis nekanonických interakcí v RNA prostřednictvím kvantově chemických výpočtů. Pozornost je věnována hlavně interakcím neobsahujícím klasické Watson-Crickovo párování bází nukleových kyselin a zahrnujícím 2'-OH skupinu ribózy přímo do interakce a dále pak interakcím mezi bázemi a fosfátovou skupinou. K odhadu jejich stability ve vakuu i ve vodě bylo použito několik kvantově chemických metod. Referenční kvantově - mechanická data jsou obzvláště důležitá pro posouzení kvality méně náročných metod, popř. mohou mít význam pro eventuální vývoj silových polí.

Řada použitých kvantově chemických výpočtů je schopna poskytnout kvantitativní odhad stabilizační energie. Referenční CBS(T) metoda ukazuje široké rozpětí energií pro jednotlivé vazebné motivy interakcí báze-fosfát, stejně jako pro jednotlivé motivy zahrnující 2'-OH skupinu ribózy. Interakční energie získané touto metodou se pro interakce báze-fosfát pohybují v rozmezí od −1 do −36 kcal/mol, vakuové energie pro páry bází zahrnující do interakce cukr v rozmezí −10 až −31 kcal/mol. Nejsilnější testované nekanonické interakce přispívají k stabilitě srovnatelně jako kanonické WC páry (C=G, −32 kcal/mol) a hrají tak velmi důležitou roli v stabilizaci RNA struktur.

Dekompozice energie na jednotlivé příspěvky za použití metody DFT-SAPT poukázala na větší vliv indukční energie u interakcí báze-fosfát v porovnání s klasickou vodíkovou vazbou v důsledku silného elektrického pole fosfátové skupiny. Srovnání referenčních výpočtů s výpočetně méně náročnými kvantově chemickými metodami prokázalo velmi dobrou spolehlivost metody DFT-D pro obě skupiny struktur.

V druhé části práce byla pozornost soustředěna na důkladnou analýzu jednotlivých energetických příspěvků silového pole. Individuální energetické příspěvky získané metodou DFT-SAPT byly srovnávány s příslušnými příspěvky silového pole Cornell a kol. Tento druh analýzy má význam především pro budoucí vývoj empirických silových polí.

Ačkoliv srovnání DFT-SAPT a empirických příspěvků není přímočaré, několik závěrů lze vyvodit ze závislosti jednotlivých energetických komponent na vzdálenosti interagujících molekul. Na velkých vzdálenostech souhlasí elektrostatická i disperzní složka relativně dobře s korespondujícími DFT-SAPT složkami. Elektrostatická energie je na vzdálenostech kolem rovnovážného minima podceněná zejména kvůli zanedbání překryvových efektů v silových polích. Až na systematické odchylky (způsobené hlavně anizotropií polarizabilit) je disperzní část empirické energie v dobré shodě s referenčními DFT-SAPT hodnotami. Výsledky ukázaly, že empirické parametry atraktivní části van der Waalsovy energie by v budoucnu mohly být získávány přímo z kvantově mechanických výpočtů. Indukční energie je v silových polí podceněná v celém rozsahu. Repulzní složka, přestože je v silových polích zdrojem největších chyb,

hraje velmi důležitou roli v celkové kompenzaci chyb. Její podcenění na rovnovážných vzdálenostech kompenzuje podcenění všech ostatních (přitažlivých) složek. Na velmi krátkých vzdálenostech je repulzní energie silně přeceněna z důvodu použití $1/r^{12}$ složky Lennard-Jonesova vztahu. Aproximace repulzní složky v silových polích exponenciální formou by mohla vést k značnému zlepšení. Navzdory chybám jednotlivých energetických složek je celková energie a tudíž i celková přesnost silových polí v relativně dobrém souladu s referenčními kvantově mechanickými výpočty.

Třetí část dizertační práce se zabývá vývojem silových polí pro nukleové kyseliny. Byla studována např. volba vhodného modelového systému pro odvození parametrů, úroveň kvantově mechanických výpočtů a možnost zahrnutí solvatačních efektů. Na základě výsledků byl navržen nový postup pro parametrizaci torzních úhlů. Ten byl následně použit na odvození nových parametrů pro glykosidický torzní úhel (úhel χ) v nukleových kyselinách.

Modifikované parametry pro glykosidický torzní úhel (χ_{OL}) jsou založeny na přesných kvantově mechanických referenčních výpočtů a zahrnují solvatační efekty, které byly v dřívějších parametrizacích zanedbávány. Molekulárně dynamické (MD) simulace s takto navrženými parametry vedly k opravě nežádoucí destabilizace ***anti*** oblasti vzhledem k high-***anti***, která v MD simulacích nukleových kyselin způsobovala nedávno identifikovaný artefakt – tzv. žebříkovou strukturu. Navržené parametry úspěšně zamezily formování žebříkových struktur a vedly k výraznému zlepšení chování nekanonických struktur nukleových kyselin v MD simulacích. χ_{OL} parametry jsou v současnosti dostupné jako výchozí parametry pro MD simulace RNA struktur v nejnovější verzi silového pole Cornell a kol. (ff10).

Předložená práce ukazuje, že současné metody kvantové chemie jsou schopny poskytnout cenný vhled do funkce silových polí. Současně však dokazuje, že mají-li být kvantově mechanické metody použity k vývoji silových polí je zohlednění výše zmíněných efektů (solvatačních, relaxačních i týkajících se jednotlivých energetických příspěvků) zcela nezbytné.

8 LIST OF PUBLICATIONS

Šponer J., **Zgarbová M.**, Jurečka P., Riley K. E., Šponer J. E., Hobza P.: Reference quantum chemical calculations on RNA base pairs directly involving the 2´-OH group of ribose, *J. Chem. Theory Comput., 5(4), 1166-1179,* **2009**
Attached as Appendix I

Zgarbová, M.; Otyepka, M.; Šponer, J.; Hobza, P.; Jurečka, P.: Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys., 12 (35), 10476-10493,* **2010**
Attached as Appendix II

Banáš, P.; Hollas, D.; **Zgarbová, M.**; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M.: Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput. 6 (12), 3836-3849,* **2010**
Attached as Appendix III

Zgarbová, M.; Jurečka, P.; Banáš, P.; Otyepka, M.; Šponer, J. E.; Leontis, N. B.; Šponer, J.: Noncanonical hydrogen bonding in nucleic acids. Benchmark evaluation of key base – phosphate interactions in folded RNA molecules using quantum-chemical calculations and molecular dynamics simulations. *J. Phys. Chem. A, submitted,* **2011**
Attached as Appendix IV

Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T.E.,III; Jurečka, P.:Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput., accepted,* **2011**
Attached as Appendix V

9 LIST OF ABBREVIATIONS

A	Adenine
BPh	Base-Phosphate
BSSE	Basis Set Superposition Error
C	Cytosine
CBS	Complete Basis Set
CC	Coupled Cluster
CI	Configuration Interaction
COSMO	Conductor like Screening Model
DFT	Density Functional Theory
DFT-D	DFT augmented with empirical dispersion
DNA	Deoxyribonucleic Acid
FF	Force Field
G	Guanine
GB	Generalized Born
GGA	Global Gradient Approximation
HB	Hydrogen Bonded
HF	Hartree Fock
LDA	Local Density Approximation
LJ	Lennard-Jones
MD	Molecular Dynamics
MM	Molecular Mechanical/Mechanics
MP	Møller-Plesset
PB	Poisson Boltzmann
PBE	Poisson Boltzmann Equation
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
QM	Quantum Mechanical/Mechanics
RESP	Restrained Electrostatic Potential
RNA	Ribonucleic Acid
SAPT	Symmetry Adapted Perturbation Theory
SASA	Solvent Accessible Surface Area
SCRF	Self Consistent Reaction Field
SE	Sugar Edge
T	Thymine
U	Uracil
vdW	van der Waals
WC	Watson-Crick
WFT	Wave Function Theory

10 REFERENCES

1. Stone, A. J. in *The Theory of Intermolecular Forces*. Oxford University Press Inc.: New York, 1997.

2. Kaplan, I. G. in *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*. John Wiley & Sons: Chichester, 2006.

3. Neidle, S. in *Principles of Nucleic Acid Structure*. Elsevier, Inc.: London, 2008.

4. Schlick, T. in *Molecular Modeling and Simulation - An Interdisciplinary Guide*. Springer: New York, 2002.

5. Schneider, B.; Berman, H. B. in *Computational studies of RNA and DNA*. Springer: Dordrecht, 2006; p 1-44.

6. Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001,** *7* (4), 499-512.

7. Leontis, N. B.; Westhof, E., Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.* **1998,** *31* (4), 399-455.

8. Leontis, N. B.; Stombaugh, J.; Westhof, E., The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002,** *30* (16), 3497-3531.

9. Lescoute, A.; Westhof, E., The A-minor motifs in the decoding recognition process. *Biochimie* **2006,** *88* (8), 993-999.

10. Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A., RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. USA* **2001,** *98* (9), 4899-4903.

11. Tamura, M.; Holbrook, S. R., Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **2002,** *320* (3), 455-474.

12. Gagnon, M. G.; Steinberg, S. V., GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* **2002,** *8* (7), 873-877.

13. Mokdad, A.; Krasovska, M. V.; Sponer, J.; Leontis, N. B., Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **2006,** *34* (5), 1326-1341.

14. Leontis, N. B.; Westhof, E., The 5S rRNA loop E: Chemical probing and phylogenetic data versus crystal structure. *RNA* **1998,** *4* (9), 1134-1153.

15. Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E., Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **2005,** *33* (8), 2395-2409.

16. Mokdad, A.; Frankel, A. D., ISFOLD: Structure prediction of base pairs in non-helical RNA motifs from isostericity signatures in their sequence alignments. *J. Biomol. Struct. Dyn.* **2008,** *25* (5), 467-472.

17. Razga, F.; Koca, J.; Mokdad, A.; Sponer, J., Elastic properties of ribosomal RNA building blocks: molecular dynamics of the GTPase-associated center rRNA. *Nucleic Acids Res.* **2007,** *35* (12), 4007-4017.

18. Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B., FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* **2008,** *56* (1-2), 215-252.

19. Sponer, J.; Jurecka, P.; Hobza, P. in *Computational studies of RNA and DNA.* Springer: Dordrecht, 2006; p 343-388.

20. Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B., Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* **2009,** *37* (15), 4898-4918.

21. Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D., Single nucleotide RNA choreography. *Nucleic Acids Res.* **2006,** *34* (5), 1481-1491.

22. Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; Daubentoncarafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; Brody, E. N.; Gold, L., CUUCGG Hairpins - Extraordinarily stable RNA secondary structures sssociated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* **1988,** *85* (5), 1364-1368.

23. Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M., Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **2010,** *6* (12), 3836-3849.

24. Correll, C. C.; Munishkin, A.; Chan, Y. L.; Ren, Z.; Wool, I. G.; Steitz, T. A., Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. USA* **1998,** *95* (23), 13436-13441.

25. Spackova, N.; Sponer, J., Molecular dynamics simulations of sarcin-ricin rRNA motif. *Nucleic Acids Res.* **2006,** *34* (2), 697-708.

26. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L., The anatomy of A-DNA, B-DNA, and Z-DNA. *Science* **1982,** *216* (4545), 475-485.

27. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981,** *78* (4), 2179-2183.

28. Conner, B. N.; Yoon, C.; Dickerson, J. L.; Dickerson, R. E., Helix geometry and hydration in an A-DNA tetramer: IC-C-G-G. *J. Mol. Biol.* **1984,** *174* (4), 663-695.

29. Gessner, R. V.; Frederick, C. A.; Quigley, G. J.; Rich, A.; Wang, A. H., The molecular structure of the left-handed Z-DNA double helix at 1.0-A atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J. Biol. Chem.* **1989,** *264* (14), 7921-7935.

30. Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D., Crystallographic structure of an RNA helix: [U(UA)6A]2. *J. Mol. Biol.* **1989,** *209* (3), 459-474.

31. Batey, R. T.; Rambo, R. P.; Doudna, J. A., Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.* **1999,** *38* (16), 2326-2343.

32. Xin, Y. R.; Laing, C.; Leontis, N. B.; Schlick, T., Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA* **2008,** *14* (12), 2465-2477.

33. Woese, C. R.; Winker, S.; Gutell, R. R., Architecture of ribosomal-RNA - constraints on the sequence of tetraloops. *Proc. Natl. Acad. Sci. USA* **1990,** *87* (21), 8467-8471.

34. Misra, V. K.; Blose, J. M.; Proctor, D. J.; Veeraraghavan, N.; Bevilacqua, P. C., Contribution of the closing base pair to exceptional stability in RNA tetraloops: Roles for molecular mimicry and electrostatic factors. *J. Am. Chem. Soc.* **2009,** *131* (24), 8474-8484.

35. Uhlenbeck, O. C., Nucleic-acid structure - tetraloops and RNA folding. *Nature* **1990,** *346* (6285), 613-614.

36. Varani, G., Exceptionally stable nucleic-acid hairpins. *Annu. Rev. Bioph. Biom.* **1995,** *24*, 379-404.

37. Marino, J. P.; Gregorian, R. S.; Csankovszki, G.; Crothers, D. M., Bent helix formation between RNA hairpins with complementary loops. *Science* **1995,** *268* (5216), 1448-1454.

38. Woodson, S. A.; Chauhan, S., Tertiary interactions determine the accuracy of RNA folding. *J. Am. Chem. Soc.* **2008,** *130* (4), 1296-1303.

39. Jaeger, L.; Michel, F.; Westhof, E., Involvement of a GNRA tetraloop in long-range tertiary interactions. *J. Mol. Biol.* **1994,** *236* (5), 1271-1276.

40. Schneider, B.; Neidle, S.; Berman, H. M., Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers* **1997,** *42* (1), 113-124.

41. Schneider, B.; Moravek, Z.; Berman, H. M., RNA conformational classes. *Nucleic Acids Res.* **2004,** *32* (5), 1666-1677.

42. Lu, X. J.; ElHassan, M. A.; Hunter, C. A., Structure and conformation of helical nucleic acids: Analysis program (SCHNAaP). *J. Mol. Biol.* **1997,** *273* (3), 668-680.

43. Dickerson, R. E., Definitions and nomenclature of mucleic acid structure components. *Nucleic Acids Res.* **1989,** *17* (5), 1797-1803.

44. Szabo, A.; Ostlund, N. S. in *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory.* Dover Publications, Inc.: New York, 1996.

45. Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P.; Schaefer, H. F., III in *Encyklopedia of Computational Chemistry*. John Wiley & Sons: New York, 1998.

46. Moller, C.; Plesset, M. S., Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934,** *46* (7), 0618-0622.

47. Grimme, S., Improved second-order Moller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *J. Chem. Phys.* **2003,** *118* (20), 9095-9102.

48. Head-Gordon, M.; Distasio, R. A., Optimized spin-component scaled second-order Moller-Plesset perturbation theory for intermolecular interaction energies. *Mol. Phys.* **2007,** *105* (8), 1073-1083.

49. Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P., Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006,** *8* (17), 1985-1993.

50. Cizek, J., On correlation problem in atomic and molecular systems . Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.* **1966,** *45* (11), 4256.

51. Paldus, J.; Shavitt, I.; Cizek, J., Correlation problems in atomic and molecular systems .4. Extended coupled-pair many-electron theory and its application to BH3 molecule. *Phys. Rev. A* **1972,** *5* (1), 50.

52. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Headgordon, M., A 5th-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989,** *157* (6), 479-483.

53. Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P., Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.* **2010,** *110* (9), 5023-5063.

54. Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Phys. Rev. B* **1964,** *136* (3B), B864.

55. Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140* (4A), 1133.

56. Zhao, Y.; Truhlar, D. G., Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008,** *41* (2), 157-167.

57. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* **1988,** *38* (6), 3098-3100.

58. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **1988,** *37* (2), 785-789.

59. Becke, A. D., Density-functional thermochemistry .3. The role of exact exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652.

60. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *J. Phys. Chem.-US* **1994,** *98* (45), 11623-11627.

61. Adamo, C.; Barone, V., Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110* (13), 6158-6170.

62. Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E., Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003,** *91* (14), 146401.

63. Koch, W.; Holthausen, M. C. in *A Chemist's Guide to Density Functional Theory*. 2nd ed.; WILEY-VCH: Weinheim, New York,Chichester, 2001.

64. Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G., Adiabatic connection for kinetics. *J. Phys. Chem. A* **2000,** *104* (21), 4811-4815.

65. Zhao, Y.; Truhlar, D. G., Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A* **2004,** *108* (33), 6908-6918.

66. Zhao, Y.; Truhlar, D. G., A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006,** *125* (19), 194101.

67. Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R., Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem.* **2007,** *28* (2), 555-569.

68. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004,** *25* (12), 1463-1473.

69. Grimme, S., Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006,** *27* (15), 1787-1799.

70. Grimme, S., Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006,** *124* (3).

71. Jansen, H. B.; Ros, P., Non-empirical molecular orbital calculations on the protonation of carbon monoxide. *Chem. Phys. Lett.* **1969,** *3* (3), 140-143.

72. Boys, S. F.; Bernardi, F., Calculation of small molecular interactions by differences of separate total energies - some procedures with reduced errors. *Mol. Phys.* **1970,** *19* (4), 553-&.

73. Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994,** *94* (7), 1887-1930.

74. Misquitta, A. J.; Jeziorski, B.; Szalewicz, K., Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.* **2003,** *91* (3), -.

75. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K., Intermolecular potentials based on aymmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005,** *123* (21), 214103.

76. Misquitta, A. J.; Szalewicz, K., Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002,** *357* (3-4), 301-306.

77. Podeszwa, R.; Bukowski, R.; Szalewicz, K., Density-fitting method in symmetry-adapted perturbation theory based on Kohn-Sham description of monomers. *J. Chem. Theory Comput.* **2006,** *2* (2), 400-412.

78. Hesselmann, A.; Jansen, G., First-order intermolecular interaction energies from Kohn-Sham orbitals. *Chem. Phys. Lett.* **2002,** *357* (5-6), 464-470.

79. Hesselmann, A.; Jansen, G., Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn-Sham density functional theory. *Chem. Phys. Lett.* **2002,** *362* (3-4), 319-325.

80. Hesselmann, A.; Jansen, G., Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.* **2003,** *367* (5-6), 778-784.

81. Hesselmann, A.; Jansen, G.; Schutz, M., Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005,** *122* (1), 014103.

82. Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K., Basis-set convergence in correlated calculations on Ne, N-2, and H2O. *Chem. Phys. Lett.* **1998,** *286* (3-4), 243-252.

83. Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J., Basis-set convergence of the energy in molecular Hartree-Fock calculations. *Chem. Phys. Lett.* **1999,** *302* (5-6), 437-446.

84. Weigend, F.; Haser, M., RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **1997,** *97* (1-4), 331-340.

85. Jurecka, P.; Nachtigall, P.; Hobza, P., RI-MP2 calculations with extended basis sets - a promising tool for study of H-bonded and stacked DNA base pairs. *Phys. Chem. Chem. Phys.* **2001,** *3* (20), 4578-4582.

86. Leach, A. in *Molecular modelling: principles and applications*. 2nd ed.; Prentice Hall: 2001.

87. Mackerell, A. D., Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004,** *25* (13), 1584-1604.

88. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A., Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Bioph. Biom.* **2001,** *30*, 211-243.

89. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

90. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.

91. van Gunsteren, W. F. *GROMOS. Groningen Molecular Simulation Program Package*, University of Groningen: Groningen, 1987.

92. Jorgensen, W. L.; Tiradorives, J., The OPLS potential functions for proteins - energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* **1988,** *110* (6), 1657-1666.

93. Lii, J. H.; Allinger, N. L., The MM3 force-field for amides, polypeptides and proteins. *J. Comput. Chem.* **1991,** *12* (2), 186-199.

94. Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J. R.; Hwang, M. J.; Li, H. Y.; Liang, C.; Maple, J.; Peng, Z. W.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X. S.; Hagler, A. T., Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. *J. Comput. Chem.* **2001,** *22* (15), 1782-1800.

95. Sun, H., COMPASS: An ab initio force-field optimized for condensed-phase applications - Overview with details on alkane and benzene compounds. *J. Phys. Chem. B* **1998,** *102* (38), 7338-7364.

96. Derreumaux, P.; Vergoten, G., A new spectroscopic molecular mechanics force field - parameters for proteins. *J. Chem. Phys.* **1995,** *102* (21), 8586-8605.

97. Halgren, T. A., Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996,** *17* (5-6), 490-519.

98. Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S., Potential energy functions: From consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers* **2003,** *68* (3), 383-394.

99. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - a generic force field for molecular simulations. *J. Phys. Chem.-US* **1990,** *94* (26), 8897-8909.

100. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., UFF, a full periodic-table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992,** *114* (25), 10024-10035.

101. Buckingham, A. D.; Fowler, P. W., A model for the geometries of van der Waals complexes. *Can. J. Chem.* **1985,** *63* (7), 2018-2025.

102. Halgren, T. A., Merck molecular force field .2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996,** *17* (5-6), 520-552.

103. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.-US* **1993,** *97* (40), 10269-10280.

104. Halgren, T. A.; Damm, W., Polarizable force fields. *Curr. Opin. Struc. Biol.* **2001,** *11* (2), 236-242.

105. Wang, J. M.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21* (12), 1049-1074.

106. Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A., A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999,** *16* (4), 845-862.

107. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007,** *92* (11), 3817-3829.

108. Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M., An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995,** *117* (48), 11946-11975.

109. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.

110. Foloppe, N.; MacKerell, A. D., All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000,** *21* (2), 86-104.

111. MacKerell, A. D.; Banavali, N. K., All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000,** *21* (2), 105-120.

112. Feig, M.; Pettitt, M., Experiment vs force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem. B* **1997,** *101* (38), 7361-7363.

113. Feig, M.; Pettitt, B. M., Structural equilibrium of DNA represented with different force fields. *Biophys. J.* **1998,** *75* (1), 134-149.

114. Cheatham, T. E.; Kollman, P. A., Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **1996,** *259* (3), 434-444.

115. Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B., DNA conformations and their sequence preferences. *Nucleic Acids Res.* **2008,** *36* (11), 3690-3706.

116. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.

117. Ode, H.; Matsuo, Y.; Neya, S.; Hoshino, T., Force field parameters for rotation around chi torsion axis in nucleic acids. *J. Comput. Chem.* **2008,** *29* (15), 2531-2542.

118. Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H., Reparameterization of RNA chi torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput.* **2010,** *6* (5), 1520-1531.

119. Mahoney, M. W.; Jorgensen, W. L., A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000,** *112* (20), 8910-8922.

120. Mahoney, M. W.; Jorgensen, W. L., Diffusion constant of the TIP5P model of liquid water. *J. Chem. Phys.* **2001,** *114* (1), 363-366.

121. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

122. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *J. Phys. Chem.-US* **1987,** *91* (24), 6269-6271.

123. Cramer, C. J.; Truhlar, D. G., Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999,** *99* (8), 2161-2200.

124. Orozco, M.; Luque, F. J., Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000,** *100* (11), 4187-4226.

125. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chem. Rev.* **2005,** *105* (8), 2999-3093.

126. Tomasi, J.; Persico, M., Molecular interactions in solution - an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **1994,** *94* (7), 2027-2094.

127. Miertus, S.; Scrocco, E.; Tomasi, J., Electrostatic interaction of a solute with a continuum - a direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981,** *55* (1), 117-129.

128. Klamt, A.; Schuurmann, G., Cosmo - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perk. T 2* **1993,** (5), 799-805.

129. Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A., Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun. Comput. Phys.* **2008,** *3* (5), 973-1009.

130. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990,** *112* (16), 6127-6129.

131. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997,** *101* (16), 3005-3014.

132. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A., Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **1996,** *273* (5282), 1678-1685.

133. Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A., The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **2000,** *289* (5481), 905-920.

134. Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V., Structure of the 30S ribosomal subunit. *Nature* **2000,** *407* (6802), 327-339.

135. Moore, P. B.; Steitz, T. A., The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.* **2003,** *72*, 813-850.

136. Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H., Structures of the bacterial ribosome at 3.5 A resolution. *Science* **2005,** *310* (5749), 827-834.

137. Korostelev, A.; Trakhanov, S.; Laurberg, M.; Noller, H. F., Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **2006,** *126* (6), 1065-1077.

138. Selmer, M.; Dunham, C. M.; Murphy, F. V., IV; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V., Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **2006,** *313* (5795), 1935-1942.

139. Battle, D. J.; Doudna, J. A., Specificity of RNA-RNA helix recognition. *Proc. Natl. Acad. Sci. USA* **2002,** *99* (18), 11676-11681.

140. Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J., Non-Watson-Crick base pairing in RNA. quantum chemical analysis of the cis Watson-Crick/sugar edge base pair family. *J. Phys. Chem. A* **2005,** *109* (10), 2292-2301.

141. Sponer, J. E.; Reblova, K.; Mokdad, A.; Sychrovsky, V.; Leszczynski, J.; Sponer, J., Leading RNA tertiary interactions: Structures, energies, and water insertion

of A-minor and P-interactions. A quantum chemical view. *J. Phys. Chem. B* **2007,** *111* (30), 9153-9164.

142. Šponer, J. E.; Leszczynski, J.; Sychrovský, V.; Šponer, J., Sugar edge/sugar edge base pairs in RNA: stabilities and structures from quantum chemical calculations. *J. Phys. Chem. B* **2005,** *109* (39), 18680-18689.

143. Šponer, J. E.; Špačková, N.; Leszczynski, J.; Šponer, J., Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs. *J. Phys. Chem. B* **2005,** *109* (22), 11399-11410.

144. Vokáčová, Z.; Šponer, J.; Šponer, J. E.; Sychrovský, V., Theoretical study of the scalar coupling constants across the noncovalent contacts in RNA base pairs: the cis- and trans-watson-crick/sugar edge base pair family. *J. Phys. Chem. B* **2007,** *111* (36), 10813-10824.

145. Oliva, R.; Cavallo, L.; Tramontano, A., Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res.* **2006,** *34* (3), 865-879.

146. Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A., On the role of Hoogsteen:Hoogsteen interactions in RNA: ab initio investigations of structures and energies. *RNA* **2010,** *16* (5), 942-957.

147. Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D., Quantum chemical studies of structures and binding in noncanonical RNA base pairs: the trans Watson-Crick:Watson-Crick family. *J. Biomol. Struct. Dyn.* **2008,** *25* (6), 709-732.

148. Sharma, P.; Sponer, J. E.; Sponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A., On the role of the cis Hoogsteen:sugar-edge family of base pairs in platforms and triplets-quantum chemical insights into RNA structural biology. *J. Phys. Chem. B* **2010,** *114* (9), 3307-3320.

149. Klein, D. J.; Moore, P. B.; Steitz, T. A., The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **2004,** *340* (1), 141-177.

150. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.

151. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003,** *24* (16), 1999-2012.

152. Cieplak, P.; Caldwell, J.; Kollman, P., Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001,** *22* (10), 1048-1057.

153. Werner, H. J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *Molpro version 2006.1*, a package of ab initio programs, 2006.

154. Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P., Highly accurate CCSD(T) and DFT-SAPTstabilization energies of H-bonded and stacked structures of the uracil dimer. *ChemPhysChem* **2008,** *9* (11), 1636-1644.

155. Hesselmann, A.; Jansen, G.; Schutz, M., Interaction energy contributions of H-bonded and stacked Structures of the AT and GC DNA base pairs from the combined density functional theory and intermolecular perturbation theory approach. *J. Am. Chem. Soc.* **2006,** *128* (36), 11730-11731.

156. Tang, K. T.; Toennies, J. P., Simple theoretical model for van der Waals potential at intermediate distances .1. Spherically symmetric potentials. *J. Chem. Phys.* **1977,** *66* (4), 1496-1506.

157. Haley, T. P.; Graybill, E. R.; Cybulski, S. M., Ab initio calculations of dispersion coefficients for nucleic acid base pairs. *J. Chem. Phys.* **2006,** *124* (20).

158. Applequist, J.; Carl, J. R.; Fung, K. K., Atom dipole interaction model for molecular polarizability - application to polyatomic-molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972,** *94* (9), 2952-2960.

159. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227-249.

160. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984,** *106* (3), 765-784.

161. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986,** *7* (2), 230-252.

162. Orozco, M.; Noy, A.; Perez, A., Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struc. Biol.* **2008,** *18* (2), 185-193.

163. Fulle, S.; Gohlke, H., Molecular recognition of RNA: challenges for modelling interactions and plasticity. *J. Mol. Recognit.* **2010,** *23* (2), 220-231.

164. MacKerell, A. D., Contribution of the intrinsic mechanical energy of the phosphodiester linkage to the relative stability of the A, B-I, and B-II forms of duplex DNA. *J. Phys. Chem. B* **2009,** *113* (10), 3235-3244.

165. Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell, A. D., Intrinsic conformational energetics associated with the glycosyl torsion in DNA: A quantum mechanical study. *Biophys. J.* **2002,** *82* (3), 1554-1569.

166. Banas, P.; Jurecka, P.; Walter, N. G.; Sponer, J.; Otyepka, M., Theoretical studies of RNA catalysis: Hybrid QM/MM methods and their comparison with MD and QM. *Methods* **2009,** *49* (2), 202-216.

167. Deng, N. J.; Cieplak, P., Free energy profile of RNA hairpins: A molecular dynamics simulation study. *Biophys. J.* **2010,** *98* (4), 627-636.

168. Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M., Extensive molecular dynamics simulations showing that canonical G8 and protonated A38H(+) forms are most consistent with crystal structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010,** *114* (19), 6642-6652.

169. Klosterman, P. S.; Shah, S. A.; Steitz, T. A., Crystal structures of two plasmid copy control related RNA duplexes: An 18 base pair duplex at 1.20 A resolution and a 19 base pair duplex at 1.55 A resolution. *Biochemistry* **1999,** *38* (45), 14784-14792.

170. Timsit, Y.; Bombard, S., The 1.3 angstrom resolution structure of the RNA tridecamer r(GCGUUUGAAACGC): Metal ion binding correlates with base unstacking and groove contraction. *RNA* **2007,** *13* (12), 2098-2107.

APPENDIX I

Šponer J., **Zgarbová M.**, Jurečka P., Riley K. E., Šponer J. E., Hobza P.: Reference quantum chemical calculations on RNA base pairs directly involving the 2´-OH group of ribose, *J. Chem. Theory Comput., 5(4), 1166-1179,* **2009**

http://pubs.acs.org/doi/abs/10.1021/ct800547k

APPENDIX II

Zgarbová, M.; Otyepka, M.; Šponer, J.; Hobza, P.; Jurečka, P.: Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys., 12 (35), 10476-10493,* **2010**

http://pubs.rsc.org/en/Content/ArticleLanding/2010/CP/c002656e

APPENDIX III

Banáš, P.; Hollas, D.; **Zgarbová, M.**; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M.: Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput. 6 (12), 3836-3849,* **2010**

http://pubs.acs.org/doi/abs/10.1021/ct100481h

APPENDIX IV

Zgarbová, M.; Jurečka, P.; Banáš, P.; Otyepka, M.; Šponer, J. E.; Leontis, N. B.; Šponer, J.: Noncanonical hydrogen bonding in nucleic acids. Benchmark evaluation of key base – phosphate interactions in folded RNA molecules using quantum-chemical calculations and molecular dynamics simulations. *J. Phys. Chem. A, submitted,* **2011**

APPENDIX V

Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T.E.,III; Jurečka, P.:Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput., accepted,* **2011**

http://pubs.acs.org/doi/abs/10.1021/ct200162x