Palacký University Olomouc

Faculty of Science

2014

Ph.D. Thesis

Physical Chemistry

Structure, Dynamics and Reaction Mechanism of RNA Enzymes

Vojtěch Mlýnský

Department of Physical Chemistry

Supervisor: Prof. RNDr. Michal Otyepka, Ph.D.

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, …… Vojtěch Mlýnský

……

Acknowledgement

It is my pleasure to thank to my supervisor Prof. Michal Otyepka for his guidance, many valuable discussions and strong support. My special thanks go to Assoc. Prof. Pavel Banáš for his helpful comments and motivating suggestions. I would like to also thank to Professors Jiří Šponer, Nils G. Walter and Adrian J. Mulholland and my colleagues at the Department of Physical Chemistry. Finally, many thanks are addressed to my parents, my family and all friends for their kind support and patience.

“The hours that make us happy make us wise.”

John Edward Masefield

Contents

Chapter 1

List of symbols and shortcuts

A#	adenine nucleotide labeled #, e.g., A38, and its N1-protonated form, i.e., $A38H^+$
Ade	N9-methyladenine nucleobase and its N1-protonated form $AdeH^+$
ADMP	atomic center density matrix propagator
AM1	Austin model 1
AM1/d	AM1 with explicit d orbital representation
AM1/d-PhoT	combined AM1/d and MNDO/d model for reactions with phosphates
AMBER	assisted model building and energy refinement
B3LYP	scheme for hybrid HF/DFT by Becke
BLYP	Becke-Lee-Yang-Parr gradient-corrected functional
BSSE	basis set superposition error
C#	cytosine nucleotide labeled #, e.g., C75, and its N3-protonated form, i.e., $C75H^+$
CBS	complete basis set
CBS(T)	MP2 calculations extrapolated for the CBS and corrected for higher-order correlation effects by CCSD(T) in a smaller basis set (typically correlation-consistent Dunning basis set cc-pVDZ)
CCSD(T)	coupled cluster with double and perturbative triple excitations
CHARMM	chemistry at Harvard molecular mechanics with subsequent force field versions, e.g., CHARMM27
CI	configuration integration
CNDO	complete neglect of differential overlap
Cyt	N4-methylcytosine nucleobase and its N3-protonated form $CytH^+$
DFT	density functional theory
DFT-D	DFT with empirical dispersion and its variants, e.g., DFT-D3
DNA	deoxyribonucleic acid
EE+vdW	PBE method including long-range van der Waals (dispersion) interactions modified by exact HF exchange
EVB	empirical valence bond
FEP	free energy perturbation
FES	free energy (hyper)surface
ff#	force field (empirical potential) and its modifications, e.g., *ff*94
G#	guanine nucleotide labeled as #, e.g., G8, and its N1-deprotonated form, i.e., $G8^-$
GlcN6P	glucose-amin-6-phosphate cofactor
glmS	glucose-amin-6-phosphate synthase
Gua	N9-methylguanine nucleobase and its N1-deprotonated form Gua^-
HDV	hepatitis delta virus
HF	Hartree-Fock
IAA	in-line attack angle
IMOMM	integrated molecular-orbital molecular mechanics
IN	intermediate
INDO	intermediate neglect of differential overlap
M06	meta-hybrid exchange functional and its variants, e.g., M06-L
MD	molecular dynamics
MINDO/3	modified INDO version 3
MM	molecular mechanic
MNDO	modified neglect of diatomic overlap

MNDO/d	MNDO with explicit d orbital representation
MP2	Møller-Plesset perturbation theory of the second-order
MPW1K	modified Perdew-Wang 1-parameter model for kinetics
MUE	mean unsigned error
NDDO	neglect of diatomic differential overlap
ONIOM	our n-layered integrated molecular orbital and molecular mechanics
optB88-vdW	Becke88 exchange functional including long-range van der Waals (dispersion) interactions
P	product
PB	Poisson–Boltzmann
PBE	generalized gradient approximations by Perdew–Burke–Ernzerhof
PES	potential energy (hyper)surface
PM3	parameterization 3 of MNDO
PME	particle-mesh Ewald
R	reactant (precursor)
RNA	ribonucleic acid
RNase	ribonuclease, e.g., RNase A
SCC-DFTB	self-consistent charge density functional tight-binding
SCF	self-consistent field
SCS-MP2	spin-component scaled MP2
QM	quantum mechanics (quantum chemistry)
QM/MM	hybrid quantum mechanical/molecular mechanical method
QM/MM–MD	QM/MM approach combined with restrained MD simulations
TS	transition state
U#	uracil nucleotide labeled as #, e.g., U–1
VS	Varkud satellite

Chapter 2

Motivation

Many essential processes in living organisms are managed by enzymes. The most important role of enzymes is to accelerate chemical reactions towards biologically effective time scales.[1] Although the overall concept of enzyme catalysis was successfully proposed more than hundred years ago, the classification of the actual mechanism (i.e. identification of catalytically active groups) and the understanding of other catalytic factors (i.e. dynamical effects) is still limited.[2,3] Even more unanswered questions are related to the ribonucleic acid (RNA) enzymes because they have not been studied as extensively as their protein counterparts.[4-8]

The chemical research of biomacromolecules requires complete description of molecular structure, function and dynamics. The initial information about sequences and active site arrangements are provided by X-ray crystallography, which is perhaps the leading experimental method in this scheme. A wide range of other techniques is available for understanding questions concerning reaction thermodynamics, kinetics, structural dynamics and mechanisms within biomacromolecules. Unfortunately, those questions are generally more difficult to answer experimentally and often only indirectly accessible. Significant contribution towards structural insights can also provide theoretical methods. Molecular dynamics simulations are able to reveal some problematic structural aspects, propose the ionization states of residues and describe the solvent behavior. Quantum mechanical calculations provide information about the reaction chemistry and in combination with simulations (as quantum mechanical/molecular mechanical approach) help to investigate plausible reaction pathways and identify reaction mechanisms of enzymes. The overall applicability and prediction power of theoretical methods is still improving because they are under continuous development and rigorous testing procedures. With the massive increase of computer power during the last two decades, theoretical methods are becoming respectable and widely used tool. It is worth noting that the contribution of theoretical approaches in the contemporary research of biomacromolecules has not gone unnoticed by the Royal Swedish Academy of Sciences. Three theoretical chemists, professors Martin Karplus, Michael Levitt and Arieh Warshel, were jointly awarded the 2013 Nobel Prize in Chemistry for their contribution in development of theoretical methods for modeling of large complex chemical systems and reactions. In overall, theoretical approaches have a potential to shine as a tool for explanation and interpretation of experimental measurements and hypothesis. On the other hand, they should be always interpreted with care, considering their accuracy and limitations.

This thesis is focused on description of the structural stability and the reaction mechanism of RNA enzymes in order to understand the catalytic mechanism by two small self-cleaving ribozymes, i.e., hairpin and HDV ribozymes. A combination of theoretical tools was used, where the conformation behavior was studied by molecular dynamics simulations and several reaction pathways were investigated systematically by combined quantum mechanical/molecular mechanical calculations. These methods seem to be useful tools in order to investigate processes in biomacromolecules and molecular complexes. Our results are discussed against available (both experimental and theoretical) observations and provide complementary data towards deciphering the reaction mechanisms of RNA catalysis.

Chapter 3

Introduction

3.1 The RNA world

Vast majority of chemical reactions in the biological systems are catalyzed by enzymes. Each enzyme acts as a chemical catalyst by enhancing the rate of specific chemical reaction or a set of closely related reactions. Enzymes are structurally complex biomacromolecules, which form multiprotein complexes with molecular weights from about 10,000 to over 1 million.[1] Until thirty years ago all enzymes were believed to be proteins composed by amino acids. The role of RNA seemed to be restricted as the carrier of the genetic information from deoxyribonucleic acid (DNA) to proteins as expressed in the central dogma of the molecular biology. However, a surprising discovery was made in the early 1980s by Cech and Altmann. They isolated RNA molecules with enzyme activity,[9,10] which represent counterparts of enzymes from protein world because they also achieve catalysis. These molecules were later termed as ribozymes and both researchers received the Nobel Prize of Chemistry in 1989 for their breakthrough discovery. RNA emerged as the first macromolecule with both informational and catalytic functions. It was suggested that first living organisms on this planet did not possibly need proteins because ribozymes could be the first self-replicating systems, playing key role in evolution ("RNA World Theory").[11] Ribozymes may be living relics from an ancient ribonucleoprotein world as shown recently on the evolution of ribosome.[12]

First known ribozymes carry out a rather limited range of reactions, involving the phosphoryl transfer, transesterification or hydrolysis reactions.[13] The initial set of reactions was later extended by the most important reaction within cells, i.e., the condensation of an amine with a sp^2-hybridized carbonyl by the rRNA component of the large ribosomal subunit.[14,15] The ribosome active site is thereby composed entirely by RNA and is in fact ribozyme.[16,17] Possible repertoire of ribozyme catalyzed chemical reactions is extending by RNA species select in the laboratory.[13]

Presently discovered ribozymes are divided into two main groups. Larger ribozymes use external nucleophiles (intermolecular nucleophilic attack) located remotely from introns (group I and II self-splicing introns).[13,18] Specific members are ribonuclease P (RNase P) that carries out the processing of tRNA in all living organisms, and peptidyl transferase of the ribosome, catalyzing the condensation of amino acids into polypeptides. Common catalytic strategy of

larger ribozymes involves two-metal-ion mechanism, where one hydrated Mg^{2+} ion is coordinated to the attacking nucleophile and the other to the leaving group.[13,18]

Small self-cleaving ribozymes (also called nucleolytic) are required for the site-specific cleavage of RNA and characterized by attack of 2'-hydroxyl group on the adjacent phosphate (intramolecular nucleophilic attack) or by 5'-hydroxyl group in the reverse reaction. The group of small self-cleaving ribozymes includes five RNA species: hairpin, hammerhead, Varkud satellite (VS), hepatitis delta virus (HDV), and glucosamine-6-phosphate synthase (*glmS*) ribozymes. Their minimum sequences essential for cleavage range typically from ~40 to 200 nucleotides (Figure 1). Hairpin, hammerhead and VS ribozymes are part of species belonging to the circular, self-replicating RNAs and are essential components of the rolling circle replication mechanism.[19] The replication cycle involves the copying of the dominant circular plus strand by a host or viral-coded RNA polymerase to give a minus strand. The long minus strand can self-cleave in vitro to give monomeric products that are subsequently circularized and copied to produce a linear plus strand. The plus strand self-cleaves to monomers, which circularize to produce the circular progeny of plant pathogen (viroid).[20] The HDV self-cleaving RNA motif was obtained from the HDV that infects humans[21,22] and has also been detected in the human genome.[23] VS and *glmS* motifs are found within the bodies of larger transcripts.[19] The *glmS* also acts as a regulatory element (riboswitch) in Gram-positive bacteria by controlling the glucosamine-6-phosphate synthase level.[24]

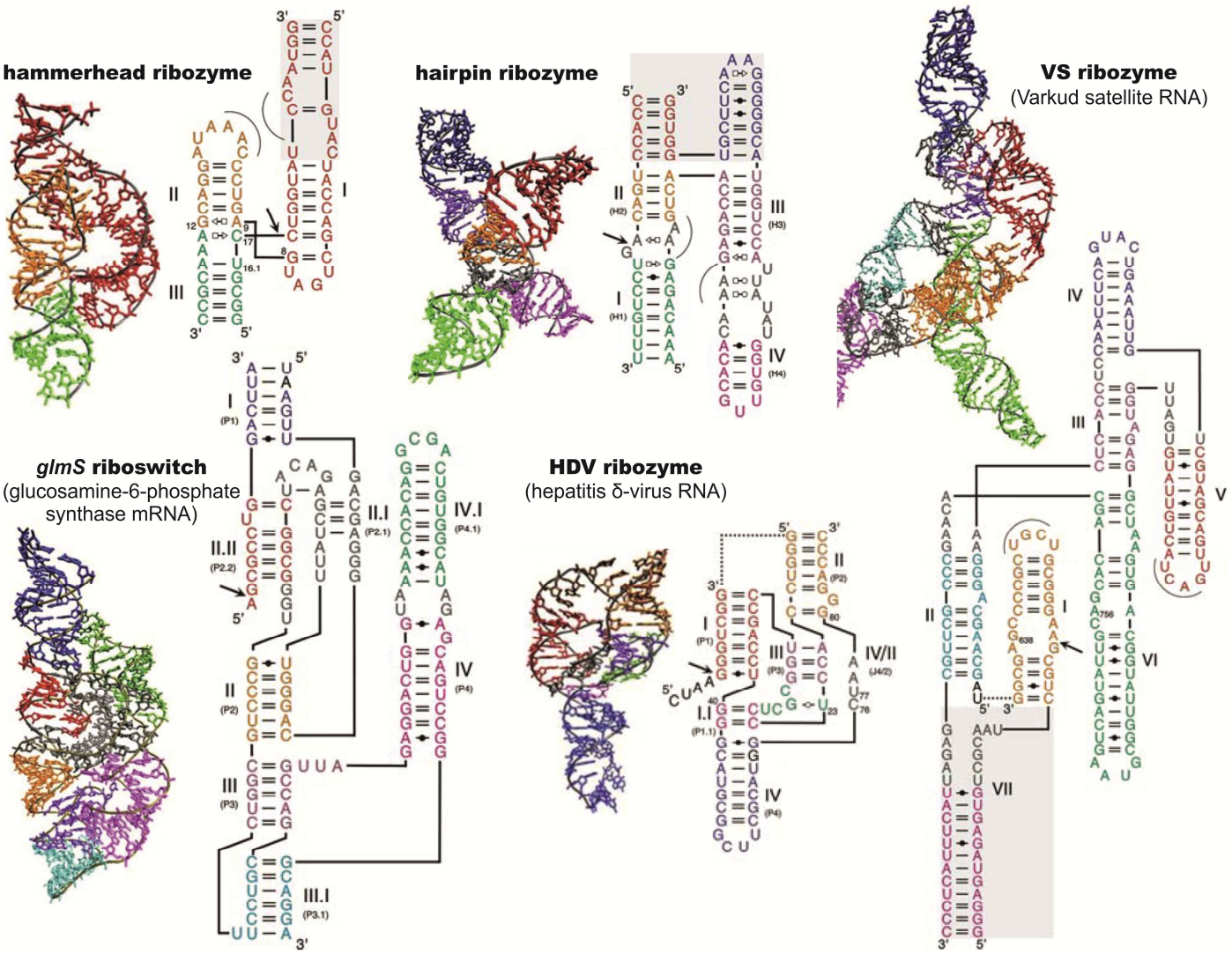

Figure 1: *Three-dimensional structures and appropriate sequences of secondary structures for members of the group of small self-cleaving ribozymes. The colors of helical stems and loops matched between certain secondary and three-dimensional structure. Black arrows indicate the cleavage sites. This figure was composed from panels originally presented in Ref. 19.*

3.1.1 RNA catalysis

The diversity of the side chains is the major difference between proteins and RNA, where four purine and pyrimidine based heterocycles stand against twenty chemically diverse amino acids.[25] Fundamental suggestions how RNA catalyzes the phosphoryl transfer reaction still come from protein enzymes.[4,13] The pancreatic ribonuclease A (RNase A) achieves cleavage of the RNA backbone via formation of a cyclic 2',3'-phosphate equivalent to the nucleolytic ribozymes with an impressive 10^{12} acceleration.[4,26] The catalytic power of RNase A originates from combination of four sources: (i) alignment of the nucleophile, scissile phosphate, and leaving group, (ii) activation of the 2′-OH nucleophile by a general base, (iii) increased affinity for the pentacoordinated phosphorane transition state, and (iv) activation of the 5′-O leaving group by a general acid.[7,26] One of proposed and well-established mechanism of RNase A is a general base/general acid catalysis, where one imidazole side chain of histidine 12 acts as general base in deprotonation of the 2'-OH group and another imidazole side chain of histidine 119 serves as general acid in protonation of leaving 5'-O group. The transition state (TS) with negatively charged phosphate is stabilized by lysine 41.[26] Imidazole heterocyclic compounds having pK_a close to neutrality are well suited for those roles.[13] Unlike amino acids, the unmodified RNA nucleosides are typically neutral at biological pH and their pK_a values are normally at least two units shifted from neutrality.[13,25,27] Those pK_a values help nucleic acids to store genetic information, but may limit potential contributions to catalysis by an electrostatic stabilization and possible proton transfers.[25] Nevertheless, recent reviews conclude that nucleobases are able to shift their pK_a's towards neutrality, participate directly in the general base/general acid chemistry, or stabilize negatively charged TS states electrostatically (Table 1, Figure 2).[7,13,25]

Table 1: *Suggested nucleobase and hydrated Mg^{2+} ion participation in the active sites of small-self-cleaving ribozymes.*

ribozyme	suggested catalytic participants	$\Delta G^{\ddagger}$ (kcal/mol) [a]
hairpin	canonical guanine, N1-deprotonated guanine canonical adenine, N1-protonated adenine	20.9 [b]
hammerhead	canonical guanine, N1-deprotonated guanine, $[Mg(H_2O)_4OH]^+$, $[Mg(H_2O)_5]^{2+}$, $[Mg(H_2O)_5OH]^+$, $[Mg(H_2O)_6]^{2+}$	19.9 [c]
HDV	canonical cytosine, N3-protonated cytosine, $[Mg(H_2O)_4OH]^+$, $[Mg(H_2O)_5]^{2+}$, $[Mg(H_2O)_5OH]^+$, $[Mg(H_2O)_6]^{2+}$	19.9 [d]
glmS	canonical guanine, N1-deprotonated guanine, glucose-amin-6-phosphate cofactor	19.8 [e]
VS	canonical guanine, N1-deprotonated guanine canonical adenine, N1-protonated adenine	20.1 [f]

[a] activation free energy barriers obtained from experimentally measured rate constants (Equation 5.4).
[b] Young *et al.*, *Nucleic Acids Res.*, 1997, 25, 3760.[28]
[c] Hertel *et al.*, *Biochemistry*, 1994, 33, 3374.[29]
[d] Shih and Been, *Biochemistry*, 2000, 39, 9055.[30]
[e] McCarty *et al.*, *Chem. Biol.*, 2005, 12, 1221.[31]
[f] Wilson *et al.*, *EMBO J.*, 2007, 26, 2489.[32]

Enzymes also use metal ions in order to catalyze phosphoryl transfer reactions. Polymerases, nucleases, transposases and phospholipases were suggested for a two metal ion mechanism, where one ion activates the nucleophile and the other acts as a Lewis acid in stabilization of the oxyanion leaving group.[13,33] Initially, it appeared that all ribozymes are special examples of metalloenzymes, where metal ions played direct roles in reaction mechanisms.[34] That assumption was overcame by more detailed studies on hairpin and VS ribozymes, where the catalytic activity was measured without any specific metal ion in the active site.[35-37] In general, metal ions play crucial roles in the folding of RNA molecules and, therefore, in formation of specific active site conformations in ribozymes.[4,7,13] Identified catalytic roles of metal ions involve (i) metal-coordinated water acting in a general base/general acid catalysis, (ii) metal ion forming an inner sphere complex with RNA and (iii) metal ion stabilizing charged TS electrostatically.[4,7,13,38,39]

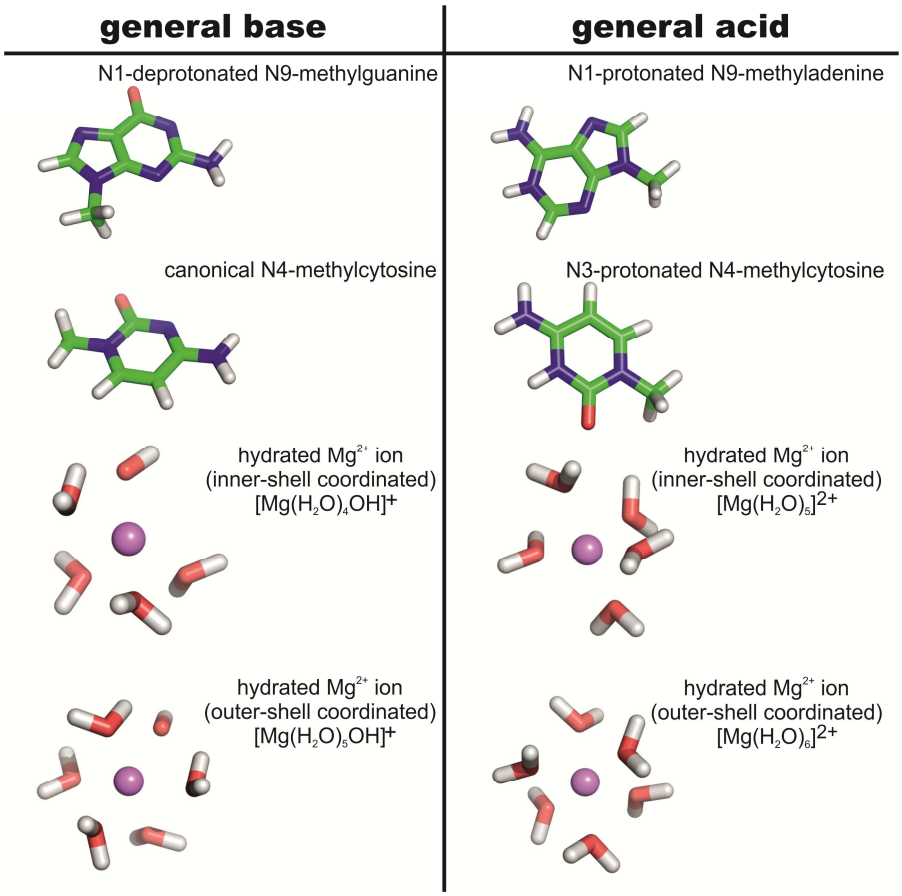

Figure 2: *Models of nucleobases and hydrated Mg^{2+} ions in protonation states suggested for participation in reaction mechanisms of small self-cleavage ribozymes.*

3.1.2 Suggested mechanism of small self-cleaving ribozymes

Although small self-cleaving ribozymes possess various secondary structures (Figure 1), they catalyze the cleavage of the RNA phosphodiester backbone through the same reaction called internal transesterification. The S_N2 type reaction starts with the 2'-hydroxyl nucleophilic attack on the adjacent scissile phosphate, proceeds throw the pentacoordinated phosphorous TS and generate products with 2',3'-cyclic phosphate and 5'-hydroxyl termini (Figure 3).[4,7,19] General catalytic strategies appear to be similar to those identified for RNase A. While the other contributions to the catalysis could not be excluded, the maximal rate enhancements are generally achieved by combination of (i) active site conformation suitably organized for the in-line attack of nucleophile toward the scissile phosphate (i.e. angle between O2', P, and O5' close to 180°), (ii) activated 2'-OH nucleophile by a general base, (iii) TS stabilization (i.e. neutralization of a negative charge located on nonbridging oxygens), and (iv) protonation of O5' by a general acid.[4,7,13,38,39] The RNA catalytic reactions are in principle well understood due to similarities with their protein counterparts. Still, several open questions concern the complete description of catalytic mechanisms in ribozymes with additional unknown factors like the pre-organization of the active sites for catalysis.

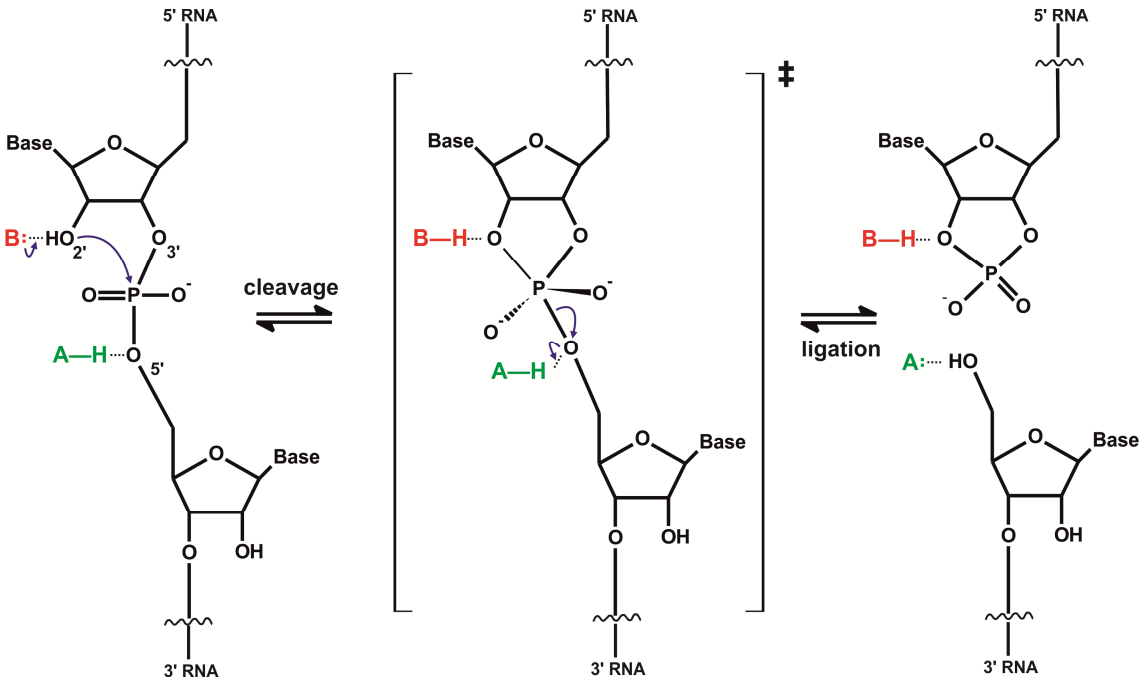

Figure 3: *The mechanism of the internal transesterification reaction catalyzed by small self-cleaving ribozymes. The 2'-OH hydroxyl is acting as nucleophile and attacks the neighboring scissile phosphate with the possible assistance of general base labeled as B. The reaction proceeds through the phosphorane TS and generates the 2'-3'-cyclic phosphate with the 5'-OH hydroxyl termini (potentially catalyzed by general acid labeled as A).*[4]

3.2 Reference reactions

The model of an uncatalyzed reaction is considered as a reference reaction and enables to understand the catalytic effect of enzymes and ribozymes (Figure 4). The catalytic effect is generally defined as the difference in reactivity between the enzyme catalyzed reaction and the corresponding uncatalyzed reaction in water.[40] According to Warshel and co-workers, the catalytic effect involves two components: (i) change of the reaction mechanism (with respect to the reference reaction in water) and (ii) the effect of enzyme environment.[3,41] The chemical effects seem to be well understood because appropriate chemical models help to distinguish between alternative mechanisms of ribozyme reactions.[41,42] Several experimental observations support protonation of the phosphodiester as a viable mechanism for RNA cleavage. Lönnenberg and co-workers demonstrated the protonation of the phosphoryl oxygen of a neutral phosphodiester followed by the nucleophilic attack of the adjacent 2'-OH group.[43-45] The first step of transesterification can be both general base/general acid catalyzed and the rate limiting transition state corresponds to the exocyclic cleavage of phosphorane concerted with the proton transfer to the departing alcoxide.[43-45] Perreault and Anslyn concluded, that two reaction mechanisms (called dianionic and monoanionic) are interchangeable scenarios for RNA cleavage.[46] Both mechanisms differ in protonation state of phosphorane intermediate and are

depended upon the leaving group ability, the pH and the availability of general bases and general acids.[46]

The effects of altering environment (called the true catalytic effect)[40] reproduce the free energy (ΔG) difference between the enzymatic and uncatalyzed reference reaction in water involving the same mechanism as in the enzyme.[3,41] Among those, the most important catalytic effect is associated with the electrostatic contribution, resulting in TS stabilization (or the ground state destabilization) by the surrounding active site.[3,40,41,47]

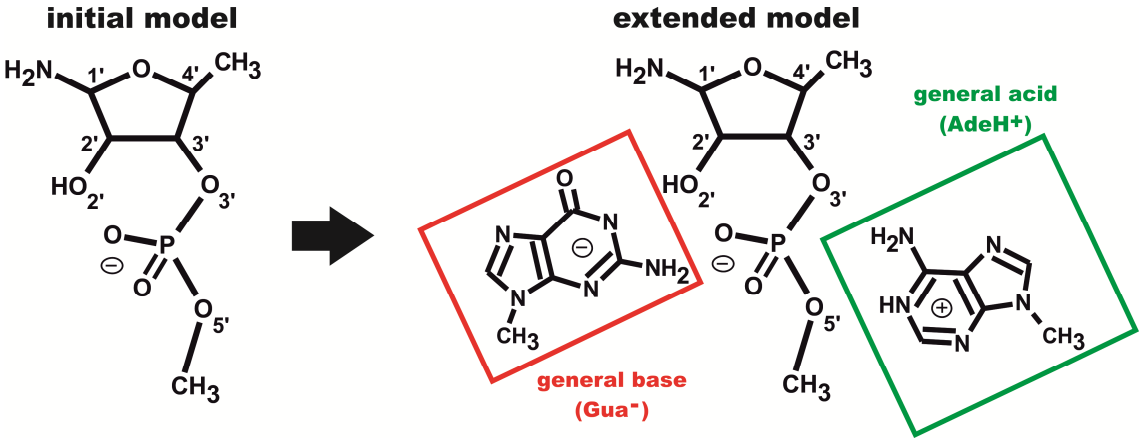

Figure 4: *The scheme of sugar-phosphate backbone model used in the reference reaction. The initial model represents the cleavage of the 3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester (27 atoms). Extended models contain nucleobases and hydrated Mg^{2+} ions (Figure 2), e.g., N1-deprotonated N9-methylguanine (Gua^-) and N1-protonated N9-methyladenine ($AdeH^+$) in the position of general base and general acid, respectively.*

3.3 Small Self-Cleaving ribozymes

3.3.1 Hairpin ribozyme

The hairpin ribozyme adopts naturally a four-way junction secondary structure,[48] but a minimal two-way junction and a junction-less forms of catalytically active ribozyme has also been derived.[19,49] In the minimal, junction-less form, the 61-nucleotide-long ribozyme is composed of two double helical A-RNA stems, which interactions form the catalytic core of the ribozyme (Figure 1).[48,50,51] The scissile phosphate is located between adenine –1 (A–1) and guanine +1 (G+1). Biochemical and structural studies identified two another purine nucleobases, i.e., guanine 8 (G8) and adenine 38 (A38), as main players in the cleavage and ligation.[4,7,13,49,52] The active site is very similar to the catalytic core of RNase A because G8 and A38 are positioned in similar manner to histidines in the active site of RNase A.

The precleavage (containing 2'-O-methyl substitution inhibiting the reaction), product (P) and TS analog crystal structures revealed that G8 is located close to the 2'-OH nucleophile, and

potentially donates a hydrogen bond from N1 nitrogen to the 2'-oxygen of A–1.[48,50,51,53] These observations, together with exogenous nucleobase rescue experiments, led to proposals that G8 electrostatically stabilizes the TS and/or activates the A–1(2'-OH) nucleophile (being deprotonated before the reaction), i.e., acting as a general base.[7,54-57] The experimental data are mostly consistent with the general base mechanism by deprotonated form of G8 ($G8^{-}$) and the recently estimated pK_a of G8 (9.5)[58] in the active site of the hairpin ribozyme fits well with kinetic profiles showing rate constants as a function of pH (Figure 5).[56,59,60] The hybrid quantum mechanical/molecular mechanical (QM/MM) calculations proposed another mechanism, where the A–1(2'-OH) group is deprotonated by the non-bridging oxygen of the scissile phosphate and canonical G8 participates indirectly in the reaction by providing electrostatic stabilization.[61] That proton-shuttling mechanism would likely not lead to any experimentally measurable pH dependence of the rate constant, thus excluding it to be the only catalytic strategy of the hairpin ribozyme.

N1 nitrogen of A38 is placed near the leaving G+1(O5') oxygen in crystal structures of TS analogs.[48,62-64] Exogenous nucleobase rescue experiments suggested, that A38(N1) is involved directly in the cleavage reaction.[65,66] Raman crystallography showed elevated pK_a of A38 (5.5,[67] up from 4.3[68]), further indicating that A38 is protonated ($A38H^{+}$) under physiological pH ~7 prior to cleavage.[67] The proposed roles of protonated $A38H^{+}$ include direct involvement (as general acid) in the reaction mechanism (Figure 5) and/or indirect participation by structural and electrostatic stabilization. Notably, the canonical A38 was also suggested for electrostatic stabilization of the negatively charged phosphorane TS and alignment of reactive groups.[65-67,69,70]

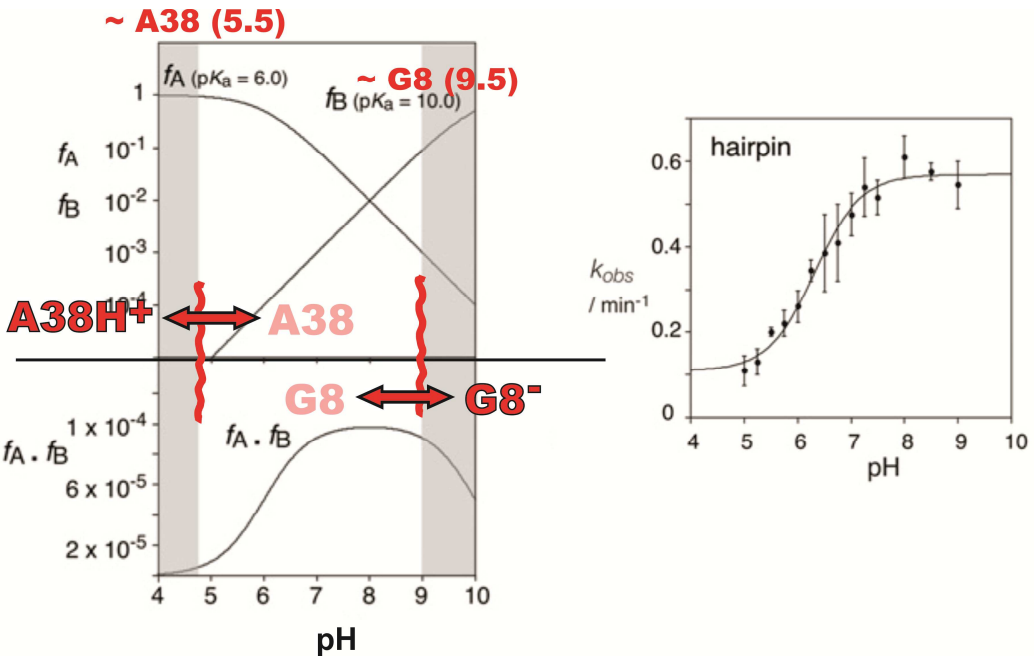

Figure 5: *Calculated pH dependence of the rate of cleavage and the experimentally measured pH dependence of reaction rates of the hairpin ribozyme. The protonated acid (f_A with pK_a = 6.0) and unprotonated base (f_B with pK_a = 10.0) were calculated as functions of pH (plotted on a logarithmic scale, upper left panel).*[60] *Those simulated pH profiles for f_A and f_B are in excellent agreement with estimated pK_a of G8 (9.5)*[58] *and A38 (5.5)*[67] *in the active site of the hairpin ribozyme. The gray areas at the sides are not accessible to experimental investigations. The product ($f_A \cdot f_B$) shows the dependence of cleavage rate with pH (plotted on a linear scale, lower left panel)*[60] *and corresponds with experimental pH dependence of the cleavage reaction for the hairpin ribozyme in its full junction form (panel on the right).*[59] *This figure was composed from panels originally presented in Ref. 60.*

3.3.2 HDV ribozyme

The HDV ribozyme is a self-cleaving RNA motif embedded in genomic and antigenomic RNA strands of HDV. Both genomic and antigenomic types are generated during virus replication and required for self-cleavage of polymeric RNA transcripts.[21,22] Additionally, two slightly different self-catalytic sequences has been synthesized, i.e., the naturally occurring *cis*-acting form, and its *trans*-acting derivate with incorporated external substrate analog inhibitor strands.[71,72] All those molecules typically contain from 85 to 95 nucleotides, depending on the sequence variant and the polarity. First crystal structure of the postcleavage state revealed that the genomic HDV ribozyme adopts a highly ordered structure composed by one stem, one pseudoknot, two stem-loops and three single-stranded junctions (Figure 1).[71] The cleavage site is located between uracil –1 (U–1) and guanine 1 (G1). The side chain cytosine 75 (C75) has been identified to be functionally essential for the catalysis.[5,6,71] HDV ribozyme was thereby the first catalytic RNA motif, where the specific nucleobase has been proposed for participation in the reaction chemistry.[73] Kinetic studies indicated that hydrated Mg^{2+} ion in the active site is directly involved in the cleavage reaction.[74,75]

Two kinetically equivalent reaction mechanisms differing in the role played by C75 have been proposed. The first reaction model was based on crystal structures of the precleavage state

inhibited by a C75/uracil mutation or chelation of Mg^{2+}.[76] C75 was identified as a general base using its N3 nitrogen to activate (deprotonate) the nucleophilic U–1(2'-OH) group. The hydrated Mg^{2+} ion was suggested to act as a general acid to protonate the G1(O5') leaving group in the corresponding manner.[76,77] The subsequent molecular dynamics (MD) simulations starting from these structures together with combined quantum mechanical/molecular mechanical (QM/MM) calculations showed that this mechanism is both chemically and structurally feasible.[73,77,78] The C75 general base mechanism does not however, explain biochemical[79] and more recent crystal structure data.[80]

In the second model, the roles (and the protonation states prior to the cleavage) of C75 and the hydrated Mg^{2+} ion are switched, i.e., the protonated N3 nitrogen of $C75H^+$ acts as a general acid to protonate the leaving group and one of deprotonated water molecules of coordinated Mg^{2+} ion activates the U–1(2'-OH) group. The $C75H^+$ general acid mechanism was proposed by Bevilacqua and co-workers[74] on the basis of crystal structure of the postcleavage state[71] and provides better agreement with biochemical and the latest structural data.[79,80] The recent high-resolution (1.9 Å) crystal structure of the *trans*-acting HDV ribozyme in the precleavage state was solved by molecular replacement from the previous postcleavage structure.[80] The active site was however, crystallographically disordered (especially the U–1 nucleotide) and required additional modeling and superposition with the cleavage site of hammerhead ribozyme.[80] The resulting structural model differs significantly in important elements from the prior precleavage structure,[76] e.g., C75 is bound more tightly to the G1 scissile phosphate, suggesting that it may be protonated prior to cleavage.[80,81] MD simulations are consistent with the role of protonated $C75H^+$ as a general acid because $C75H^+$ appears to assist in the local organization of the active site.[82] Subsequent QM/MM calculations by Hammes-Schiffer and co-workers revealed a concerted reaction mechanism in the presence of an active site Mg^{2+} ion, where combined nucleophilic attack (from U–1(O2') to G1(P)) and proton transfer step (from $C75H^+$ to G1(O5') leaving group) resulted in a single phosphorane-like TS.[83] The monovalent metal ion substitution of the catalytic Mg^{2+} favored a sequential mechanism.[83]

3.3.3 Hammerhead ribozyme

The hammerhead ribozyme was the first nucleolytic ribozyme discovered[84] and the first catalytic RNA with determined crystal structure.[85] The first crystal structures however, contained only a "minimal" three-junction core and provided only negligible information of the hammerhead cleavage site (Figure 1).[52] An innovative crystal structure of a "full-length" hammerhead ribozyme of *Schistosoma mansoni* showed potentially important (and previously neglected) tertiary contact between helices I and II.[86,87] Subsequent crystal structures of precleavage and postcleavage states with these contacts revealed an active site, which is similar

to the active site of the hairpin ribozyme.[19,52] The scissile phosphate is located between two cytosines, numbered 1.1 (C1.1) and 17.1 (C17.1).[7]

Initial cleavage models based on minimal hammerhead structures proposed direct involvement of one or two hydrated Mg^{2+} ions in the reaction mechanism.[88-90] In contrary, recent structural and biochemical data based on the extended full-length version revealed that divalent metal ions are not essential for the catalytic activity.[7,52] Possible roles of Mg^{2+} ions in the hammerhead ribozyme were suggested to be rather minor, i.e., participation during folding and charge (electrostatic) stabilization.[52,90] Structural data instead identified two guanine nucleobases (G8 and G12) in the active site, where G12 was localized within hydrogen bond distance of nucleophilic C17.1(2'-OH) group and 2'-hydroxyl of G8 having hydrogen bond contact to the C1.1(O5') leaving group. The combined general base/general acid mechanism involving both G12 and G8 nucleobases is the most consistent model with kinetic profiles[91] and mechanistic experiments.[92,93] That mechanism may still be accompanied with an active participation of hydrated Mg^{2+} ion forming innershell or outershell coordination with nonbridging oxygen of the scissile phosphate.[94] Recent semiempirical QM/MM calculations showed that the presence of Mg^{2+} in the active site affects geometries, provides electrostatic stabilization, and decreases energy barriers along the reaction pathway.[95]

3.3.4 *glmS* ribozyme

The *glmS* ribozyme was identified as an catalytically active part of the 5'-untranslated region of mRNA using a small-molecule metabolite cofactor (glucose-amin-6-phosphate, GlcN6P) as an allosteric activator or an coenzyme.[24] The overall structure contains ~120 nucleotides and is composed of eight small stems (Figure 1).[96,97] The scissile phosphate is located between adenine –1 (A–1) and guanine 1 (G1). Another guanine nucleobase labeled as G40 (or G33 depending on the obtained crystal structure)[96,97] is located within hydrogen bond distance from the A–1(2'-OH) group. Structural and biochemical data showed that the presence of both G40 and GlcN6P cofactor simultaneously is critical for the catalytic activity.[7,52] Mutation of G40 to any other nucleobase in the presence of GlcN6P resulted in a significant decrease of the cleavage rate.[97,98] The presence of GlcN6P in the active site directly activates the self-cleavage of *glmS* ribozyme,[31] although it does not lead to any detectable conformational rearrangements.[8,24] Specific interactions between the phosphate moiety of GlcN6P and G1 nucleobase appears to be crucial for proper cofactor positioning in the binding site.[99]

A number of reaction mechanisms were suggested for *glmS* self-cleavage. The first scenario is similar to other small self-cleaving ribozymes and involves the N1-deprotonated form of G40 ($G40^{-}$). $G40^{-}$ was proposed to act as general base accepting the proton from the A–1(2′-OH) nucleophile. Subsequent protonation of G1(O5') leaving group would be facilitated by the ammonium form of GlcN6P (acting as general acid).[7,96] In the second mechanism suggested by

Ferré-D'Amaré and co-workers[98] proton donors and acceptors are GlcN6P cofactor and water molecules. The amino form of GlcN6P accepts the proton from the A–1(2′-OH) nucleophile via two tightly bound water molecules and consequently, as protonated ammonium form, transfers the proton to the leaving G1(O5′) oxygen.[98] Recent MD simulations proposed an alternative mechanism, where the A–1(2'-OH) group is deprotonated by one of nonbridging oxygens of the scissile phosphate.[100] The canonical G40 (N1-protonated) participates in electrostatic stabilization of the negatively charged phosphorane TS state and the ammonium form of GlcN6P donates the proton to the G1(O5') group.[100] Similar results were observed by fluorescence measurements using nucleobase mutation of the active site guanine (G33), where the canonical form of guanine appears to be the major protonation state.[101] Latest kinetic experiments are consistent with the model, where GlcN6P directly participates in proton transfers during the self-cleavage reaction.[102]

3.3.5 VS ribozyme

The complete crystal structure of the VS ribozyme has not been published yet. The secondary structure and initial information about general fold were obtained from biophysical and small-angle X-ray scattering studies.[103,104] The VS ribozyme is the largest known small self-cleaving ribozyme with ~140 nucleotides forming seven helical segments connected by three three-way helical junctions (Figure 1).[19] The cleavage site is located between guanine (G620) and adenine (A621) in the internal loop and displays topologically similar arrangement with the active site of the hairpin ribozyme.[60] Furthermore, two other critical purines (guanine (G638) and adenine (A756)) were identified by nucleotide substitution and implicated for the chemical reaction.[32,103,105,106] Similar to the hairpin ribozyme, both G638 and A756 nucleobases are considered as the key players in the catalysis.[60]

Possible roles of G638 and A756 may lie in electrostatic stabilization, but experimental data are mostly consistent with combined general base/general acid mechanism. The measured pH dependence of the rate of substrate cleavage in the presence of Mg^{2+} ions revealed bell-shaped profile, which corresponded to a double ionization model[56] with pK_a values of ~5.2 and ~8.4.[32] Nucleotide substitution experiments combined with 5'-phosphorothiolate analog data identified a correlation between the pK_a of the nucleobase at position 638 and the observed pK_a of the cleavage reaction.[13,107] Thus, the available experimental data are mostly consistent with the mechanism, where deprotonated form of G638 acts as the general base to activate the 2'-OH nucleophile and protonated form of A756 plays the role of general acid by transferring its proton to the O5'-leaving group.[60]

Chapter 4

Theory

In this chapter, a brief theoretical background about quantum mechanical and molecular mechanical approaches is presented. Considerable attention is given to those theoretical methods, which were further used for computations of small model systems as well as for investigations in ribozymes. At the last part, three computational techniques, i.e. quantum mechanical calculations, classical MD simulations, and combined QM/MM calculations, are described with the emphasis on basic principles, strengths and potential limitations.

4.1 Quantum mechanical approach

4.1.1 *Ab initio* methods

The high-level quantum mechanical (QM) methods involve solving of the Schrödinger equation by using various physically evincible approximations. Their accuracy can be assessed at a certain level of quality without performing highest level calculations because their results converge systematically towards the correct solution.[108] This behavior contrasts with molecular mechanics (empirical force fields) calculations, which can fail in an unpredictable manner by moving away from systems that were used for parameterization.[108,109] The starting point of quantum chemistry is the Schrödinger equation, which time-independent form is:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + \upsilon\right\}\Psi(\hat{\mathrm{r}}) = E\Psi(\hat{\mathrm{r}}), \tag{4.1}$$

where E is energy of single particle of mass m, which is moving through space (given by a position vector $\hat{\mathrm{r}} = x\vec{\mathrm{i}} + y\vec{\mathrm{j}} + z\vec{\mathrm{k}}$) and under the influence of an external field υ (which might be the electrostatic potential due to the nuclei of molecule).[110] Ψ is the wavefunction that characterizes the particle's motion and ∇ is a vector differential operator defined by:

$$\nabla = \vec{\mathrm{i}}\frac{\partial}{\partial x} + \vec{\mathrm{j}}\frac{\partial}{\partial y} + \vec{\mathrm{k}}\frac{\partial}{\partial z}\ , \tag{4.2}$$

where $(\vec{\mathrm{i}}, \vec{\mathrm{j}}, \vec{\mathrm{k}})$ are unit vectors. The left side of Equation 4.1 can be abbreviated by $\hat{H}\Psi$, where $\hat{H}$ is the Hamiltonian operator, reducing the formula to $\hat{H}\Psi = E\Psi$. The Schrödinger equation is thereby a partial differential eigenvalue equation, where the operator acts on a function (eigenfunction) and returns the function multiplied by a scalar (eigenvalue).[110] Precise solutions are however, possible for only a few systems (e.g., hydrogen atom). Any solution for

polyelectronic atoms or molecules, i.e., systems involving three and more particles, can only be approximations ("three-body problem").[110] Another complication for polyelectronic atoms is the electron spin, which is also incorporated into solutions of the Schrödinger equation via spin orbitals. The wavefunction is required to either remain unchanged, when two electrons are exchanged, or it must change sign ("the antisymmetry principle"). The initial approximation assumes that the motion of the electrons is decoupled from the motion of the nuclei ("the Born-Oppenheimer approximation"). The electrons having less significant mass than nuclei (proton is 1836 times heavier than electron) can adjust almost instantaneously to any changes in the position of the nuclei, which allows to concentrate on the electronic motions under the fixed nuclei position(s).[110]

All QM methods that solve Schrödinger equation are called *ab initio* (from first principles) QM methods. The most complete *ab initio* QM method is full configuration interaction (full-CI) that represents exact solution of non-relativistic Schrödinger equation within a finite basis set of atomic orbitals. It is however, so demanding that it cannot be used for chemically interesting systems.[111] A number of *ab initio* QM methods are based on the Hartree–Fock (HF) approximation, which considers correlated motions of electrons. The initial set of electronic solutions is gradually refined (self-consistent field, SCF) and corresponds to lower energies until the point (unchanged results for all electrons) is reached.[110] The missing term, i.e. electron correlation energy, creates the attractive part of the van der Waals interactions (dispersion term).[108] Fortunately, there are various affordable techniques to include large amounts of the correlation energy at a reasonable computational cost (post-HF methods).[108] The most widely used are Møller–Pleset second-order perturbation (MP2) theory and coupled-clusters with single, double and perturbative triple excitations (CCSD(T)). MP2 is not fully balanced for intermolecular forces, but that problem can be partially eliminated by spin-component scaled MP2 method (SCS-MP2).[112] CCSD(T) is the reference method, when applied with large basis set of atomic orbitals ("the golden standard of quantum chemistry"). However, CCSD(T) is still applicable only to the smallest systems (~40 atoms) due to huge computer demands. Other methods that aim to provide an exact solution of the Schrödinger equation, e.g., variation and diffusion Monte Carlo, have still enormous computational demands.[108,111]

4.1.2 Density functional theory

The *ab initio* QM methods are presently often replaced by various density functional theory (DFT) approaches relying on the electron density rather than on wave function. DFT offers a wide spectrum of methods of diverse quality, applicability, and computer efficiency.[113] The major DFT benefit is that it is offering significantly faster results than *ab initio* methods having comparable accuracy.[114-116] Commonly used DFT functionals (e.g., BLYP, B3LYP, PBE)[117-120] provide decent estimations of molecular geometries and chemical reactions, but they fail in

description of non-covalent interactions due to the inability to include the London dispersion energy.[108,111] Another DFT functionals, e.g., M06-2X, M06-L, M06-HF,[121] optB88-vdW[122] and EE+vdW[123], were optimized or at least improved for non-covalent interactions and provide promising results.[124,125] Recently developed functionals can also contain empirical dispersion terms (marked by suffix -D, i.e., DFT-D)[114-116]. DFT-D methods are parameterized against benchmark *ab initio* QM data and practical results indicate that the latest improved variants (such as DFT-D3)[126] seem to be the best option for QM studies on biomolecular systems.[111] Despite all the efforts, there is no single DFT functional accurate for all applications comparable to the best *ab initio* QM methods. Various DFT functionals still achieve different accuracy for various classes of systems and chemical problems.[111]

4.1.3 Semiempirical methods

Approximate molecular orbital theories (semiempirical methods) have been developed because *ab initio* (or even DFT) QM approaches are expensive in terms of computer resources. Semiempirical methods have affordable computational cost and are thereby useful for investigation of large systems (e.g., biomacromolecules) and/or complicated reaction schemes (e.g., construction of complete free energy surfaces).[108,110] However, semiempirical methods are heavily dependent on parameterization. Semiempirical schemes explicitly consider only valence electrons of the system and the core electrons are subsumed into the nuclear core.[110] The most demanding part in comparison with the *ab initio* HF SCF calculation, i.e., calculation and manipulation with integrals, is neglected or approximated.[110] The most known semiempirical methods are those developed in groups of Pople and Devar. The initial methods developed in Pople's group (i.e., the complete neglect of differential overlap model (CNDO), the intermediate neglect of differential overlap model (INDO), and the neglect of diatomic differential overlap model (NDDO)) in 1970s were further modified in Dewar's group (the modified INDO method (MINDO/3), the modified neglect of diatomic overlap method (MNDO), the Austin model 1 (AM1), and the Parameterization 3 of MNDO method (PM3)) in 1980s and 1990s.[110] The later are still being reparameterized for a given (class of) reactions and used in various computational studies, e.g., the combined AM1/d (AM1 method with an explicit d orbital representation) and MNDO/d model with new parameters for H, O and P atoms (AM1/d-PhoT)[127] parameterized for reactions with phosphates. Another frequently used semiempirical method is based on a second-order expansion of the Kohn-Sham total energy in DFT, known as the Self-Consistent Charge version of Tight Binding Hamiltonian (SCC-DFTB).[128,129] The SCC-DFTB method incorporates only two-center Hamiltonian integrals in the calculations, whereas the usual short-range repulsive interaction appearing in tight binding models is fitted to self-consistent data derived in the local density approximation.[129]

4.2 Molecular mechanical approach

Molecular mechanics (MM, also called empirical potential or force field) is the most approximate computational method introduced in this thesis. MM is based on classical Newtonian physics which deals with molecules as classical objects.[130] MM force fields are basically interpreted in a terms of relatively simple four-component picture. Molecular potential energy contains bonded terms (typically bond stretching, angle bending, torsions and out-of-plane deformations or improper torsions) and interactions between non-bonded parts of the system.[110] The non-bonded term is usually modeled using a Coulomb potential for electrostatic interactions and a Lennard-Jones potential for van der Waals interactions. The example of a functional form for such force field is:

$$E_{\mathrm{MM}} = \sum_{\text{bonds}} k_d (d - d_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi + \delta)] + \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \left(\varepsilon_{\mathrm{AB}} \left[\left(\frac{\sigma_{\mathrm{AB}}}{r_{\mathrm{AB}}} \right)^{12} - \left(\frac{\sigma_{\mathrm{AB}}}{r_{\mathrm{AB}}} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_{\mathrm{A}} q_{\mathrm{B}}}{r_{\mathrm{AB}}} \right) , \quad (4.3)$$

where the symbols d, θ, and ϕ indicate bond lengths, angles, and torsions, respectively (d_0 and θ_0 are the corresponding equilibrium values); n and δ are the torsion multiplicity and phase, respectively. The bonded force constants are k_d, k_θ, and k_ϕ; r_{AB} is the nonbonded distance between atoms A and B, and $\varepsilon_{\mathrm{AB}}$ and σ_{AB} are the Lennard–Jones parameters; q_{A}, q_{B} are atomic partial charges; and ε_0 is the vacuum permittivity (dielectric constant).[110] The force field is pair-additive, i.e., it neglects polarization of atoms (non-additive effects) in the electric field from the surroundings.[108] More physically accurate polarization force fields are possible and under development, e.g., polarized CHARMM (the Chemistry At Harvard Molecular Mechanics)[131] force field based on the classical Drude oscillator model,[132,133] but no such advanced force field has been successfully parameterized for nucleic acids.[108,111] Despite sophisticated parameterization, force fields remain approximate due to fundamental limitation using those simple analytic atomistic functions (Equation 4.3) in relating structure and potential energy.[134,135] In addition, force fields are primarily parameterized and tested for description of canonical structures and they may be less accurate for non-canonical elements (e.g., for residues within RNA catalytic centers).[136]

4.2.1 Force Fields

The transferability of parameters from one molecule to another is a key characteristic of any force field. There are two force fields that have been extensively tested on RNA and DNA

systems; AMBER (the Assisted Model Building and Energy Refinement)[137] and CHARMM[131], which share similar functional form but differ in parameterization. Both AMBER and CHARMM offer high-quality protein force fields for consistent description of nucleic acid/protein complexes.[135] The AMBER force field family includes the innovative Cornell *et al.* parm94 (*ff*94), a second-generation force field containing particle-mesh Ewald (PME) treatment[138] of long-range electrostatics developed in the 1990s by Kollman and co-workers,[134] and its subsequent modifications, e.g., *ff*99[139], *ff*99bsc0[140], *ff*99bsc0χ_{OL3}[141]. The actual success of force fields starting from *ff*94 is closely related to the compensation of errors, e.g., the atomic point charges of *ff*94 force field were calculated by the HF method in the 6-31G* basis set (taken from the electrostatic potential), where the molecular dipole moments are overestimated. On the other hand, the overestimation in molecular dipole moments is able to indirectly compensate the missing polarization in condensed matter simulations.[108] The *ff*99 differs from the original *ff*94 only marginally, but *ff*99bsc0 contains first of significant (and necessary) improvements. MD simulations of B-DNA on time scales reaching ~10 ns with the *ff*94–99 variants revealed irreversible α/γ backbone substates that resulted in a severe structural deformation in B-DNA simulations.[140,142] Even longer (>100 ns) MD simulations revealed formation of spurious ladder-like RNA structures with a shift of the glycosidic torsion χ toward the region typical for B form (high-anti, ~270°).[141,143,144] Such a behavior was eliminated by an inclusion of χ_{OL3} parameters[141] and the *ff*99bsc0χ_{OL3} version of AMBER Cornell *et al.* force field (internally abbreviated as *ff*10 or *ff*12 in the recent AMBER code versions) is presently the most widely tested RNA force field. The latest reparameterized *ff*99bsc0$\chi_{OL3}\varepsilon\zeta_{OL1}$ version[145] was suggested primarily for B-DNA, where the updated $\varepsilon\zeta_{OL1}$ torsion parameters improve BI/BII populations and helical twist.[145]

CHARMM27 is another type of force field successfully designed for nucleic acids.[146,147] Longer MD simulations (~100 ns) showed that CHARMM27 describes the canonical B-DNA structure relatively well with only certain problems with groove widths and helical twists.[148] First systematic tests of CHARMM27 on RNA predicted however, very fast base pair breathing of A-RNA, where the strength of base pair interactions appeared to be underestimated.[149,150] Reparameterized 2'-OH dihedrals in CHARMM36 force field[151] revealed similar problems (albeit less frequent) with base pairs breaking and/or formation of noncanonical Watson-Crick pairs during A-RNA simulations, resulting in lower structural stability of A-RNA duplexes.[152]

4.3 Computational Techniques

4.3.1 QM calculations

The best *ab initio* QM calculations cannot directly be used to study enzyme reactions because they are computationally demanding. They are mostly applied on complete isolated

systems consisting ~30-50 atoms in the gas phase (in vacuum) and serve as benchmarks (reference reactions) in a similar manner as experimental data.[73,108,111] The high-level *ab initio* QM calculations in vacuum (e.g. MP2, CCSD(T)) of the reference reaction enable to assess the performance of less accurate QM method (typically DFT) used in subsequent investigations of reactions in biomacromolecules. The presence of solvent is significant especially for charged systems, e.g., nucleic acid models with negatively charged phosphates, and the inclusion of solvent screening effects enables to perform the free energy calculations. The energy difference between QM calculations in the gas phase and QM calculations in the solvent could be used for extrapolation of the free energy corrections, which correspond to reactant (R), intermediate (IN), TS, and P states along identified reaction coordinate in the enzyme.[73]

The easiest way is to perform approximate QM continuum solvent model calculations that are analogous to classical Poisson–Boltzmann (PB) theory used in molecular modeling.[111] However, the continuum solvent QM computations are not as accurate as parameter-free QM gas phase calculations. Continuum solvent methods neglect explicit water interactions and calculated hydration energies are dependent on parameterization. The key parameters appear to be the atomic radii, which are parameterized for a definition of the molecular cavity. Molecular cavities separate molecular interior and the surrounding dielectric continuum, which is mimicking the solvent response.[153] The inclusion of (unphysical) atomic radii indicates that continuum solvent models may require various optimal parameters for different types of systems.[108,111] Calculated free energies using continuum solvent models tends to be unrealistic in absolute values, but the relative trends are supposed to be correct.[111]

4.3.2 Classical MD simulations

Molecular dynamics (MD) calculates the real dynamics of the system, from which time averages of properties are calculated.[110,135] State of the system at any future time can be predicted from its current state (MD is deterministic method). The studied molecule initially assumes certain *xyz* geometry (starting structure), which has generally major impact on the subsequent simulation. Thus, high-resolution experimental crystal structures are required for reliable characterization of biomacromolecular systems.[111,135] Sets of atomic positions are solved numerically using differential equations emerging from Newton's second law of motion:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \ , \tag{4.4}$$

where a particle of mass m_i moves along one coordinate (x_i) under the force F_{x_i} in that direction.[110] The modified Verlet algorithm ("Velocity-Verlet") is applied in order to propagate the position ($\vec{\mathrm{x}}$), velocity vectors ($\vec{\mathrm{v}}$, the first derivative of the positions in time) and accelerations ($\vec{\mathrm{a}}$, the second derivative) in a coupled mode:[110]

$$\vec{\mathrm{x}}(t+\delta t) = \vec{\mathrm{x}}(t) + \vec{\mathrm{v}}(t)\delta t + \frac{1}{2}\vec{\mathrm{a}}(t)\delta t^2 \tag{4.5}$$

$$\vec{\mathrm{v}}(t+\delta t) = \vec{\mathrm{v}}(t) + \frac{1}{2}[\vec{\mathrm{a}}(t) + \vec{\mathrm{a}}(t+\delta t)]\delta t \tag{4.6}$$

The equations of motion are integrated by breaking the calculation into a series of very short time steps (typically 1.0 fs). The forces on the atom are computed at each step and combined with the current positions and velocities to generate new positions and velocities in the next step. The atoms are then moved to new positions and updated set of forces is computed. MD trajectory is generated describing changes of the dynamics variables during time.[110] Typical MD simulations are presently run on hundreds of nanoseconds or even microsecond time scales for systems containing ~10^4 atoms (the 1 μs simulation using a 1 fs time step requires 10^9 steps). Although the overall simulation length is constantly improving, accessible time scales are still inadequate for the majority of biological processes, e.g., protein folding, which occurs in the real world on time scales of seconds. Another limitation is the sampling of conformational space. Standard unrestrained simulations reveal dynamics that a real single molecule may undergo from the starting structure in the finite simulation timescale (i.e., the problem of local conformational traps). There are other techniques to enhance sampling, e.g., metadynamics, replica exchange MD, steered MD, but all of them include additional approximations and bias.[111,135]

4.3.3 Combined QM/MM calculations

The combined QM/MM methods are using a concept introduced by Warshel and Levitt in 1976.[154] The studied system is divided typically into two parts, which are calculated by different methodological approaches and both parts are allowed to communicate with each other. The first part of the system, the QM core, describes the region where the chemical reaction takes place. It is calculated at the QM (e.g., *ab initio*, DFT, semiempirical) level, which is able to perform SCF treatment in the presence of external point charge field. The second (surrounding) part that generates steric and polarization constraints on the core, is calculated at a less rigorous level (typically an empirical force field).[108,155] Any system can be, in principle, divided into more than two parts, where the largest part is treated by the computationally cheapest method (i.e., the most approximate) and the smallest part by an accurate and expensive high-level QM method.[108]

There are two QM/MM schemes (additive and subtractive), which are equivalent and differ in some technical details.[108,155] The total energy of the system in the additive scheme (Equation 4.7) is represented as a sum of the QM energy of the QM core, E_{QM}, the MM energy of the MM region (without the QM core), E_{MM}, and the coupling term representing the interaction between both regions, $E_{\mathrm{QM/MM}}$. The subtractive scheme (Equation 4.8) divides the system into layers and

subtracts double counted energies of the smaller layer. $E_{\text{real}}^{\text{MM}}$ denotes the whole real system energy with both the QM and MM regions, which is calculated by the lower level (MM) method. $E_{\text{model}}^{\text{QM}}$ represents energy of the QM core calculated at a higher level QM method and $E_{\text{model}}^{\text{MM}}$ is the energy of the same part computed by the lower level (MM) method (Equation 4.8).[108] The subtractive scheme was introduced as integrated molecular-orbital molecular mechanics (IMOMM)[156] but is rather termed as ONIOM scheme (our n-layered integrated molecular orbital and molecular mechanics).[157]

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \tag{4.7}$$

$$E_{\text{tot}} = E_{\text{real}}^{\text{MM}} + E_{\text{model}}^{\text{QM}} - E_{\text{model}}^{\text{MM}} \tag{4.8}$$

The key factors, which significantly affect accuracy of the QM/MM calculation, are boundaries and couplings between the QM and MM regions.[108,155] The number of atoms in the QM core and its boundaries has to be chosen carefully. The most challenging cases are found in studies of biomacromolecules, where the border between regions is situated between covalently bounded atoms. The cutting of covalent bounds is usually treated by: (i) a link atom saturating the cut bond (mostly hydrogen atom), or (ii) a localized orbital at the boundary between the regions.[108,155] Both strategies significantly perturb the studied system and affect the convergence of the calculation. QM/MM boundary should be placed as far from the chemically active region as is feasible in terms of computational effort (at least three bonds away from the boundary).[108,155,158]

The QM/MM coupling describes the non-bonded interactions between the QM and MM regions, where the crucial component is the description of the electrostatic part.[108] The non-bonded interactions can be treated at different level of mutual polarization, classified as (i) mechanical embedding, (ii) electronic embedding and (iii) polarized embedding. Mechanical embedding is the simplest and the most approximate QM/MM coupling, where the electrostatic interactions between the QM and MM regions are calculated at the MM level. The polarization of the QM wavefunction is completely neglected because the QM calculation of the inner part is executed without MM the surrounding.[108,155] Electronic embedding performs QM calculation in the presence of MM charged model by incorporating MM point charges as one-electron terms in the QM Hamiltonian.[108,155] The electronic structure of the QM region can adapt to modified charged environment and is automatically polarized by it. The polarization of the QM wavefunction represents a crucial improvement, but the simplified MM charges placed in proximity to the QM electron density can result in overpolarization. This bias can cause serious inaccuracies and possible extension of the QM region should be considered in such case.[108,155] Another way to avoid the inadequate polarization of the QM wavefunction is the three-layer ONIOM scheme developed by Morokuma and co-workers,[158] where the middle layer (described at a medium level of theory) keeps the problematic QM/MM boundary far from any bond

involved in the reaction. The polarized embedding involves flexible MM charge models polarized by the QM charge distribution. Since the development of polarizable protein force fields is still in progress, polarizable QM/MM calculations were essentially restricted to explicit solvation.[155] The robust tests of polarized QM/MM schemes are still lacking because such implementations appear to be less critical for biomacromolecules and at the same time significantly increase computational costs.[108,155]

Apart from boundaries and couplings, another limitation of a typical QM/MM scheme arises from conformational sampling. The most accurate QM/MM calculations (high-level *ab initio* QM method, large QM core, presence of explicit solvent and ions) are only possible with a very limited sampling mostly at the level of simple energy minimization. Those studies are generally combined with other method, for example, classical MD, which enables to achieve at least limiting sampling by generating a set of different starting geometries.[108,155] The QM/MM schemes with semiempirical methods provide a true sampling (combined with restrained MD simulations as QM/MM–MD approach) involving dynamics on a hundred of picoseconds time scales. Empirical valence bond (EVB) methods are able to provide a robust sampling on a nanosecond timescales. The QM/MM–MD schemes have been initially developed for high-level QM methods but still remain computationally demanding with DFT and *ab initio* approaches. The QM/MM–MD schemes are generally divided into two groups according to the MD protocol. Firstly, QM energy and forces come from converged SCF calculations, known as Born-Oppenheimer MD, and such schemes are frequently used with semiempirical methods. The alternative protocol involves on-the-fly dynamics, where the wavefunction is propagated in the each step directly by extended Lagrangian equation.[108] The example of such model is the Atom-centered Density Matrix Propagation (ADMP), developed by Schlegel and Iyengar.[159] Additional techniques, e.g., free energy perturbation (FEP) schemes, provide alternative ways to increase the conformational sampling. The free energies are calculated along predefined reaction coordinate, where the MM degrees of freedom are sampled for the specific (fixed) position of QM atoms.[155]

Chapter 5

Results

This chapter of the Ph.D. thesis is divided into three parts. Firstly, essential corrections for comparison of reaction barriers with experiments are mentioned. The second part describes QM calculations of the uncatalyzed reaction in water providing (i) comparison of the performance of QM methods used in our QM/MM calculations and (ii) estimation of the free energy corrections corresponding to the states along the reaction coordinate of the specific reaction mechanism in ribozymes. The main results are presented in the third part and reveal our studies of RNA self-cleavage reactions within the hairpin and HDV ribozymes using classical MD simulations and combined QM/MM calculations.

5.1 Comparison of calculated barriers with experiments

5.1.1 Free energy corrections

Typical QM/MM schemes evaluate the potential energy surface (PES) and provide electronic energy changes. The energies are determined either for a set of fixed geometries or with the use of gradient geometry optimization (energy minimization). Energy barriers located on the PES are not directly comparable to experimental data. Corrections for free energies, which contain zero point vibration energy (*ZPVE*), enthalpy correction for finite temperature ($H_{T=0K}$) and entropy contributions ($T\Delta S$), need to be included.

$$\Delta G = \Delta E + \Delta ZPVE + H_{T=0K} - T\Delta S \tag{5.1}$$

There are at least three ways to estimate the free energies and each approach has certain benefits and drawbacks. The first (and probably the simplest) option involves an extrapolation of free energy corrections from the model of reference reaction, which shares the same reaction mechanism as the corresponding QM/MM pathway, by using the harmonic approximation (ideal gas, rigid rotor, and harmonic oscillator) in the canonic ensemble.[73] This approach usually requires QM calculations within implicit solvent, which affects the accuracy of typically precise gas phase QM calculation (see Methods). The second way involves a direct estimation of free energies by either the construction of an entire free energy surface (FES) of the enzymatic reaction or by the sampling of the FES along the reaction pathway. FES computations are however, still too expensive for high quality (DFT, *ab initio*) QM methods.[108,160] The third possibility is to calculate the minimum free energy pathway (FEP) along the predefined reaction

coordinate, where the MM degrees of freedom are sampled for the specific (fixed) position of QM atoms. Those calculations would however, either neglect the explicit polarization of QM region by MM point charges, which is assumed to be crucial for polyanionic RNA environment,[108,111] or require recalculations of the QM part every integration step resulting in similar computational demands as the second option.

5.1.2 Corrections for rare protonation states

The ribozyme active sites include titrable residues, nucleobases and/or specific water molecules coordinated to divalent Mg^{2+} ions. The major protonation states under physiological conditions (pH ~7) are expected be the canonical (neutral) forms of nucleobases (e.g. G8 and A38) and the doubly positively charged, hexa-water-coordinated Mg^{2+} ion ($[Mg(H_2O)_6]^{2+}$). The calculated free energy (and thus all energies calculated along the QM/MM pathway) of the precleavage ribozyme with different protonation states for these residues must be corrected for the presence of only a minor equilibrium population of these catalytically competent protonation states. The pK_a corrections, e.g., for N1-protonated form of A38 ($A38H^+$) and N1-deprotonated form of G8 ($G8^-$) are calculated as:[40]

$$\Delta G^{\mathrm{corr}}_{\mathrm{A38H^+}} = RT\ln 10\left(\mathrm{pH} - \mathrm{p}K_{\mathrm{a}}^{\mathrm{A38}}\right) \tag{5.2}$$

$$\Delta G^{\mathrm{corr}}_{\mathrm{G8^-}} = RT\ln 10\left(\mathrm{p}K_{\mathrm{a}}^{\mathrm{G8}} - \mathrm{pH}\right) \tag{5.3}$$

yielding 2.1 and 3.4 kcal/mol (at 300 K and pH 7), respectively. Note that the total correction is independent of pH and the terms involving pH mutually cancels each other.

5.1.3 Relation between free energy and rate constant

The uncatalyzed reaction in water (reference reaction) is characterized by the reference rate constant. The reaction rates are also being compared in the terms of free energies, since the activation free energy barrier ($\Delta G^{\ddagger}$) and rate constant (k) are related via Eyring equation:

$$k = \kappa\left(\frac{k_{\mathrm{B}}T}{h}\right)e^{-\Delta G^{\neq}/RT}\ , \tag{5.4}$$

where κ is the transmission coefficient; k_B and h are Boltzmann and Planck constants, respectively; T is temperature and R universal gas constant. The transmission coefficient κ is considered to be equal to 1 in most cases with complete exclusion of tunneling effects. Possible tunneling throw potential energy barriers with inclusion of computed transmission coefficients is determined, e.g., by Skodje and Truhlar formula:[161]

$$\kappa = \left(\frac{\beta\pi/\alpha}{\sin(\beta\pi/\alpha)}\right) - \left(\frac{\beta}{\alpha-\beta}\right)e^{[(\beta-\alpha)(\mathrm{V}_0-\mathrm{V}_2)]}\ , \tag{5.5}$$

assuming that $\beta \leq \alpha$, and where $\beta = 1 / k_B T$, $\alpha = 2\pi / \hbar\omega$, V_2 is the zero-point-including potential energy difference between TS and R states. V_2 is 0 for an exoergonic reaction and the (positive) zero-point-including potential energy difference between R and P states for an endoergonic reaction.[161] In our case, the calculated transmission coefficients for the reference reaction did not differ significantly from 1 with the maximal values of κ reaching ~1.3, which resulted in negligible decrease (by ~0.3 kcal/mol) of $\Delta G^{\ddagger}_{300K}$.

5.2 QM calculations of uncatalyzed reaction in water

The uncatalyzed (reference) reaction enables to (i) examine the performance of DFT and semiempirical approaches against higher level *ab initio* methods on the small model and (ii) estimate the free energy corrections on the extended models corresponding to the specific state along the reaction pathway for different reaction mechanism in ribozymes. QM calculations on extended model systems were also analyzed in order to investigate potential inaccuracies emerging from the small basis sets.

The complete scheme of intramolecular phosphodiester cleavage displays that seven main reaction pathways are plausible at physiological pH ~7 (Figure 6). Four pathways (I-IV) represent the dianionic reaction mechanisms, where the reactions proceed via (at least) one dianionic phosphorane TS state. Three other paths (V-VII) correspond to monoanionic reaction mechanisms, where reactions provide monoanionic IN and TS states. The reaction paths on the grey background in Figure 6 involve rare protonation states of R and IN states at pH ~7, i.e., neutral or positively charged phosphates and phosphoranes.

5.2.1 The minimal model and the accuracy of various QM methods

The cleavage of the minimal (uncatalyzed) 3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester model (Figure 4, Paths I and V in Figure 6) was used for assessing the performance of DFT functionals (mainly MPW1K, Table 2) and semiempirical Hamiltonians. Initial geometries were taken from Ref. 73 and reoptimized (at MPW1K/6-31+G(d,p) level) in the polarizable conductor calculation model (CPCM, $\varepsilon_r = 78.4$). Subsequently, single point calculations in gas phase were compared against energies by CBS(T) method calculated on the same geometries. CBS(T) indicates Møller–Plesset (MP2) calculations extrapolated for the complete basis set (CBS)[162,163], further corrected for higher-order correlation effects by CCSD(T) in a small basis set (cc-pVDZ).[164,165] The mean unsigned errors (MUE) of the MPW1K/6-31+G(d,p) method with respect to CBS(T) for all states along the reaction pathway are 0.7 and 2.2 kcal/mol for the monoanionic and dianionic reaction mechanism, respectively (Table 2). The MUE value for TS states from both mechanisms is 0.6 kcal/mol (Table 2). Thus, the gas-phase energies show that the hybrid DFT MPW1K functional optimized for reaction

kinetics[166,167] provides reasonably accurate energies in comparison with post Hartree–Fock wave function theory methods and can be recommended for QM/MM calculations on corresponding reactions in ribozymes.

Table 2: *The MPW1K functional errors obtained as a difference between MPW1K/6-31+G(d,p) and CBS(T) gas phase energies of the 3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester self-cleavage reaction for both dianionic and monoanionic reaction mechanisms.* [a]

monoanionic (endo/*pro*-R_P) [b]	R	TS_1	IN_1	TS_2	IN_2	TS_3	P	P`
MPW1K/6-31+G(d,p)	0.0	29.4	26.7	36.2	30.1	36.4	2.3	17.2
CBS(T) [c]	0.0	28.9	27.1	36.8	30.0	37.4	4.1	19.0
MPW1K error	0.0	-0.5	0.4	0.5	0.0	1.0	1.8	1.8
monoanionic (exo/*pro*-S_P)								
MPW1K/6-31+G(d,p)	0.0	27.4	21.2	32.3	26.1	36.8	2.3	17.2
CBS(T)	0.0	27.1	21.7	32.4	26.6	37.9	4.1	19.0
MPW1K error	0.0	-0.4	0.4	0.1	0.5	1.1	1.8	1.8
dianionic								
MPW1K/6-31+G(d,p) [d]	0.0	20.7	23.9			35.6 [e]	-49.1	
CBS(T)	0.0	21.9	25.6			36.6	-44.0	
MPW1K error	0.0	1.1	1.7			1.0	5.1	

[a] all geometries were optimized at CPCM (ε_r = 78.4)/MPW1K/6-31+G(d,p) level.
[b] monoanionic scenario can be realized via two (endo and exo) microscopic pathways involving *pro*-R_P (endo) and *pro*-S_P (exo) nonbridging oxygens of the scissile phosphate.
[c] MP2/CBS energies corrected to higher-order correlation effects using CCSD(T) energies (see Methods section).
[d] CPCM (ε_r = 78.4)/MPW1K/6-31+G(d,p) calculations revealed marginally stable IN state.
[e] second TS in the dianionic path is topologically closer to TS_3 rather than TS_2 from the monoanionic pathway (involving the departure of the alcoholate).

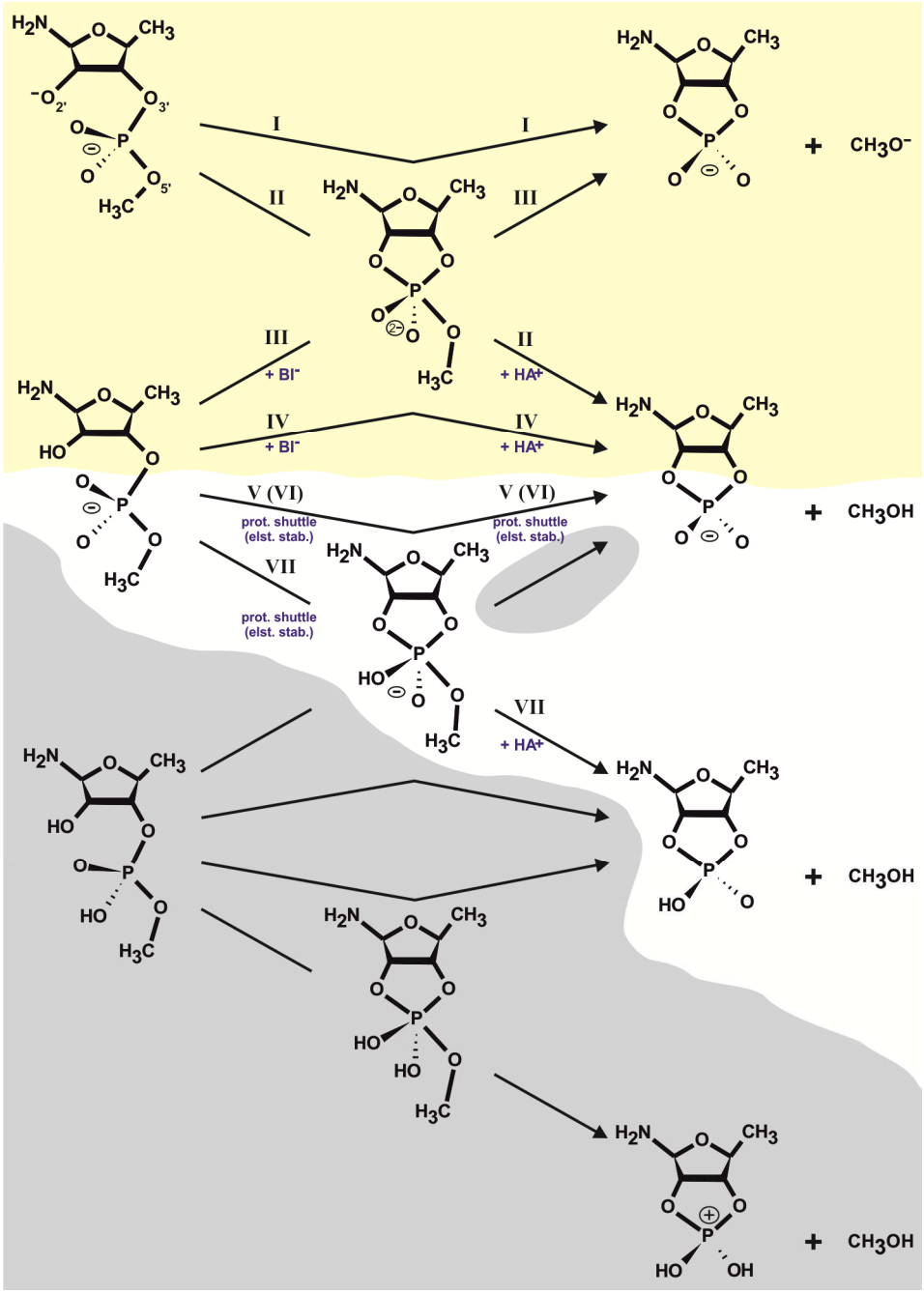

Figure 6: *Schematic diagram of all possible reaction paths of the intramolecular phosphodiester cleavage. Reaction paths (I, II, III and IV) on yellow background belong to the dianionic reaction mechanism (considering dianionic phosphorane); whereas paths on white background (V, VI and VII) involve monoanionic phosphorane (the monoanionic reaction mechanism). Remaining pathways on grey background involve rare protonation states of phosphates and phosphoranes at pH ~7.*

5.2.2 Extended models and the estimation of free energy corrections

The reaction paths II, III, IV, VI and VII in Figure 6 represent cleavage of extended models. The initial 3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester model was surrounded by nucleobases and hydrated Mg^{2+} ions (Figure 4) in protonation states that were suggested for participation in the reaction mechanisms of small self-cleaving ribozymes (Table 1, Figure 2). Nucleobases and Mg^{2+} can in principle act (i) directly as general bases/general acids in the

cleavage and/or (ii) provide only the electrostatic stabilization in general base/general acid positions (Figure 6). The calculated overall barriers for each pathway would (in principle) allow evaluate the effects of external species on the intramolecular phosphodiester cleavage. Such comparison however, could not be done in gas phase because the solvation term is critical, especially for charged systems (see Methods). It is worth noting that the estimation of solvation energies for conformationally variable phosphorane anion and its interaction with other charged centers is not trivial. Calculated solvation terms were found to be significantly different, when the reaction mechanism is changed even with the same external nucleobase/hydrated Mg^{2+} ion presented in position of general base/general acid (Table 3). As a result, the $\Delta G^{\ddagger}$ barriers between different reaction pathways could not be straightforwardly compared. Nevertheless, the relative trends along each specific reaction pathway are believed to be correct and the calculated free energy barriers could be used for extrapolation of the free energy corrections.[111] Still, the present estimation of solvation energies represent the weakest point in estimation of free energies of the reference reaction in water. Table 3 shows the example of calculated free energy corrections for the extended model system with canonical (or N1-deprotonated) N9-methylguanine (Gua or Gua^{-}, respectively) and N1-protonated N9-methyladenine ($AdeH^{+}$). Those corrections were further added to QM/MM energy profiles of the hairpin ribozyme self-cleavage in order to estimate the free energy changes. The obtained free energy corrections have only minor contribution to overall barriers, and more importantly, their absolute values are comparable between different reaction mechanisms (Table 3). Typical values of free energy corrections were around 1 kcal/mol, which is agreement with previous studies on small model systems.[73,89,168,169]

Table 3: *The MPW1K/6-31+G(d,p) gas phase energies, solvation terms, free energy corrections (calculated at CPCM(ε_r=78.4)/MPW1K/6-31+G(d,p) level), and total free energy profiles of the extended self-cleavage reaction.* [a]

monoanionic (endo/*pro*-R_P) [b]	R	TS_1	IN_1	TS_2	IN_2	TS_3	P
MPW1K/6-31+G(d,p)	0.0	25.6	14.4	26.8	19.4	24.3	-9.1
solvation energy [c]	0.0	-2.2	2.1	-1.2	0.6	5.0	5.6
free energy correction [d]	0.0	0.0	-0.8	-1.5	-0.7	-1.9	-4.2
MPW1K free energy in water [e]	0.0	23.4	15.7	24.1	19.3	27.3	-7.7
monoanionic (exo/*pro*-S_P)							
MPW1K/6-31+G(d,p)	0.0	23.7	3.4	13.5	10.5	16.4	-9.1
solvation energy	0.0	-3.8	6.6	3.2	1.8	6.7	3.7
free energy correction	0.0	-0.8	1.2	0.0	0.5	-2.4	-3.0
MPW1K free energy in water	0.0	19.1	11.2	16.7	12.8	20.7	-8.4
dianionic [f]							
MPW1K/6-31+G(d,p)	0.0	0.2					-41.4
solvation energy	0.0	13.5					33.8
free energy correction	0.0	-0.1					-4.0
MPW1K free energy in water	0.0	13.6					-11.6

[a] all geometries were optimized at CPCM (ε_r = 78.4)/MPW1K/6-31+G(d,p) level. All energies and energy corrections are in kcal/mol and are related to R state.

[b] endo/exo-3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester sugar-phosphate backbone model was extended by Gua and $AdeH^{+}$. The proton of 2'-OH hydroxyl was shuttled via *pro*-R_P (endo path) or *pro*-S_P (exo path) non-bridging oxygen.

[c] solvation term was calculated as the difference between CPCM(ε_r=78.4)/MPW1K/6-31+G(d,p) and gas phase MPW1K/6-31+G(d,p) SCF energies.
[d] free energy corrections were calculated at the CPCM(ε_r=78.4)/MPW1K/6-31+G(d,p) level.
[e] no pK_a corrections were included in total free energies.
[f] endo/exo-3'-(1'-amino-4'-methylribose)-5'-methylphosphodiester sugar-phosphate backbone model extended by Gua^- and $AdeH^+$. The Gua^- acts as a general base, while the $AdeH^+$ acts as a general acid.

5.2.3 Molecular complexes and the basis set superposition error

The overall free energy barriers for extended models (Figure 4) calculated by hybrid DFT MPW1K functional in medium basis set, i.e., 6-31+G(d,p), are also influenced by inaccuracies from the basis set superposition error (BSSE). The BSSE results in the overstabilization of molecular complexes, where the interacting monomers use each other orbitals to compensate for their own incomplete basis sets and the electronic energies are thereby artificially improved.[108,111] The BSSE artifact could be eliminated entirely, e.g., for gas phase calculations, by extrapolation to the CBS of atomic orbitals, or at least reduced by applying local electron correlation methods. In our case, three different reaction pathways (with Gua, $AdeH^+$, and both Gua and $AdeH^+$, Table 4) revealed that intramolecular BSSE affects the overall free energy barriers of all states (R, TS, IN, and P) along the reaction pathway with both Gua and $AdeH^+$ by ~3 kcal/mol. The relative differences between each of two states are however, minimal (up to 0.6 kcal/mol, Table 4) indicating that the relative free energies (even for the largest models) are not influenced significantly by the BSSE artifact.

Table 4: *Summary of calculated intramolecular BSSE contributions.* [a]

extended models [b]	R	TS_1	IN_1	TS_2	IN_2	TS_3	P
Gua	-1.3	-1.6	-1.5	-1.5	-1.6	-1.3	-1.2
$AdeH^+$	-1.4	-1.4	-1.6	-1.6	-1.7	-1.8	-1.3
Gua and $AdeH^+$	-2.4	-2.8	-2.8	-3.0	-2.9	-2.9	-2.5

[a] intramolecular BSSE contribution (kcal/mol) calculated as the difference between MPW1K/6-31+G(d,p) gas phase SCF energies, where (i) all nucleobase atoms were marked as dummy (not counted), and (ii) respective nucleobase was not presented (in identical geometry of sugar-phosphate backbone model).
[b] all structures were optimized by CPCM (ε_r = 78.4)/MPW1K/6-31+G(d,p)), where Gua and $AdeH^+$ acted indirectly in reaction mechanism (for electrostatic stabilization in position of general base and general acid).

5.3 Investigations of reaction mechanisms in ribozymes

The identification of reactive conformations is necessary for evaluation of catalytic mechanisms. MD simulations are able to model native (reactive) conformations of active sites on the basis of available crystallographic data, which are often limited in predicting power due to the necessity of using inhibited and/or mutated structures.[69,135,170-172] MD simulations generate catalytically interesting snapshots that were further used as suitable starting points for the systematic identification of plausible reaction pathways in the hairpin and HDV ribozyme by hybrid QM/MM calculations.

5.3.1 The reactive conformation and overall stability of various protonation states of key residues within the active site of the hairpin ribozyme

MD simulations of the hairpin ribozyme showed different conformation of A–1 ribose ring capped with inactivating A–1(2'-methoxy) modification (used in starting X-ray structure to prevent self-cleavage) and with the native A–1(2'-OH) group. MD simulation with A–1(2'-methoxy) group contained the lowest root mean square deviation values (RMSD), suggesting that this simulation is in best agreement with the crystallographic geometry. On the other hand, simulations with native A–1(2'-OH) group revealed rapid changes of A–1 sugar pucker. Such reconformation indicates that the methoxy group distorts the active site. This observation agrees with previous MD simulations, where C2'-endo/C3'-endo repuckering of A–1 and repositioning of its 2'-OH were observed.[69]

MD simulations also identified a potential artifact of the AMBER *ff*99 and *ff*99bsc0 force fields leading to the generation of 'ladder-like' structures in one stem of the hairpin ribozyme. This 'ladder-like' distortion appeared on a tens-of-nanoseconds timescale and was observed to be irreversible on the accessible (~500 ns) timescales. The artifact affected only considerable part of the H4 helix and fortunately did not propagate into the catalytically relevant components of the simulated structures. The 'ladder-like' helix deformation was characterized by a significant shift of glycosidic χ torsion angles, loss of helical twist, a change of the sugar-pucker, and by an increase in slide and P-P distances. Further tests revealed that the 'ladder-like' distortions could be attributed to the χ torsion parameterization and an inclusion of new χ_{OL3} parameters eliminated that behavior.[141]

Simulations with the canonical G8 form in the active site indicate that G8 is structurally consistent with crystallographic data. The G8(N1) imino group of the canonical G8 formed stable hydrogen bonds with the G+1(*pro*-S_P) or G+1(*pro*-R_P) nonbridging oxygens of the scissile phosphate and/or a hydrogen bond with A–1(O2'). On the other hand, the deprotonated $G8^-$ form quickly left the active site, which caused large structural distortions of the active site likely due to electrostatic repulsion with the scissile phosphate (Figure 7). This observation is not consistent with a catalytic role of G8 as the general base. The G8 enol tautomer (G8t)

remained in contact with the active site and might also be considered for a potential structural role in catalysis (Figure 7). However, the role of G8t in catalysis is questionable because guanine tautomers are highly unfavorable in water.[61,173] Thus, within the limits of classical MD simulations the canonical G8 is structurally and energetically feasible for the mechanism of self-cleavage (Figure 7).

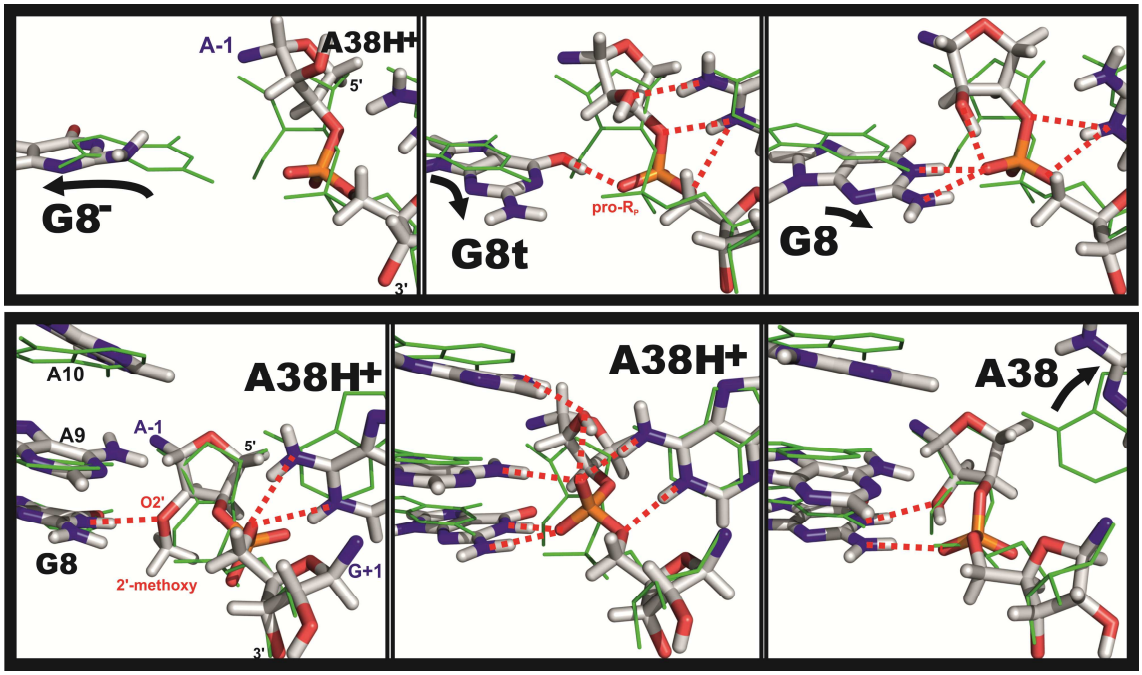

Figure 7: *Active site structures from MD simulations showing the behavior of different G8 and A38 protonation states. The structures shown in sticks are averaged over the last nanosecond of MD simulation and superimposed with the starting crystal structure (green lines).*

In contrary to G8, the protonated (noncanonical) form A38 ($A38H^+$) is most consistent with the available crystal structures. Three types of behavior were observed for the (unprotonated) canonical A38 during MD simulations: (i) A38 departed from the scissile phosphate, which led to large structural changes in the S-turn bearing the A38 base; (ii) A38 shifted from the scissile phosphate but remained at an ~7-Å distance after losing its base pairing with A24; and (iii) A38 established a hydrogen bonding contact with A–1(2'-OH) and remained close to the scissile phosphate. Once established, the contact between the A–1(2'-OH) nucleophile and the A38 base remained stable until the end of the MD simulation. This contact might be catalytically relevant because A38 was suggested as a potential shuttle capable of accepting a proton from the nucleophile and transferring it to the G+1(O5') oxygen of the leaving alcoholate.[69] All simulations with the protonated $A38H^+$ generally agree well with the crystal structure data (Figure 7).

All together, MD simulations of the hairpin ribozyme revealed that canonical G8 and protonated $A38H^+$ are the most consistent protonation states with the available structural data. In case of deprotonated $G8^-$ form, its departure from the active site did not affect the global fold of the hairpin ribozyme. Thus, it remains questionable, whether the empirical force field could

sufficiently describe the repulsion between two negatively charged centers (deprotonated G8$^-$, the scissile phosphate). Additional classical MD simulations revealed only metastable interaction between G8$^-$ and the scissile phosphate tolerated on time scales of hundreds of ps. While MD simulations can reveal the structural stability of catalytically relevant conformations, they cannot address their reactivity. As a result, both canonical G8 and protonated A38H$^+$ do not have to be reactive states if other forms, e.g. G8$^-$, are sufficiently reactive. Liu and co-workers used recently the 8-azaguanosine/G8 substitution in the active site and observed that pK_a of the substituted nucleobase is not perturbed by the active site.[58] Thus, only a negligible fraction of G8 will be deprotonated at a physiological pH ~7, underlining that general base role of G8$^-$ requires inherent sufficient reactivity.[58]

The results of this study were published in the Journal of Physical Chemistry B in 2010. The manuscript is available in the Appendix A.

The structural dynamics of the hairpin ribozyme containing canonical G8 and protonated A38H$^+$ forms in the active site (the most stable protonation states) was further investigated by three 500-ns-long MD simulations with different force fields and simulation settings; i.e., (i) unrestrained simulation with the *ff*99bsc0χ_{OL3} force field based on AMBER *ff*99 Cornell *et al.* force field[134,139] corrected by Barcelona α/γ[140] and Olomouc χ_{OL3}[141] reparameterizations, (ii) restrained simulation in the same force field with τ_2 torsion of A–1 ribose ring restraint keeping its C2'-endo pucker, and (iii) the unrestrained simulation with *ff*99bsc0$\chi_{OL3}\varepsilon\zeta_{OL1}$ force field, which is corrected by the most recent Olomouc $\varepsilon\zeta_{OL1}$ reparameterization[145]. Consistently with older simulations using *ff*99bsc0 force field,[143] the A–1 ribose changed (within first few ns of MD simulation) its sugar-pucker from C2'-endo to C3'-endo conformation also with the *ff*99bsc0χ_{OL3} force field, leading to loss of the G8(N1H)…A–1(O2') hydrogen bond. The C3'-endo state was characterized by fast fluctuations between two conformations of the scissile phosphate (Figure 8, top) with only marginal population of conformations suitable for the self-cleavage reaction, i.e. those having high values (above ~140°) of the in-line attack angle (IAA, A–1(O2')-G+1(P)-G+1(O5')) accompanied with A38H$^+$(N1H)…G+1(O5') and G8(N1H)…A–1(O2') hydrogen bonds. In contrast, the subsequent *ff*99bsc0χ_{OL3} simulation with restrained C2'-endo pucker of A–1 ribose flipped (after few tens of ns) the scissile phosphate into a conformation that is well suited for the catalysis. Such active site geometry revealed firm hydrogen bonds between the scissile phosphate and the catalytically important nucleobases G8 and A38H$^+$, namely G8(N1H)…A–1(O2'), G8(N2H)…G+1(*pro*-S$_P$), A38H$^+$(N1H)…G+1(O5'), and A38H$^+$(N6H)…G+1(*pro*-R$_P$). The simulation also populated high values of IAA (with mean value of ~140° with fluctuations up to 170°). Nonetheless, the simulation maintained this geometry only for ~150 ns. Then the scissile phosphate lost the reactive conformation and adopted another fluctuating (bistable) state (Figure 8, middle) characterized by +*gauche* and *trans* conformation of A–1 of the ε and ζ torsions, respectively.[136] This is rare non-canonical

conformation of RNA sugar-phosphate backbone. It does not correspond to any annotated suite in database of RNA backbone conformations,[174,175] however, some occurrences of this sugar-phosphate backbone were found in the structural database (Figure 3B in Ref. 174).

Within third simulation, the effect of the $\varepsilon\zeta_{OL1}$ correction on the simulation behavior was tested. The structural dynamics of the hairpin ribozyme active site in *ff*99bsc0$\chi_{OL3}\varepsilon\zeta_{OL1}$ force field was changed significantly. The hairpin ribozyme maintained the reactive active site conformation (the same as described above for the restrained simulation) including the population of high values of IAA on the whole simulation time scale, i.e., 0.5 μs (Figure 8, bottom). It is worth noting that in this simulation the A–1 pucker dominantly populates C2'-endo conformation even without the necessity of additional restrain to the A–1 ribose pucker.[136] Thus it seems that the preference of the C3'-endo conformation in *ff*99bsc0χ_{OL3} and in preceding *ff*99bsc0 and *ff*99 simulations[69,143] is dictated by the local conformation of the scissile phosphate. Once the ε/ζ torsions adopted the +*gauche/trans* conformation with the aid of the $\varepsilon\zeta_{OL1}$ correction, the A–1 ribose stably fluctuated in the C2'-endo region.[136]

The RNA self-cleavage reaction requires a rare noncanonical conformation of the scissile sugar-phosphate backbone segment. As the empirical force fields are primarily designed to describe the canonical regions,[176] descriptions of non-canonical RNA conformations are challenging. In case of the conformational behavior of the scissile phosphate, key parameters seem to be dihedrals affecting the ribose pucker and ε/ζ torsions.[136]

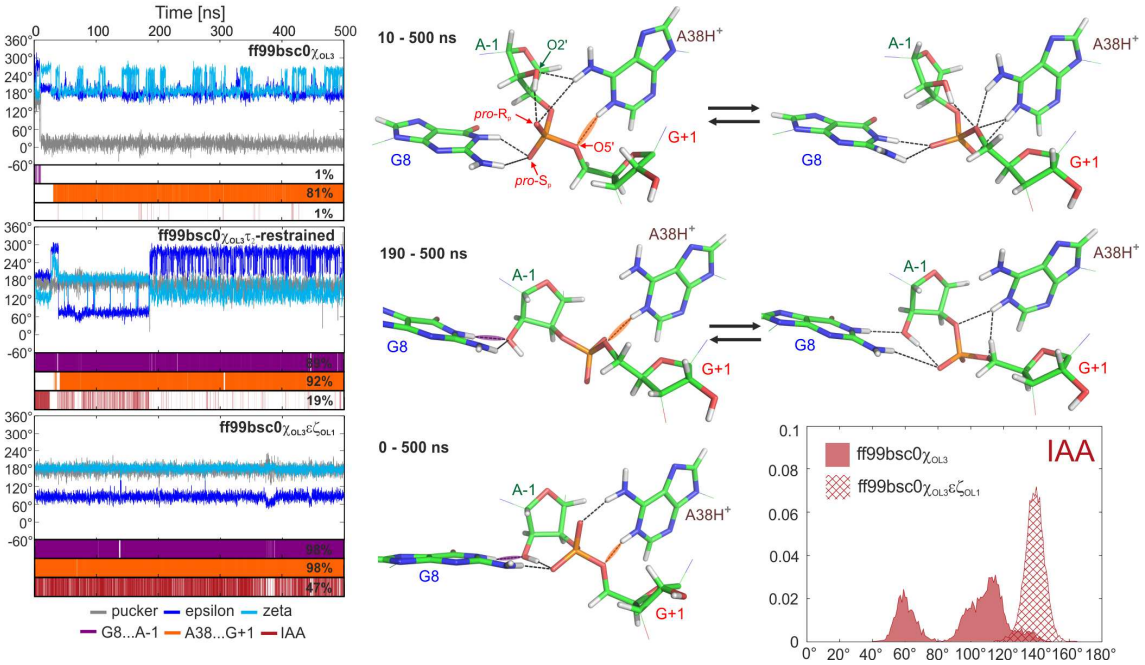

Figure 8: *Structural dynamics of the active site using MD simulations with (i) ff99bsc0χ_{OL3} AMBER RNA force field (top), (ii) the same force field with restrained A–1 C2'-endo ribose pucker (middle), and (iii) ff99bsc0$\chi_{OL3}\varepsilon\zeta_{OL1}$ force field, i.e., including the latest ε/ζ reparameterization.*[145] *The time courses of pucker, ε and ζ of A–1 determining conformation of the scissile phosphate are shown in lines, while the occurrence of G8(N1H)…A-1(O2') and A38H$^+$(N1H)…G+1(O5') hydrogen bonds (with heavy atom distance below 3.5 Å) and favorable IAA (above 140°) are shown in stripes with % of population indicated. The histograms of IAA (A–1(O2')-G+1(P)-G+1(O5') in ff99bsc0χ_{OL3} and ff99bsc0$\chi_{OL3}\varepsilon\zeta_{OL1}$ simulations are shown in bottom-right graph.*[136]

5.3.2 QM/MM calculations of the hairpin ribozyme suggest the feasibility of multiple competing reaction mechanisms

The representative snapshots (containing canonical G8 and protonated A38H$^+$) were selected as starting structures for QM/MM calculations on the basis of structural criteria for the reactive conformation, namely a high value for the IAA of A–1(O2')…G+1(P)…G+1(O5'), typically above 150°, with the presence of both A38H$^+$(N1)…G+1(O5') and G8(N1H)…A–1(O2') hydrogen bonds. Combined QM/MM calculations using DFT functionals (BLYP for initial geometry optimizations and MPW1K for subsequent single-point recalculations) show that the self-cleavage pathway of the hairpin ribozyme may follow several competing microscopic reaction mechanisms, which provide activation barriers in good agreement with experimental data (20-21 kcal/mol). The initial nucleophilic attack of the A–1(2'-OH) group on the scissile phosphate is predicted to be rate-limiting in all these mechanisms. Most importantly, identified pathways involve both main reaction scenarios suggested for phosphodiester cleavage: dianionic and monoanionic mechanisms.[46] The main reaction pathway (the active site contains canonical G8 and protonated A38H$^+$) is the monoanionic (proton-shuttling) mechanism, characterized by proton transfers via the G+1(*pro*-R$_P$) nonbridging oxygen of the scissile phosphate and overall barrier of 21.0 kcal/mol (Figure 9). Three TS states were identified along

the reaction coordinate, where two IN states corresponded to the single protonated phosphorane form. Both canonical G8 and protonated $A38H^+$ did not participate in any proton transfer and served predominantly for alignment of reactive groups and electrostatic stabilization of the negatively charged phosphorane. The $A38H^+$ was also identified as possible general acid in the second part of the reaction (the exocyclic cleavage step, TS_3 state), which resulted in P state significantly less stable (by 9.5 kcal/mol) due to protonation of the cyclic phosphate. Subsequent deprotonation of the cyclic phosphate might be accompanied by protonation of A38, which should lead to energy relaxation of the product. Thus, both scenarios with $A38H^+$ are chemically equivalent, highlighting the versatility of the hairpin ribozyme. The canonical A38 also provided a feasible activation barrier but required a properly structured active site derived from simulations with a protonated $A38H^+$ form. The protonation state of A38 does not perturb or switch the rate-limiting step, which is the nucleophilic attack of the 2'-OH on the scissile phosphate.

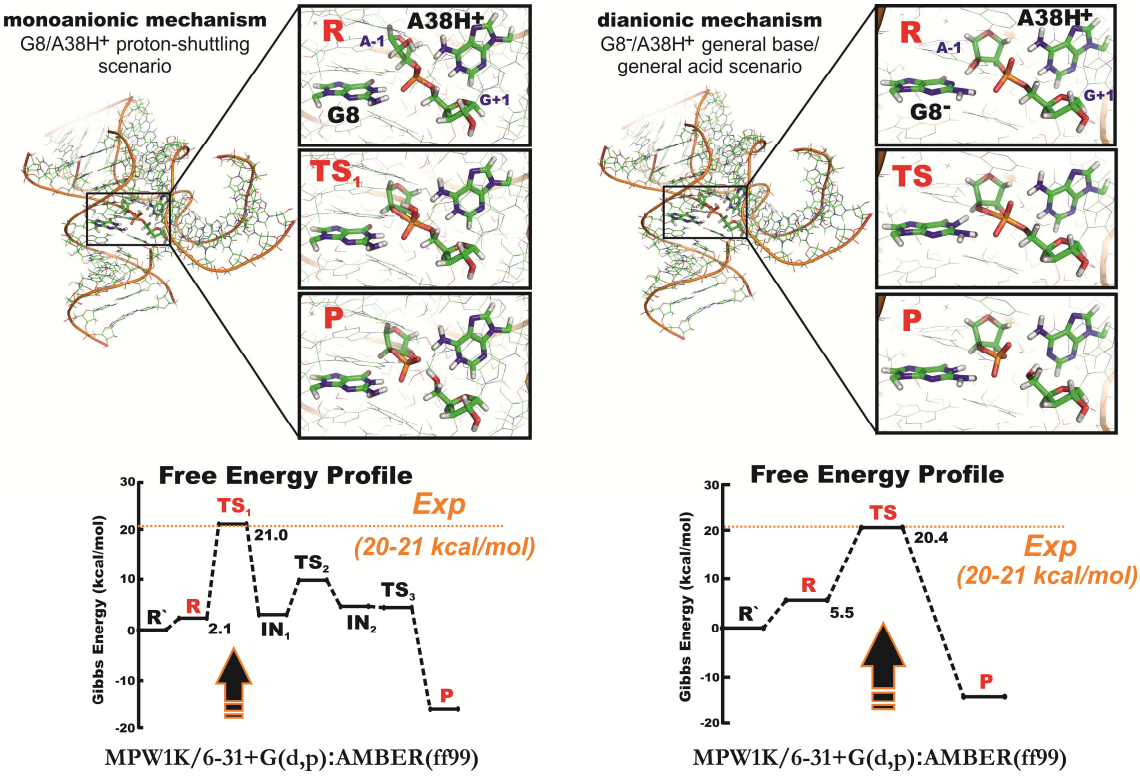

Figure 9: *Schematic diagrams for the relative free energy profiles along monoanionic and dianionic reaction pathways obtained by QM/MM calculations of the hairpin ribozyme. Structures in boxes show details of the active site (with the QM core highlighted as sticks) in the R, the rate limiting TS, and P states, respectively.*

The dianionic mechanism was realized via general base/general acid reaction, where deprotonated $G8^-$ and protonated $A38H^+$ act as a general base and a general acid, respectively. The single TS involved deprotonated (double negatively charged) phosphorane and corresponded with the overall barrier of 20.4 kcal/mol (Figure 9). The combined general base/general acid mechanism is in excellent agreement with experimentally measured pH

profiles. Estimated pK_a's of G8 (9.5)[58] and A38 (5.5)[67] within the active site of the hairpin ribozyme fits well with experimental profiles showing reaction rates as functions of pH (Figure 5). However, the initial snapshot was taken from MD simulation with canonical G8 form because the deprotonated $G8^-$ form quickly (after initial minimizations) left the active site during MD simulations. The R state conformation with $G8^-$ form in the active site was further investigated by preliminary ADMP simulations (*ab initio* QM/MM–MD). The interaction between $G8^-$ and the scissile phosphate remained stable on the time-scale of several ps. Unfortunately, longer time-scales are still out of reach due to enormous computational costs for systems containing ~50,000 atoms.

The experimentally observed pH-rate profiles cannot be straightforwardly explained by the proton shuttling (monoanionic) mechanism as it should not lead to any detectable pH dependence. One proposal might be, that at least one other mechanism competes with the proton shuttling mechanism to produce the experimentally observed pH-rate profile. The dianionic $G8^-/A38H^+$ general base/general acid mechanism is a plausible candidate since it may compete effectively. The main obstacle for this mechanism is the observation that so far classical MD simulations do not reveal a tendency of $G8^-$ to easily establish a catalytically relevant geometry. Therefore, it appears that either the $G8^-$ is considerably more reactive (in consistency with calculated data) or that G8 has a high propensity of transiently forming the proper hydrogen bond with A–1(2'-OH) immediately after deprotonation.

QM/MM calculations of the hairpin ribozyme were published in the Journal of Physical Chemistry B in 2011 and are given in the Appendix B.

5.3.3 Comparison of various QM methods for description of catalytic mechanism in the hairpin ribozyme

In order to further differentiate between the monoanionic (proton shuttling mechanism) and the dianionic (combined general base/general acid mechanism) reaction scenarios, the QM DFT (BLYP, MPW1K) approach was compared with additional semiempirical (AM1/d-PhoT, SCC-DFTB) and *ab initio* (SCS-MP2) QM methods. We constructed the potential (PES) and, when available, free energy (FES) surfaces of the self-cleavage reaction, where we focused mainly on the first reaction step, i.e., the nucleophilic attack of the A–1(2'-OH) group to the G+1 scissile phosphate with simultaneous proton transfer from A–1(2'-OH) to either one of nonbridging oxygens of the scissile phosphate (monoanionic) or to $G8^-$(N1) nitrogen (dianionic mechanism).

All tested QM/MM methods provided activation barriers with acceptable agreement with the experimental data for both monoanionic and dianionic mechanisms. On the other hand, we identified large differences in the nature of the reaction scenario predicted by the different methods. Semiempirical methods suggested sequential (step-wise) reaction pathways, where the first proton transfer from A–1(2'-OH) to either the G+1(*pro*-R_P) nonbridging oxygen or

G8$^-$(N1) nitrogen was separated from the subsequent nucleophilic attack of A–1(O2') to G+1(P) (Figure 10). This is in a sharp contrast with the concerted pathway (proton transfer from the A–1(2'-OH) to either the G+1(*pro*-R_P) nonbridging oxygen or G8$^-$(N1) nitrogen occurred simultaneously with nucleophilic attack of A–1(O2') to G+1(P)) favored by the *ab initio* SCS-MP2 calculations (our reference method) and both DFT functionals (BLYP, MPW1K) (Figure 10).

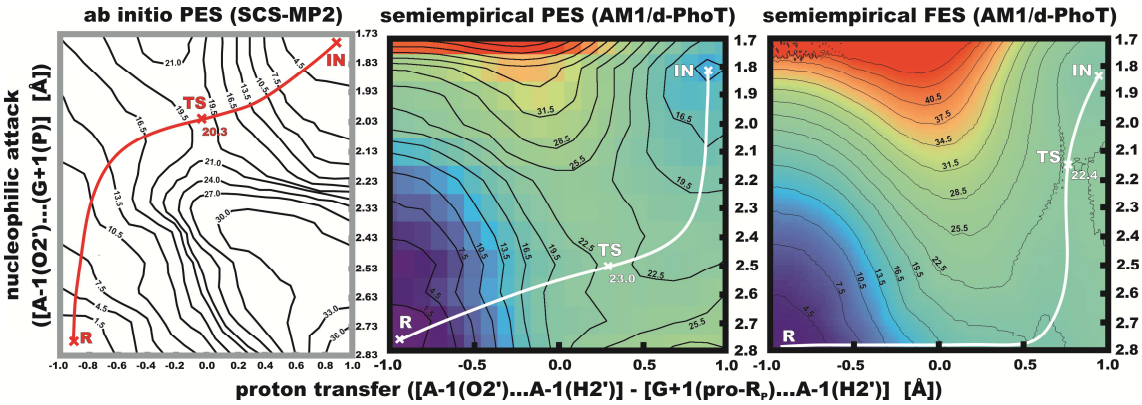

Figure 10: *Two-dimensional QM/MM maps of the PES (calculated by SCS-MP2 and AM1/d-PhoT) and FES (AM1/d-PhoT) for the monoanionic mechanism. Red and white lines indicate the reaction paths with R, TS, and IN states.*

As the approximate semiempirical methods supported the sequential pathway at both the PES and FES levels of description, the preference for the sequential scenario can be primarily attributed to the semiempirical potential rather than application of the free energy MD approach. In addition, gas phase QM calculations ruled out a possibility that the above-noted discrepancy reflected some QM/MM specific settings, e.g., QM/MM couplings. The inconsistent estimation of PES shapes originated in semiempirical Hamiltonians. Since the results obtained by robust *ab initio* and DFT methods were mutually consistent, it appears that semiempirical methods, although clearly superior in terms of sampling, may visibly distort the calculated PES and FES of ribozyme reactions. Semiempirical methods are still widely used for investigation of reaction mechanism catalyzed by biomacromolecules. Significantly different cleavage pathways predicted for the hairpin ribozyme indicates that mechanistic predictions based on semiempirical methods should be interpreted with great care.

Comparison of QM methods for the investigation of the reaction mechanism in the hairpin ribozyme was published in the Journal of Chemical Theory and Computation in 2014 and is given in the Appendix C.

5.3.4 MD simulations of the HDV ribozyme reveal several positions of the catalytically active Mg^{2+} ion

Explicit solvent 80-ns long MD simulations were used for inspection of position and bounding patterns of the active site ion in the HDV ribozyme. MD simulations contain (i) monovalent ions only, (ii) one divalent Mg^{2+} ion in the active site, and (iii) four divalent Mg^{2+} ions in positions with presumed important structural and/or functional roles. All those positions were taken from the latest crystal structure of *trans*-acting HDV ribozyme with 2'-deoxy-mutation of U–1.[80] The simulations with Mg^{2+} in the active site revealed high tendency to form triple-innershell coordination to U–1(O2'), G1(*pro*-R_P), and U23(*pro*-S_P), populating 99.3% and 30.1% of all states in simulation with one and four divalent ions, respectively. Such coordination of the Mg^{2+} ion corresponds to the active site arrangement of the starting X-ray structure,[80] and was used in recent QM/MM calculations by Hammes-Schiffer group.[83,177] The triple-innershell coordination shifts to the double-innershell coordination to [G1(*pro*-R_P), U23(*pro*-S_P)] nonbridging oxygens in the MD simulation with more divalent ions (populating remaining 69.9% of the simulation). It should be noted that divalent ions are poorly described by the approximate, non-polarizable force field and susceptible to accumulation of simulation artifacts during MD simulations.[178] In addition, the first-shell ligands of the Mg^{2+} ion have life-time on microsecond time-scale and thus the presented MD simulations with Mg^{2+} ion were not able to sufficiently sample its position in the active site. In most cases, it is better to avoid the usage of divalent ions in classical MD simulations, specifically when they do not play any significant structural role and can be replaced by monovalent ions.[108,135]

MD simulations with the active site Na^+ ion revealed significantly different and more dynamic behavior, where the most populated state (36.4%) was the outershell coordination with six water molecules in the first coordination shell (forming a canonical hexacoordination). In addition, we observed 4 different triple-innershell coordinations, 10 distinct double-innershell coordinations, and 6 other single-innershell coordinations with the population of each state at least 0.1% during respective MD simulation. In total, 16 different positions and coordinations of the Mg^{2+} ion in the active site were considered for QM/MM calculations. Starting snapshots were subsequently selected from geometries achieving best orientation of key residues in the active site, where catalytically important hydrogen bond distances (U–1(2'-OH)…WAT(O) and $C75H^+$(N3H)…G1(O5')) were below 2.8 Å with the highest available value (typically greater than 160°) of the IAA (U–1(O2')…G1(P)…G1(O5').

5.3.5 The activation barrier of the HDV ribozyme is affected by the specific position and coordination of the Mg^{2+} ion

QM/MM calculations were used to localize the self-cleavage reaction path of the HDV ribozyme, where $C75H^+$ acts as the general acid and hydroxide ion is coordinated to active site

Mg^{2+} ion. 16 different positions and coordinations of the Mg^{2+} ion were considered and free energies along these reaction pathways were estimated. The QM/MM minimizations and scans revealed 8 complete reaction pathways (Table 5). In the initial part of the reaction, the U–1(2'-OH) nucleophile was deprotonated/activated by a hydroxide ion coordinated to the Mg^{2+} ion, which abstracted the proton from the U–1(2'-OH) group, i.e., acted as the general or specific base. The deprotonation of the U–1(2'-OH) nucleophile and nucleophile attack are consecutive so that the deprotonation of the U–1(2'-OH) precedes the nucleophile attack in all investigated reaction pathways with different coordination of the Mg^{2+} ion. The nucleophile attack is then concurrent with the second proton transfer from the protonated $C75H^+$ to the leaving G1(O5') group. In contrast, an equivalent combined general base/general acid mechanism of the hairpin ribozyme shown that the initial activation of the 2'-OH nucleophile via deprotonation by the $G8^-$ and nucleophilic attack are simultaneous events, corresponding to the TS of the rate-limiting step.

The calculated overall free energy barriers range from 14.2 to 28.8 kcal/mol and are to some extent influenced by the specific coordination of active site Mg^{2+} ion during the reaction. The snapshot with double-innershell coordination to [U–1(O2), U–1(O2')], used for detail exploration of the potential energy surface, displayed the lowest activation barrier of 14.2 kcal/mol. Four other coordinations, i.e., double-innershell coordinations to [G1(*pro*-R_P), U23(*pro*-S_P)], [G1(*pro*-R_P), U20(O2)], and [G25(O6), G25(N7)], and the triple-innershell coordination to [U–1(O2'), G1(*pro*-R_P), U23(*pro*-S_P)] revealed comparable and slightly higher activation barriers of 15.6, 17.6, 18.2, and 15.7, kcal/mol, respectively. The remaining three reaction pathways, where the Mg^{2+} ion has single-innershell coordination to [U–1(O2)], triple-innershell coordination to [G1(*pro*-R_P), U20(O2), G25(O6)], and double-innershell coordination to [U–1(O2), G25(N7)], showed significantly higher free energy barriers of 21.9, 28.0, and 28.8 kcal/mol, respectively (Table 5).

The experimentally measured rate constants indicate an activation barrier of 19-20 kcal/mol for the *trans*-acting HDV ribozyme under physiological conditions (298 K, pH 7).[30,179] The activation barriers estimated here, i.e., between 14 and 29 kcal/mol, cover the typical range of many enzymatic reactions (10-20 kcal/mol)[3] as well as the experimental measurements for the HDV ribozyme and other small self-cleaving ribozymes (19-21 kcal/mol).[28,29,31,32] On the other hand, the wide range of estimated barrier heights suggests sensitivity of the reaction to the specific arrangement of the active site, particularly to the position and the coordination of the active site Mg^{2+} ion. Beside that, the variation in the calculated free energy profiles may be also caused by: (i) limited accuracy of the QM/MM approach, (ii) indirect estimation of the free energy corrections, which were extrapolated from the uncatalyzed reaction of a small model system; and/or (iii) by an uncertainty in the additional correction corresponding to deprotonation of the partially hydrated Mg^{2+} ion, which relies on a roughly estimated pK_a and the specific localization within the active site of the HDV ribozyme.

Table 5: *Free energy barriers (in kcal/mol) of reaction pathways representing the combined general base/general acid reaction mechanism in the HDV ribozyme.* [a]

coordination of Mg^{2+} [b]	R	R' [c]	TS	P
double-innershell to [U–1(O2), G25(N7)]	8.3	6.1	28.8	11.1
triple-innershell to [G1(*pro*-R_P), U20(O2), G25(O6)]	8.3	11.5	28.0	-1.1
single-innershell to [U–1(O2)]	8.3	8.7	21.9	4.4
double-innershell to [G25(O6), G25(N7)]	8.3	4.1	18.2	2.2
double-innershell to [G1(*pro*-R_P), U20(O2)]	8.3	4.4	17.6	-9.4
triple-innershell to [U–1(O2'), G1(*pro*-R_P), U23(*pro*-S_P)]	8.3	3.7	15.7	-4.8
double-innershell to [G1(*pro*-R_P), U23(*pro*-S_P)]	8.3	5.0	15.6	-21.2
double-innershell to [U–1(O2), U–1(O2')]	8.3	7.0	14.2	1.1

[a] energies are calculated at the MPW1K/6-31+G(d,p) level and includes all corrections, i.e., pK_a for the rare protonation states of the active site (8.3 kcal/mol) with the free energy corrections taken from the model of reference reaction (-0.7 and -5.0 kcal/mol for TS and P state, respectively).
[b] active site groups participating in the coordination.
[c] R' state contains the deprotonated U-1(2'-O^-) group.

5.3.6 QM/MM calculations of the HDV ribozyme predict significant pK_a shift of the U–1(2'-OH) group

Two titrable groups in rare protonation states are involved in the reaction mechanism of the HDV ribozyme. The first group is the protonated cytosine $C75H^+$, for which the pK_a constant (within the environment of the active site) was measured by Raman crystallography.[180] The second group is either a hydroxide anion coordinated to the catalytic Mg^{2+} ion or the already deprotonated U–1(2'-O^-) nucleophile. The experimental values of the pK_a constant of the ribose 2'-hydroxyl are ambiguous, ranging from 12 up to 15.[42,181-186] In addition, the direct, first-shell coordination of the 2'-OH group to the active site Mg^{2+} ion most likely significantly shifts the pK_a of this hydroxyl within the HDV ribozyme active site to lower values, as suggested by proton inventory experiments[75] and NMR spectroscopic measurements.[177] By contrast, the pK_a value of a hydrated Mg^{2+} ion is unambiguously measured (pK_a of 11.4)[187] and is expected to be less affected by the active site environment. Therefore, the correction terms for the rare protonation states used in this study were estimated from the pK_a of C75 (already shifted in the HDV ribozyme active site environment) and the pK_a of a hydrated Mg^{2+} ion. The pK_a shift of the hydrated Mg^{2+} ion within the HDV ribozyme active site is, however, unknown, and thereby figures as a limiting factor in any prediction.

The calculated reaction profiles suggest that both the precleavage state R (its rare protonation form) and the intermediate state R', i.e., the state with native U–1(2'-OH) and a hydroxide coordinated to the Mg^{2+} ion and the state with already deprotonated U–1(2'-O^-) and a water molecule coordinated to Mg^{2+}, respectively, are typically close in free energy (Table 5). The estimated pK_a values of the U–1(2'-OH) group range from 8.8 up to 14.5 by using free energy differences between two precursors (R and R') from Table 5. If we do not take into account the reaction profiles with highest activation barriers (21.9, 28.0, and 28.8 kcal/mol) that most likely correspond to the unfeasible reaction paths, then the interval for estimated pK_a's is reduced from 8.8 up to 11.2. This finding implies that the pK_a of the 2'-OH group in the environment of HDV

ribozyme active site is comparable to and most likely smaller than the pK_a of the solvated active site Mg^{2+} ion. Considering the experimental predicted pK_a's of the ribose 2'-hydroxyl (from 12 up to 15)[42,181-186], with probably the most relevant value of 12.8 for UpG monophosphate dinucleotide from NMR spectroscopic measurements,[182] the pK_a of the U–1(2'-OH) group is possibly lowered by ~1.6 – 4.0 units (close to or even below a pK_a of ~11.4[187] for the solvated Mg^{2+} ion).

Results from MD simulations and QM/MM calculations of the HDV ribozyme were submitted to the Physical Chemistry Chemical Physics journal. The manuscript with Supporting Information is available in the Appendix D.

Chapter 6

Summary

The presented thesis describes the structural dynamics and reaction mechanisms of RNA self-cleavage reactions catalyzed by two small self-cleaving ribozymes, i.e., hairpin and HDV ribozymes. The classical MD simulations in explicit solvent were used to study conformational variability and to identify protonation states of the active site nucleobases under crystalline conditions. The MD simulations were also used to obtain suitable starting structures for subsequent multiscale QM/MM methods in order to study the reaction mechanisms.

In the hairpin ribozyme, canonical and protonated forms of the key active site nucleobases G8 and A38H$^+$, respectively, were most consistent with available X-ray structures. The investigated system with G8 and A38H$^+$ in the active site remains stable and compact in MD simulations on the microsecond time scale. QM/MM calculations of the hairpin ribozyme suggested two reaction mechanisms, i.e., combined general base/general acid and proton-shuttling mechanisms, with overall barriers (20.4 and 21.0 kcal/mol, respectively) in excellent agreement with the activation barrier derived form experimentally measured rate constants (20–21 kcal/mol).[28,188] While the combined general base/general acid mechanism corresponds with experimentally measured pH profiles the proton shuttling mechanism can not be explained by experimental observations as it would not lead to any apparent pH dependence of the rate constant. Both mechanisms are energetically close and might be in competition.

In the HDV ribozyme, MD simulations suggested possible active site arrangements with specific position and coordination of the catalytically essential Mg^{2+} ion, which were used as suitable starting points in subsequent QM/MM calculations. QM/MM calculations of the HDV ribozyme showed that specific coordination of the Mg^{2+} ion in the active site significantly affects the activation barrier of the self-cleavage reaction. The QM/MM energies indicate significant pK_a shift of U–1(2'-OH) group in the active site of HDV ribozyme that would contribute to catalysis by activating the U–1(2'-OH) nucleophile.

We also identified an artifact inherent to AMBER *ff*99/*ff*99bsc0 force fields that affected RNA helical structures with the potential to accumulate on longer (tens-of-nanoseconds) time scales. This finding ignited additional parameterizations leading to introduction of currently well performing *ff*99bsc0χ_{OL3} version[141] of AMBER Cornell *et al*. force field[134] that became standard AMBER force field since 2010 (it corresponds to RNA part of standard *ff*10, *ff*12 and/or *ff*14 force field in AMBER).

It should be noted that each utilized computational method has its merit but also its limitations. While MD simulations can reveal the structural stability of catalytically relevant conformation they cannot address their reactivity. As a result, the most stable protonation states do not have to be reactive states if the other (rare) protonation forms are sufficiently reactive. In addition, force fields are still under development and fine tuning of the parameters could possibly affect reactive conformations as well as the stability of specific protonation states of residues. The description of divalent ions during MD simulations also remains problematic and would preferably require introduction of a well tested polarizable force field for RNA systems that is however missing in contemporary force fields.

Utilized QM/MM scheme reveal just energy profile of the reactions, so it requires corrections for free energies in order to compare the activation barriers with experiments. Free energy corrections were derived from QM calculations of the uncatalyzed reaction in water rather than ribozyme environment using continuum solvent model. The continuum solvent models are however, parameter-dependent and introduce uncertainties in the estimation of free energy corrections. Another possible way how to calculate free energies in the context of full ribozyme is to construct the entire FES and obtain free energies directly. Computationally affordable semiempirical methods favored however, atypical and most likely incorrect reaction pathways, namely in the hairpin ribozyme the semi-empirical methods predict the sequential mechanisms, where the initial proton transfer was separated from the nucleophilic attack step, while the DFT and ab initio methods uniformly revealed concerted scenarios. Unfortunately, the complete construction of FES by DFT or ab initio approaches is still computationally demanding.

Altogether, the source of RNA catalytic power is still not widely understood and theoretical approaches (especially MD simulations combined with QM/MM calculations) are able to provide valuable structural insight and description of reaction mechanisms. The presented computational results provide complementary data against both experimental and theoretical observations. We also encountered certain limitations of utilized approaches, which indicated potential directions of further improvement.

Chapter 7

List of publications

1. Mlýnský V., Banáš P., Hollas D., Réblová K., Walter N.G., Šponer 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* 114(19), 6642-6652, (2010).
   http://pubs.acs.org/doi/abs/10.1021/jp1001258

2. Mlýnský V., Banáš P., Walter N.G., Šponer J., Otyepka M. QM/MM Studies of Hairpin Ribozyme Self-Cleavage Suggest the Feasibility of Multiple Competing Reaction Mechanisms. *J. Phys. Chem. B* 115(47), 13911-13924, (2011).
   http://pubs.acs.org/doi/abs/10.1021/jp206963g

3. Mlýnský V., Banáš P., Šponer J., van der Kamp M.W., Mulholland A.J., Otyepka M. Comparison of *Ab initio*, DFT, and Semiempirical QM/MM Approaches for Description of Catalytic Mechanism of Hairpin Ribozyme. *J. Chem. Theory Comput.* 10(4), 1608-1622 (2014).
   http://pubs.acs.org/doi/abs/10.1021/ct401015e

4. Mlýnský V., Walter N.G., Šponer J., Otyepka M., Banáš P. Deciphering the Role of Active Site Mg^{2+} Ion in the HDV Ribozyme Self-Cleavage: An Insight from QM/MM Calculations. *Phys. Chem. Chem. Phys.*, submitted.

Chapter 8

References

(1) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry, 5th edition*; W H Freeman: New York, 2002.

(2) Miller, B. G.; Wolfenden, R. *Annu. Rev. Biochem.* 2002, *71*, 847.

(3) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. *Chem. Rev.* 2006, *106*, 3210.

(4) Lilley, D. M. *Trends Biochem. Sci.* 2003, *28*, 495.

(5) Fedor, M. J.; Williamson, J. R. *Nat. Rev. Mol. Cell Biol.* 2005, *6*, 399.

(6) Doudna, J. A.; Lorsch, J. R. *Nat. Struc. Mol. Biol.* 2005, *12*, 395.

(7) Cochrane, J. C.; Strobel, S. A. *Acc. Chem. Res.* 2008, *41*, 1027.

(8) Fedor, M. J. *Annu. Rev. Biophys.* 2009, *38*, 271.

(9) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* 1981, *27*, 487.

(10) Guerriertakada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* 1983, *35*, 849.

(11) Gilbert, W. *Nature* 1986, *319*, 618.

(12) Harish, A.; Caetano-Anolles, G. *PLoS One* 2012, *7*.

(13) Lilley, D. M. J. *Philos. Trans. R. Soc. London, B* 2011, *366*, 2910.

(14) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* 1992, *256*, 1416.

(15) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* 2000, *289*, 920.

(16) Steitz, T. A.; Moore, P. B. *Trends Biochem. Sci.* 2003, *28*, 411.

(17) Scott, W. G. *Curr. Opin. Struct. Biol.* 2007, *17*, 280.

(18) Lonnberg, T. *Chem.--Eur. J.* 2011, *17*, 7140.

(19) Reymond, C.; Beaudoin, J. D.; Perreault, J. P. *Cell. Mol. Life Sci.* 2009, *66*, 3937.

(20) Symons, R. H. *Nucleic Acids Res.* 1997, *25*, 2683.

(21) Been, M. D. *Trends Biochem. Sci.* 1994, *19*, 251.

(22) Been, M. D.; Wickham, G. S. *Eur. J. Biochem.* 1997, *247*, 741.

(23) Salehi-Ashtiani, K.; Luptak, A.; Litovchick, A.; Szostak, J. W. *Science* 2006, *313*, 1788.

(24) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* 2004, *428*, 281.

(25) Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C. *Acc. Chem. Res.* 2011, *44*, 1270.

(26) Raines, R. T. *Chem. Rev.* 1998, *98*, 1045.

(27) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* 2004, *73*, 90.

(28) Young, K. J.; Gill, F.; Grasby, J. A. *Nucleic Acids Res.* 1997, *25*, 3760.

(29) Hertel, K. J.; Herschlag, D.; Uhlenbeck, O. C. *Biochemistry* 1994, *33*, 3374.

(30) Shih, I. H.; Been, M. D. *Biochemistry* 2000, *39*, 9055.

(31) McCarthy, T. J.; Plog, M. A.; Floy, S. A.; Jansen, J. A.; Soukup, J. K.; Soukup, G. A. *Chem. Biol.* 2005, *12*, 1221.

(32) Wilson, T. J.; McLeod, A. C.; Lilley, D. M. *EMBO J.* 2007, *26*, 2489.

(33) Yang, W.; Lee, J. Y.; Nowotny, M. *Mol. Cell* 2006, *22*, 5.

(34) Pyle, A. M. *Science* 1993, *261*, 709.

(35) Hampel, A.; Cowan, J. A. *Chem. Biol.* 1997, *4*, 513.

(36) Nesbitt, S.; Hegg, L. A.; Fedor, M. J. *Chem. Biol.* 1997, *4*, 619.

(37) Murray, J. B.; Seyhan, A. A.; Walter, N. G.; Burke, J. M.; Scott, W. G. *Chem. Biol.* 1998, *5*, 587.

(38) Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. *RNA* 2003, *9*, 907.

(39) Breaker, R. R.; Emilsson, G. M.; Lazarev, D.; Nakamura, S.; Puskarz, I. J.; Roth, A.; Sudarsan, N. *RNA* 2003, *9*, 949.

(40) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, 1991.

(41) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* 2006, *106*, 1737.

(42) Li, Y. F.; Breaker, R. R. *J. Am. Chem. Soc.* 1999, *121*, 5364.

(43) Oivanen, M.; Lonnberg, H. *J. Org. Chem.* 1989, *54*, 2556.

(44) Kosonen, M.; Hakala, K.; Lonnberg, H. *J. Chem. Soc., Perkin Trans. 2* 1998, 663.

(45) Mikkola, S.; Kosonen, M.; Lonnberg, H. *Curr. Org. Chem.* 2002, *6*, 523.

(46) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* 1997, *36*, 432.

(47) Warshel, A.; Florian, J. *Proc. Natl. Acad. Sci. U. S. A.* 1998, *95*, 5950.

(48) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferre-D'Amare, A. R. *Science* 2002, *298*, 1421.

(49) Lilley, D. M. J.; Eckstein, F. *Ribozymes and RNA catalysis*; RSC Biomolecular Sciences, 2008.

(50) Rupert, P. B.; Ferre-D'Amare, A. R. *Nature* 2001, *410*, 780.

(51) Salter, J.; Krucinska, J.; Alam, S.; Grum-Tokars, V.; Wedekind, J. E. *Biochemistry* 2006, *45*, 686.

(52) Ferre-D'Amare, A. R.; Scott, W. G. *Cold Spring Harbor Perspect. Biol.* 2010, *2*.

(53) Alam, S.; Grum-Tokars, V.; Krucinska, J.; Kundracik, M. L.; Wedekind, J. E. *Biochemistry* 2005, *44*, 14396.

(54) Lebruska, L. L.; Kuzmine, I. I.; Fedor, M. J. *Chem. Biol.* 2002, *9*, 465.

(55) Kuzmin, Y. I.; Da Costa, C. P.; Fedor, M. J. *J. Mol. Biol.* 2004, *340*, 233.

(56) Bevilacqua, P. C. *Biochemistry* 2003, *42*, 2259.

(57) Walter, N. G. *Mol. Cell* 2007, *28*, 923.

(58) Liu, L.; Cottrell, J. W.; Scott, L. G.; Fedor, M. J. *Nat. Chem. Biol.* 2009, *5*, 351.

(59) Nahas, M. K.; Wilson, T. J.; Hohng, S. C.; Jarvie, K.; Lilley, D. M. J.; Ha, T. *Nat. Struct. Mol. Biol.* 2004, *11*, 1107.

(60) Wilson, T. J.; Lilley, D. M. *RNA* 2011, *17*, 213.

(61) Nam, K. H.; Gao, J. L.; York, D. M. *J. Am. Chem. Soc.* 2008, *130*, 4680.

(62) Torelli, A. T.; Krucinska, J.; Wedekind, J. E. *RNA* 2007, *13*, 1052.

(63) Macelrevey, C.; Salter, J. D.; Krucinska, J.; Wedekind, J. E. *RNA* 2008, *14*, 1600.

(64) Torelli, A. T.; Spitale, R. C.; Krucinska, J.; Wedekind, J. E. *Biochem. Biophys. Res. Commun.* 2008, *371*, 154.

(65) Kuzmin, Y. I.; Da Costa, C. P.; Cottrell, J. W.; Fedor, M. J. *J. Mol. Biol.* 2005, *349*, 989.

(66) Cottrell, J. W.; Kuzmin, Y. I.; Fedor, M. J. *J. Biol. Chem.* 2007, *282*, 13498.

(67) Guo, M.; Spitale, R. C.; Volpini, R.; Krucinska, J.; Cristalli, G.; Carey, P. R.; Wedekind, J. E. *J. Am. Chem. Soc.* 2009, *131*, 12908.

(68) Lide, D. R. *Ed. CRC Handbook of Chemistry and Physics*; 83rd ed.; CRC, Boca Raton, 2003.

(69) Ditzler, M. A.; Sponer, J.; Walter, N. G. *RNA* 2009, *15*, 560.

(70) Cottrell, J. W.; Scott, L. G.; Fedor, M. J. *J. Biol. Chem.* 2011, *286*, 17658.

(71) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. *Nature* 1998, *395*, 567.

(72) Wrzesinski, J.; Legiewicz, M.; Smolska, B.; Ciesiolka, J. *Nucleic Acids Res.* 2001, *29*, 4482.

(73) Banas, P.; Rulisek, L.; Hanosova, V.; Svozil, D.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* 2008, *112*, 11177.

(74) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* 2000, *287*, 1493.

(75) Nakano, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* 2001, *123*, 11333.

(76) Ke, A. L.; Zhou, K. H.; Ding, F.; Cate, J. H. D.; Doudna, J. A. *Nature* 2004, *429*, 201.

(77) Krasovska, M. V.; Sefcikova, J.; Spackova, N.; Sponer, J.; Walter, N. G. *J. Mol. Biol.* 2005, *351*, 731.

(78) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. *Biophys. J.* 2006, *91*, 626.

(79) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* 2005, *1*, 45.

(80) Chen, J. H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* 2010, *49*, 6508.

(81) Golden, B. L. *Biochemistry* 2011, *50*, 9424.

(82) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* 2011, *115*, 8346.

(83) Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. Lett.* 2011, *2*, 2906.

(84) Prody, G. A.; Bakos, J. T.; Buzayan, J. M.; Schneider, I. R.; Bruening, G. *Science* 1986, *231*, 1577.

(85) Pley, H. W.; Flaherty, K. M.; Mckay, D. B. *Nature* 1994, *372*, 68.

(86) De la Pena, M.; Gago, S.; Flores, R. *EMBO J.* 2003, *22*, 5561.

(87) Martick, M.; Scott, W. G. *Cell* 2006, *126*, 309.

(88) Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. *Nucleic Acids Res.* 1997, *25*, 3421.

(89) Torres, R. A.; Himo, F.; Bruice, T. C.; Noodleman, L.; Lovell, T. *J. Am. Chem. Soc.* 2003, *125*, 9861.

(90) Leclerc, F. *Molecules* 2010, *15*, 5389.

(91) Osborne, E. M.; Schaak, J. E.; DeRose, V. J. *RNA* 2005, *11*, 187.

(92) Thomas, J. M.; Perrin, D. M. *J. Am. Chem. Soc.* 2008, *130*, 15467.

(93) Thomas, J. M.; Perrin, D. M. *J. Am. Chem. Soc.* 2009, *131*, 1135.

(94) Nelson, J. A.; Uhlenbeck, O. C. *RNA* 2008, *14*, 605.

(95) Wong, K. Y.; Lee, T. S.; York, D. M. *J. Chem. Theory Comput.* 2011, *7*, 1.

(96) Klein, D. J.; Ferre-D'Amare, A. R. *Science* 2006, *313*, 1752.

(97) Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. *Chem. Biol.* 2007, *14*, 97.

(98) Klein, D. J.; Been, M. D.; Ferre-D'Amare, A. R. *J. Am. Chem. Soc.* 2007, *129*, 14858.

(99) Klein, D. J.; Wilkinson, S. R.; Been, M. D.; Ferre-D'Amare, A. R. *J. Mol. Biol.* 2007, *373*, 178.

(100) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* 2010, *114*, 8701.

(101) Viladoms, J.; Scott, L. G.; Fedor, M. J. *J. Am. Chem. Soc.* 2011, *133*, 18388.

(102) Viladoms, J.; Fedor, M. J. *J. Am. Chem. Soc.* 2012, *134*, 19043.

(103) Lilley, D. M. J. *RNA* 2004, *10*, 151.

(104) Lipfert, J.; Ouellet, J.; Norman, D. G.; Doniach, S.; Lilley, D. M. J. *Structure* 2008, *16*, 1357.

(105) Lafontaine, D. A.; Wilson, T. J.; Norman, D. G.; Lilley, D. M. J. *J. Mol. Biol.* 2001, *312*, 663.

(106) Lafontaine, D. A.; Wilson, T. J.; Zhao, Z. Y.; Lilley, D. M. J. *J. Mol. Biol.* 2002, *323*, 23.

(107) Wilson, T. J.; Li, N. S.; Lu, J.; Frederiksen, J. K.; Piccirilli, J. A.; Lilley, D. M. J. *Proc. Natl. Acad. Sci. U. S. A.* 2010, *107*, 11751.

(108) Banas, P.; Jurecka, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *Methods* 2009, *49*, 202.

(109) Sponer, J.; Spackova, N. *Methods* 2007, *43*, 278.

(110) Leach, A. R. *Molecular Modelling: Principles and Applications*; 2nd ed.; Pearson Education: Harlow, UK, 2001.

(111) Sponer, J.; Sponer, J. E.; Mladek, A.; Banas, P.; Jurecka, P.; Otyepka, M. *Methods* 2013, *64*, 3.

(112) Grimme, S. *J. Chem. Phys.* 2003, *118*, 9095.

(113) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. *Chem. Rev.* 2012, *112*, 289.

(114) Grimme, S. *J. Comput. Chem.* 2004, *25*, 1463.

(115) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chem. Rev.* 2010, *110*, 5023.

(116) Grimme, S. *Wires Comput. Mol. Sci.* 2011, *1*, 211.

(117) Becke, A. D. *Phys. Rev. A* 1988, *38*, 3098.

(118) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* 1988, *37*, 785.

(119) Becke, A. D. *J. Chem. Phys.* 1993, *98*, 5648.

(120) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* 1996, *77*, 3865.

(121) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* 2008, *120*, 215.

(122) Klimes, J.; Bowler, D. R.; Michaelides, A. *J. Phys-Condens. Mat.* 2010, *22*.

(123) Granatier, J.; Lazar, P.; Otyepka, M.; Hobza, P. *J. Chem. Theory Comput.* 2011, *7*, 3743.

(124) Cooper, V. R.; Thonhauser, T.; Langreth, D. C. *J. Chem. Phys.* 2008, *128*.

(125) Lazar, P.; Karlicky, F.; Jurecka, P.; Kocman, M.; Otyepkova, E.; Safarova, K.; Otyepka, M. *J. Am. Chem. Soc.* 2013, *135*, 6372.

(126) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* 2010, *132*.

(127) Nam, K.; Cui, Q.; Gao, J. L.; York, D. M. *J. Chem. Theory Comput.* 2007, *3*, 486.

(128) Porezag, D.; Frauenheim, T.; Kohler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* 1995, *51*, 12947.

(129) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* 1998, *58*, 7260.

(130) Deuflhard, P.; Hermans, J.; Leimkuhler, B.; Mark, A. E.; Reich, S.; Skeel, R. D. *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer: Berlin, 1997.

(131) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* 1983, *4*, 187.

(132) Lamoureux, G.; Roux, B. *J. Chem. Phys.* 2003, *119*, 3025.

(133) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. *Chem. Phys. Lett.* 2006, *418*, 245.

(134) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* 1995, *117*, 5179.

(135) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* 2010, *43*, 40.

(136) Mlynsky, V.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Otyepka, M.; Sponer, J.; Banas, P. *in preparation* 2014.

(137) Case, D. A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A.W., Kolossváry, I., Wong, K.F., Paesani, F., Vanicek, J., Wolf,

R.M., Liu, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D.R., Mathews, D.H., Seetin, M.G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, P.A. *AMBER 12.0*; University of California: San Francisco, 2012.

(138) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* 1993, *98*, 10089.

(139) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* 2000, *21*, 1049.

(140) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* 2007, *92*, 3817.

(141) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P. *J. Chem. Theory Comput.* 2011, *7*, 2886.

(142) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* 2004, *32*, 4269.

(143) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* 2010, *114*, 6642.

(144) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* 2010, *6*, 3836.

(145) Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P. *J. Chem. Theory Comput.* 2013, *9*, 2339.

(146) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* 2000, *21*, 86.

(147) MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* 2000, *21*, 105.

(148) Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* 2008, *36*, 2379.

(149) Pan, Y. P.; MacKerell, A. D. *Nucleic Acids Res.* 2003, *31*, 7131.

(150) Reblova, K.; Fadrna, E.; Sarzynska, J.; Kulinski, T.; Kulhanek, P.; Ennifar, E.; Koca, J.; Sponer, J. *Biophys. J.* 2007, *93*, 3932.

(151) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. *J. Comput. Chem.* 2011, *32*, 1929.

(152) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. *J. Phys. Chem. B* 2012, *116*, 9899.

(153) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* 2005, *105*, 2999.

(154) Warshel, A.; Levitt, M. *J. Mol. Biol.* 1976, *103*, 227.

(155) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed. Engl.* 2009, *48*, 1198.

(156) Maseras, F.; Morokuma, K. *J. Comput. Chem.* 1995, *16*, 1170.

(157) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* 1996, *100*, 19357.

(158) Morokuma, K.; Wang, Q. F.; Vreven, T. *J. Chem. Theory Comput.* 2006, *2*, 1317.

(159) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* 2001, *114*, 9758.

(160) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. *Chem. Soc. Rev.* 2012, *41*, 3025.

(161) Skodje, R. T.; Truhlar, D. G. *J. Phys. Chem.* 1981, *85*, 624.

(162) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* 1998, *286*, 243.

(163) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* 1999, *302*, 437.

(164) Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* 2002, *365*, 89.

(165) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* 2003, *125*, 15608.

(166) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* 2000, *104*, 4811.

(167) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* 2001, *105*, 2936.

(168) Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. *J. Phys. Chem. B* 2006, *110*, 11525.

(169) Chval, Z.; Chvalova, D.; Leclerc, F. *J. Phys. Chem. B* 2011, *115*, 10943.

(170) Cheatham, T. E. *Curr. Opin. Struct. Biol.* 2004, *14*, 360.

(171) Auffinger, P.; Hashem, Y. *Curr. Opin. Struct. Biol.* 2007, *17*, 325.

(172) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* 2007, *85*, 169.

(173) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* 1996, *118*, 6811.

(174) Murray, L. J. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. *Proc. Natl. Acad. Sci. U. S. A.* 2003, *100*, 13904.

(175) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* 2008, *14*, 465.

(176) Sponer, J.; Cang, X. H.; Cheatham, T. E. *Methods* 2012, *57*, 25.

(177) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* 2014, *136*, 1483.

(178) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* 2003, *107*, 8669.

(179) Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* 2013, *52*, 6499.

(180) Gong, B.; Chen, J. H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. *J. Am. Chem. Soc.* 2007, *129*, 13335.

(181) Izatt, R. M.; Hansen, L. D.; Rytting, J. H.; Christensen, J. J. *J. Am. Chem. Soc.* 1965, *87*, 2760.

(182) Acharya, S.; Foldesi, A.; Chattopadhyaya, J. *J. Org. Chem.* 2003, *68*, 1906.

(183) Jarvinen, P.; Oivanen, M.; Lonnberg, H. *J. Org. Chem.* 1991, *56*, 5396.

(184) Usher, D. A.; Richardson, D. I., Jr.; Oakenfull, D. G. *J. Am. Chem. Soc.* 1970, *92*, 4699.

(185) Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M. *J. Am. Chem. Soc.* 2002, *124*, 6594.

(186) Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* 2000, *122*, 166.

(187) Dahm, S. C.; Derrick, W. B.; Uhlenbeck, O. C. *Biochemistry* 1993, *32*, 13040.
(188) Fedor, M. J. *J. Mol. Biol.* 2000, *297*, 269.

Chapter 9

Results – Appendix