CHAPTER 6

Molecular Dynamics Simulations of RNA Molecules

J. ŠPONER,*[a] M. OTYEPKA,[b] P. BANÁŠ,[b] K. RÉBLOVÁ[a] AND N. G. WALTER[c]

[a] Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic; [b] Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, Olomouc, Tr. 17. Listopadu 12, 771 46 Olomouc, Czech Republic; [c] Department of Chemistry, Single Molecule Analysis Group, University of Michigan, 930 N. University Ave., Ann Arbor, MI 48109-1055, USA
*Email: sponer@ncbr.chemi.muni.cz

6.1 Molecular Simulations of RNA: General Considerations

The central role of RNA in numerous biological processes including translation,[1] protein localization,[2] gene regulation,[3] RNA processing,[4] and viral replication[5] calls for a detailed understanding of RNA function, structure, and conformational dynamics.[6] Accompanying and enhancing our increasing appreciation of RNA is the rapidly expanding availability of high-resolution structures of RNAs and RNA-protein (RNP) complexes. These atomic resolution snapshots provide detailed rationalization for existing biochemical data. However, biological function depends on the dynamic evolution of structures along functional pathways. A complete understanding of the

RSC Biomolecular Sciences No. 24
Innovations in Biomolecular Modeling and Simulations: Volume 2
Edited by Tamar Schlick

Published by the Royal Society of Chemistry, www.rsc.org

relevant structural dynamics exhibited by RNA requires monitoring timescales from picoseconds to hours through the application of a correspondingly broad range of techniques,[6] with careful consideration given to the scope and limitation of each approach.

Provided they are judiciously applied, computational methods can complement experimental techniques and provide insights that are not accessible by experiment.[7,8] While reproduction of experimental data is desirable for assessing calculation accuracy, the main goal of computations is to obtain new insights that lead to experimentally testable predictions. Molecular dynamics (MD) simulations can identify problematic aspects of experimental structures,[8–12] reveal functionally significant stochastic fluctuations and molecular flexibility,[8,13,14] predict the impact of base substitutions, modifications and ionization on RNA structure and dynamics,[12,15] and characterize solvent behavior and monovalent ion binding.[9,12,16,17] Combining simulations with quantum mechanical (QM) calculations in QM/MM approaches expands the repertoire of applications to mechanistic questions concerning the reaction chemistry of catalytic RNAs (ribozymes).[18–22]

However, despite their popularity, computational methods are accompanied by numerous limitations, which must be taken into account for the successful interpretation of the results. Inexpert use of computational methods, not respecting their genuine limitations, hurts the long-term reputation of computational chemistry. Thus, in this chapter, we pay a substantial attention to the limitations of the computation methods, namely to limitations of molecular mechanical force fields that are used in explicit-solvent molecular dynamics simulation studies. Sections 2–4 introduce the general scope of the RNA molecular dynamics field with detailed description of the main limitations, namely the force field artifacts and influence of the starting structure. We are mainly focused on the AMBER force field family and explain the latest efforts to tune its performance for nucleic acids. This on one hand reflects our own experience with this force field family but also the current balance in the RNA simulation literature where variants of the original pair-additive Cornell *et al.* force field dominated in the past.[23] This force field has been broadly used by many groups for RNA simulations, including studies aimed at extensive testing, and until now it appears to perform best for RNA molecules. The last two sections briefly comment on the scope and limitations of QM and QM/MM methods that are frequently combined with MD techniques to address the chemical reaction in ribozymes. We also recommend the beginners to read the most recent reviews about applications of the computational methods to nucleic acid research.[7,8,24–34]

6.2 The General Scope and Limitations of MD Simulations

When assessing the outcome of simulations, it is important to understand what exactly the technique does. Explicit solvent MD is an atomistic single-molecule technique dealing with a solute molecule or molecular complex in aqueous

solution. The studied system possesses an exact starting geometry (set of atomic coordinates), which usually originates from X-ray crystallography or NMR studies. Its quality critically affects the subsequent simulations. The biomolecular system surrounded by an environment of water and ions then undergoes 1–1000+ ns of dynamics simulated at ambient temperature and pressure starting from the initial geometry. The method thus mimics the genuine thermal fluctuations of the initial structure. Next, we are trying to deduce from these simulation runs (which are still very short compared with most real dynamics) useful information about the studied molecule.

The greatest advantage of the simulation technique is the unsurpassed level of detail of all aspects of the time evolution (with sub-ps time resolution) of the three-dimensional structure, including the positions of all water molecules (including their hydrogens) and ions. However, MD simulations are faced with two significant limitations. First, the sampling of conformational space reflects the short timescale of MD compared to biochemically relevant biomolecular dynamics. This limitation is slowly being overcome with the continuing emergence of more powerful computers and algorithms. Second, a fundamental limitation not waning with faster computers is the approximate nature of the present biomolecular force fields, which are simple, atomistic analytical functions relating structure with potential energy.[23,35] They assign potential energy to each single Cartesian XYZ geometry of the system. Despite sophisticated parameterization, such a force field consists of sets of harmonic springs for both bond lengths and valence angles, supplemented by torsion profiles for dihedral angles.[23,35–37] Atoms are approximated as Lennard-Jones spheres with constant point charges localized at the atomic centers.[36,38] Such a description trivially suppresses covalent bond breaking and making, which constrains a studied system in its chemical starting state, including the protonation states of specific acid and base moieties. In addition, the description of non-covalent interactions is obviously physically incomplete with many effects (such as all types of polarization and charge transfer effects) neglected by definition. Since force fields neglect all non-additive contributions, they are called ''effective pair-additive force fields''. The neglected contributions must be included in the force fields indirectly, *i.e.*, they are somehow effectively implemented by parameters of those force field terms that are explicitly used.[39,40]

Probably the most difficult part of the parameterization of current pair-additive force fields (see also below) is that of the torsion profiles. Bond and angle parameters can be derived from structural data, IR and microwave spectroscopy, and/or high level QM. There are relatively straightforward procedures or protocols available to determine inter-molecular parameters, *i.e.*, van der Waals radii, well-depths and partial charges. Van der Waals radii and well-depths can be derived by matching experimental densities and heats of vaporization, whereas atomic charges can be parameterized through fits to QM-derived electrostatic potentials or energetics. Fitting of the torsional parameters is difficult and their actual physical meaning is not clearly defined. Rather than being related to real electronic structure contributions, they represent *ad hoc* functions used for the ultimate tuning of force field behavior.

There are two widely used nucleic acid (NA) force fields, Cornell *et al.*[23] (known also as AMBER, albeit the AMBER program suite contains also other force fields) and CHARMM27,[35] which share similar functional form but differ in parameterization.[41] Note that parameterization of NA force fields is probably more challenging than parameterization of protein force fields due to a more complex balance of forces in nucleic acids. It thus appears that the performance of NA force fields is not as matured yet as that of protein force fields (although even the description of protein force fields is far from perfect and different protein force fields are biased towards different types of protein structures).[42–47] We strongly discourage utilization of any force field for nucleic acids that has not been carefully validated. Variants of the Cornell *et al.* (AMBER) force field (parm94,[23] parm98,[48] parm99[49] and the latest parmbsc0[37] and parmOL[50,51]) have been extensively tested for many folded RNAs and non-canonical DNAs.[9–11,14–17,28–30,37,52,53] CHARMM27 describes B-DNA well,[54] but has not yet been systematically tested for either folded RNAs or non-canonical DNAs. While some studies report successful application of the CHARMM27 force field to RNA, these studies usually do not provide a sufficiently detailed analysis to verify the force field. In addition, several other studies reported rather unstable A-RNA trajectories, casting doubt on the applicability of the CHARMM27 force field in its present form to RNA. The main problem of the CHARMM27 simulations of RNA is evidently an underestimated stability of base pairs in the A-RNA duplexes in the simulations, which leads to anomalously frequent base pair breathing and subsequent large-scale stem disruptions.[50,52,55,56] Most likely, tuning of this force field would be possible and may bring basic stability to the simulations. In fact, the CHARMM RNA force field has been substantially revised in 2011 by complete reparametrization of the dihedral parameters dictating the structure/energy relation of the 2′-hydroxyl proton of the ribose, which was evidently inadequate in the preceding version of the force field.[57] Both AMBER and CHARMM offer high-quality protein force fields for a consistent description of NA-protein complexes. A few years ago, the GROMOS nucleic acids force field was released.[58] This force field, however, does not contain parameters for RNA, and even for B-DNA it yields very unstable trajectories.[59]

QM calculations show that base stacking is the best approximated term in NA simulations when using the AMBER force fields,[60] followed by base pairing, including non-Watson-Crick interactions utilizing the 2′-OH group (Figure 6.1).[61,62] Some other interactions, such as the highly conserved base-phosphate H-bonds may be more difficult to describe and the force field is yet to be tested (Figure 6.1).[63] The Lennard-Jones form of the van der Waals term dominating in the stacking interaction may be non-optimal, so that a potentially better description could be obtained by a physically meaningful exponential repulsion term.[40,64] It is, however, difficult to evaluate if changing the van der Waals form would dramatically affect force field performance. In contrast, the partial charges included in the electrostatics term of the Cornell *et al.* force fields, which are derived to reproduce the electrostatic potential around the NA building blocks, seem to be the best choice within the

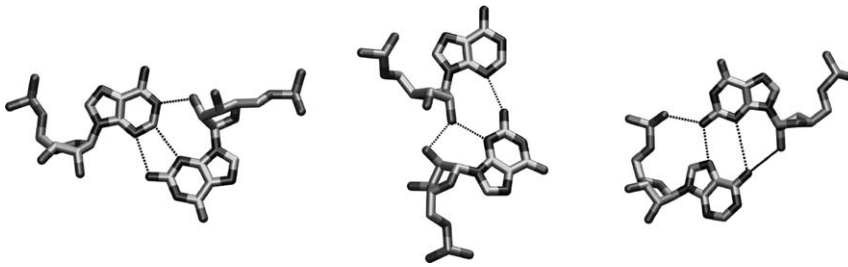

Figure 6.1 Folded RNA molecules are stabilized by numerous non-canonical interactions. A prominent role in non-canonical RNA base pairing is played by very diverse base pairing patterns directly involving the 2′-OH hydroxyl group, such as seen in the trans sugar-edge sugar-edge GA base pair (left) and cis sugar-edge sugar-edge AG base pair (middle). RNA base pairing can be further extended by direct, highly sequence-specific and conserved base-phosphate hydrogen bonds (right), which involve ~12% of ribosomal nucleotides.[63] Despite their complexity, all these non-canonical interactions appear to be rather well-described by the Cornell *et al.* force field model.

approximation of constant atomic charges.[36,65] Note, however, the partial charges *per se* do not have any physical meaning, they are adjustable parameters of the force field and do not correspond to any observable property of the system (*i.e.*, there is no quantum-chemical operator corresponding to point charges).[33] The description of the flexible backbone is less straightforward.[37,66] In particular, the anionic phosphate group is highly polarizable and the many dihedral angles of the backbone conformation may adopt multiple sub-states with different combinations of the individual angles. The backbone description would therefore certainly benefit from geometry-dependent electrostatic and polarization terms.

Ions are simplified as Lennard-Jones spheres with constant point charges at their centers. Monovalent ions and solute-solvent interactions are thought to be reasonably well described, while the description of divalent ions is outside the applicability of force fields. This originates from the fact that the total amount of non-additive effects (neglected by the force field) in the first ligand shell of divalent metal ions is worth around 70 kcal/mol.[67,68] This is an ~14-times larger energy effect than the gas phase binding interaction energy of a water dimer. The first-shell ligands are thus highly polarized, which leads to inter-ligand polarization repulsion (anti-cooperativity) within this first shell around the cation. In turn, the activated polarized first-shell ligands are capable of forming very strong hydrogen bonds outside the first shell, which is an example of cooperative non-additive effects. Properties of a water molecule bound to a divalent cation therefore differ significantly from properties of a common bulk water molecule.[67,68] All these contributions are neglected by the force field, which is why we usually do not recommend to include divalent ions into NA simulations, although such simulations may at first sight (when ignoring the force field limitations) look more biochemically relevant. However, even

monovalent ions in combination with an incorrect usage of combination rules and ionic parameters may lead to significant simulation artifacts like crystallization of salts, *i.e.*, spurious pairing of anions and cations in excess salt simulations.[69–71] Note, however, that in contrast to proteins, NA simulations can be performed using a net-neutral cation atmosphere, since NAs have a charge of -1 per phosphate so that net-neutralization gives a reasonable concentration of $\sim$0.15 M cations. Thus, there is no strong reason for NA simulations with excess salt[72–74] and there has been no convincing report yet showing a substantial difference between net-neutral and excess-salt simulations for NA (numerous such comparisons can be found in our recent studies).[72–74] We find that both net-neutral and low-excess salt ion conditions are viable and close to equivalent. Opposite claims occasionally found in the literature (without showing, however, any convincing comparison of a representative set of simulations) should be disregarded. It does not mean, of course, that the description of ions is perfect. It means that the results are not dramatically sensitive to the specific details of the ion treatments and all ion treatments suffer from similar inaccuracies.

It is also notable that, although there are diverse parameter sets for monovalent ions available in the literature for use in pair-additive force fields, the actual differences among them (as changes in well depth and radius compensate for each other) are rather small, all deviating from reference QM computations in the same manner.[75]

The quality of a force field's performance inherently relies on the mutual compensation of errors, which in turn depends on a balance between forces in the system under study and the accuracy and completeness of the parameterization. There are two basic scenarios for what may happen during the course of a simulation: (i) The compensation of errors is sufficient and the force field finds the correct global minimum of the simulated system. In this case, not all details are necessarily correct, but the overall description is meaningful. The more qualitative the computational task, the more likely the force field description is sufficient. (ii) The force field does not give the correct global minimum and then the simulated system eventually degrades.[25,37,76,77] The degradation may be visible within a few nanoseconds or it may be entirely hidden over the computationally affordable timescale. The latter occurs when the starting structure is correct and separated by a sufficiently large energy barrier from incorrect and degraded conformations. Assuming first order kinetics and the applicability of the Eyring equation, the maximum free energy barrier that can be overcome within a simulation time Δt in ns with at least 90% probability is $\Delta G^{\ddagger} \approx 4.7 + 1.37 \cdot \log(\Delta t)$ kcal/mol (Δt is in ns). If the force field does not correctly predict the global minimum, it can still often be used to satisfactorily derive certain properties of known structures. However, any attempts (despite their frequency in the literature) to combine in this case computational methods with enhanced sampling (such as replica exchange MD) to, for example, fold an entire NA (e.g., RNA hairpin) molecule are futile.

Biomolecular force fields are intentionally parameterized as multipurpose, with delicate trade-offs in parameterization. Tremendously challenging efforts

are being expended to develop more physically accurate multipurpose polarization force fields.[68,78–82] Major problems in parameterization of these sophisticated biomolecular force fields are to achieve a satisfactory overall balance between all their parameters (*i.e.*, if a force field is too sophisticated and requires too many parameters, it can be difficult to balance) and to avoid parameter-correlation artifacts. The ultimate utility of future multipurpose polarization force fields is far from evident at the moment. Thus, it is well possible that we will have to rely for some time on the present generation of biomolecular force fields.

6.3 Refinement of the Cornell *et al.* Force Field for Nucleic Acids: From Parm94 to Parmbsc0 and ParmOL

The performance of a force field can be well illustrated with the continuous refinement of the Cornell *et al.* NA force field. For a long time, the original parm94 Cornell *et al.* force field was assumed to give sufficiently good description of B-DNA although, for example, the helical twist was known to be notoriously underestimated compared to X-ray crystallography data as well as appropriate NMR experiments.[83] (Note that the apparently lower helical twist in many NMR B-DNA structures is due to limitations of the experiment and the refinement protocol. Once an accurate NMR experiment is used there is no discrepancy between the B-DNA helical twists in crystal and NMR structures.)[83] Minor refinements of the force field considering subtle tuning of pucker and χ parameters were attempted. The resulting force fields, known as parm98 and parm99, respectively, perform rather similarly to parm94. Larger problems were, however, reported in simulations of guanine quadruplex DNA (G-DNA) that consists of four-stranded stems formed by cation-stabilized guanine quartets complemented by single-stranded hairpin loops. The parm99 force field provides a global minimum consistent with the experimental structures for the G-DNA stems, but not for the loops since their structure degrades in sufficiently long simulations.[75,76] Thus, in a given simulation different parts of a molecule can be described with varying success. In this particular case, the G-DNA stem is described well while the loops are not. The G-DNA loop problem may, at first sight, appear as a relatively marginal problem since it concerns non-canonical DNA segments. However, the situation dramatically changed within less than a year. Entirely unanticipated problems were detected in B-DNA simulations once slightly longer (15–50+ ns) MD simulations of B-DNA became available starting in $\sim$2004.[84,85] Specifically, there occurred an accumulation of irreversible, experimentally unobserved backbone sub-states with concomitant progressive degradation of the entire structure. This degradation, however, was not openly reported in these first studies, so readers were left in the dark regarding the significant magnitude of the problem. The main feature of these sub-states was a γ-trans geometry of the backbone (Figures 6.2–6.4). Such γ-trans topologies can sometimes occur in DNA-protein complexes but should not occur in naked B-DNA.

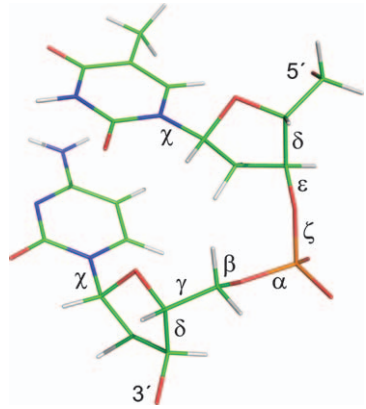

Figure 6.2 The basic nomenclature of dihedral angles of the nucleic acid backbone.

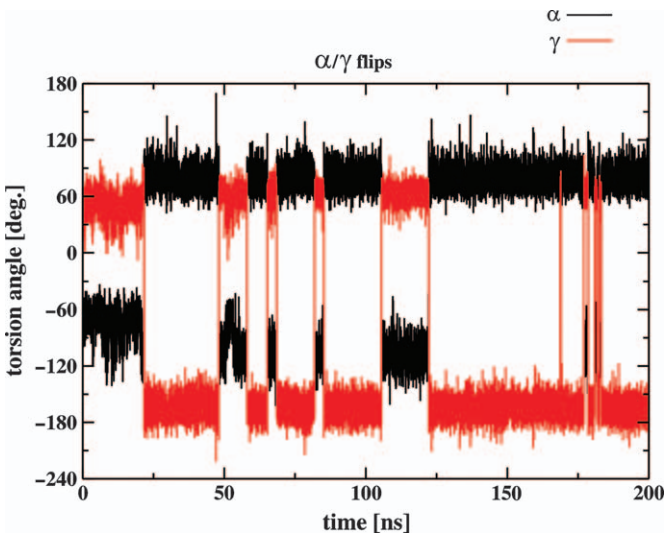

Figure 6.3 An example of the α/γ phosphate flip from canonical gauche-/gauche+ to artificial gauche+/trans in simulations lacking the parmbsc0 reparametrization. The backbone might sample both native and artificial γ-trans state for some time, but it finally always irreversibly flips into the γ-trans state.

The profile of α and γ backbone torsions was subsequently substantially reparametrized, leading to parmbsc0.[37] This improved force field allows stable microsecond-timescale simulations of B-DNA and even repairs partially degraded B-DNA structures, indicating that B-DNA is now the global minimum.[53] Fortunately, the γ-trans degradation has never substantially affected RNA simulations, as the γ-trans sub-states in RNA turn out to be reversible

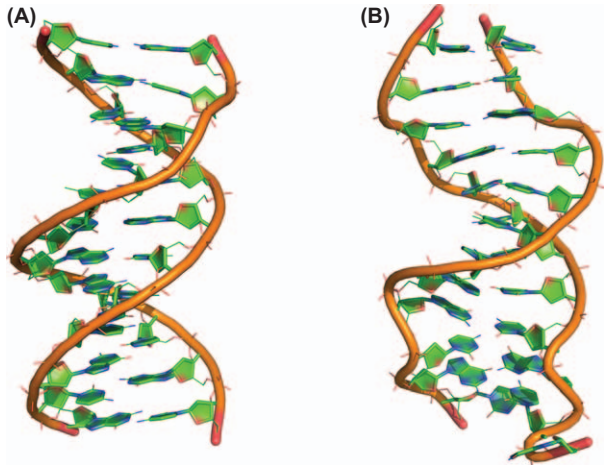

Figure 6.4 (A) The crystal structure of Dickerson dodecamer, and (B) the same dodecamer structure after 120 ns of MD simulation using the parm99 force field. Ten of the twenty four γ torsions are already in the trans conformation. The structure is visibly distorted and under-twisted. Ultimately, all nucleotides would adopt the γ-trans conformation.

and correspond to minor sub-states observed in experimental RNA structures (Figure 6.5).[86]

Initially, it was not clear whether the parmbsc0 would bring substantial improvement for RNA (whereas for DNA it must be used). All force field variants were assumed to provide the correct global minimum for A-RNA with a rather similar detailed performance.[72] Ultimately, however, a major degradation was discovered also for RNA structures, occurring with all these force field variants. On sufficiently long timescales (typically 50–100+ ns), short A-RNA molecules undergo irreversible, entirely degrading transitions to a meaningless ladder-like structure characterized by shifting the glycosidic torsion angle χ from the *anti* to the high-*anti* region (Figure 6.6).[77]

Fortunately, this artifact was ultimately eliminated by a one-dimensional (1D) reparameterization of the χ torsion angle using highest-quality QM calculations and considering the influence of solvent on both the QM and MM components of the procedure.[51] This force field variant is known as parmOL.[50,51] Besides eliminating the ladder-like degradation, the parmOL force field improves the *syn* region and the *syn–anti* balance. For RNA, the best performance is achieved when combining parmOL with parmbsc0. The parmOL should not be applied to DNA as it is not possible to simultaneously fix the χ profile for DNA and RNA by a 1D reparametrization, *i.e.*, without simultaneously modifying some other force field terms. Thus, presently the best option is parmbsc0 for DNA and parmbsc0+OL for RNA, as these are the only force fields preventing massive degradation on longer time scales (currently tested up to the microsecond timescale). The first degrading flips in B-DNA (when the parmbsc0 parameters are not applied) can occur even during

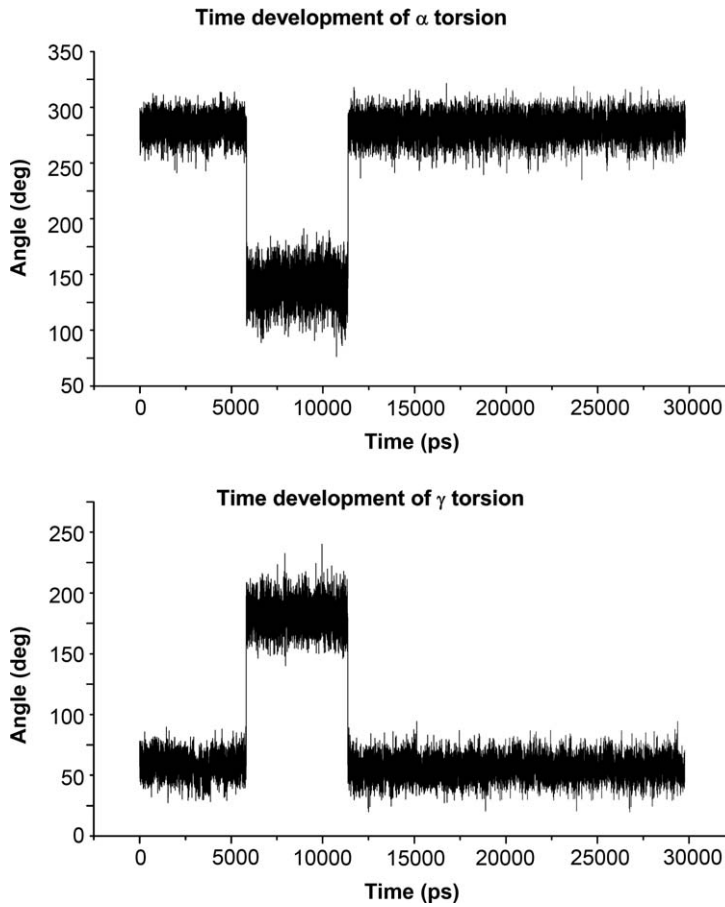

Figure 6.5 Characteristic correlated time development of α and γ dihedral angles in AMBER Cornell *et al.*-based RNA simulations without using the parmbsc0 correction. The starting α/γ combination corresponds to canonical A-RNA. The γ-trans flips are entirely reversible and typically short-lived. Thus, in contrast to DNA, γ-trans flips do not degrade RNA. With parmbsc0, the γ-trans flips are essentially suppressed entirely.[72]

equilibration while the onset of high-*anti* χ RNA transitions (when the parmOL parameters are not applied) is slower, depending on the simulated system. However, it was observed that there are specific systems, such as reverse Kink-turn motif, which cannot be investigated without the parmOL force field even on pretty short time scale.[87]

It should be noted that even after the most recent adjustments, force fields are far from perfect. It is likely that an increased sampling, expected from an improvement in computational hardware and algorithms, may in the future uncover additional force field artifacts that are currently hidden at (sub)-microsecond simulation timescales due to a high free energy barrier separating

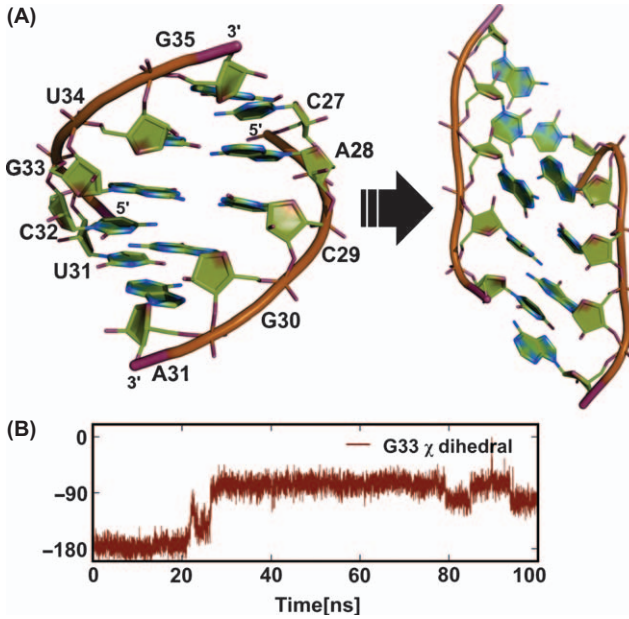

Figure 6.6 (A) Formation of ladder-like structure observed on a tens-of-nanosecond timescale during simulation of a very short A-RNA duplex. (B) The corresponding shift of the glycosidic torsion to *high*-anti conformation. The ladder–like structure is the global minimum ultimately predicted by all variants of the Cornell *et al.* (AMBER) force field without applying the latest χ_{OL} glycosidic torsion Reparameterization.[50,51] Unless χ_{OL} is applied, essentially all RNA structures would ultimately be degraded in sufficiently long simulations.

the starting native state and the potential artificial state. The G-DNA loop problem has not been solved and, even though the description of UNCG and GNRA tetraloops in RNA has been visibly improved, it is likely that their simulation behavior is still not perfect. Single-stranded hairpin loops are especially difficult targets for simulations due to the complex balance of various energy contributions shaping them. In studies of RNAs, one has to be especially careful with simulation methods attempting to enhance sampling, such as replica-exchange molecular dynamics, locally enhanced sampling and targeted MD. Such methods, on the one hand, are obviously highly desirable and useful. On the other hand, however, their application is not unproblematic. First, these methods are always based on (significant) approximations additional to those used in standard unrestrained simulations.[88] For example, forcing a conformational change through a drastic time-dependent root-mean-square distance (RMSD) penalty function in targeted MD is not the same as to observe the change spontaneously during unrestrained simulations. Targeted MD should be safely applicable to simple pathways such as the A-to-B-DNA transition,[89] but more complex changes may easily go beyond the applicability

of the method. The approximations underlying all enhanced sampling methods are not always fully acknowledged and respected in the literature, sometimes making it challenging for the non-expert to fully understand the significance of the computations and results presented in a publication. There is no silver bullet for solving the sampling problem that would be penalty-free. These otherwise highly valuable methods in no case can replace conventional, unrestrained simulations that remain the gold standard. The applicability of sophisticated enhanced sampling methods, which are specifically designed to overcome energy barriers and to ultimately allow RNA folding, obviously requires that the force field yields the correct global minimum. Thus, although a series of recent simulation papers reported successful folding of short RNA stem-loop hairpin systems,[90] we suggest that none of these studies folded the RNAs to their native structures since all were done with force fields that do not reproduce the native structures as global minima.[50] Closer inspection of the published data in fact confirms that. Note that GNRA and especially UNCG tetraloops in RNA are precisely structured, recurrent RNA motifs that have clearly defined 3D shapes with salient signature interactions that have been strictly conserved in the course of evolution.[31,91,92] These 3D shapes ought to be dominantly sampled by any suitable simulation method with a sufficiently accurate force field, which is not the case for the preceding RNA tetraloop simulation studies.

Nevertheless, despite the remaining limitations, the performance of the refined AMBER Cornell *et al.* force field for RNA is remarkable considering its striking simplicity. Given the difficulties in the development of polarization force fields noted above and the expensive nature of sufficiently accurate QM descriptions, it is likely that refined pair-additive force fields will still dominate NA simulation studies over the next decade. As noted above, local conformational traps associated with standard MD simulations can be overcome by enhanced sampling techniques such as locally enhanced sampling, replica exchange, or targeted MD.[29,76] Broad sampling is also critical to obtain reliable results from free energy calculations that can provide useful information on the thermodynamics of RNA conformations, but their in-depth assessment is beyond the scope of our review. It always needs to be considered that sophisticated methods that aim to overcome limitations of standard unrestrained simulations may be prone to over-interpretation.

6.4 MD Simulations and the Starting Structure

As pointed out above, MD relies on the availability of accurate high-resolution structures. If a reasonably accurate starting structure is available, MD in many cases can locally improve molecular interactions and backbone conformations in the experimental structure.[93] Due to force field and sampling limitations MD is unable to predict RNA structures without experimental input.[94] Additionally, unrealistic models often become swiftly distorted during a simulation, revealing their inadequacies. If the starting structure is in an incorrect

conformation confined by large (>5–7 kcal/mol) energy barriers, an MD simulation cannot easily move it away from its starting geometry.[52] Reliable characterization of the dynamics of an RNA therefore requires the use of high-resolution experimental starting structures. After collecting enough simulations (current state-of-the-art is multiple simulations of ~20–250+ ns duration each),[52] a careful comparison of the MD time trajectories with the experimental structure is needed. A simplified analysis of only few hetero-atom distances of interest accompanied by generally uninformative RMSD plots may mask considerable problems. The simulation behavior results from a mixture of factors, including the actual stochastic flexibility of the RNA, experimental artifacts introduced through crystal contacts,[10] disorder or chemical modification, and force field artifacts. If this mixture is properly resolved, the analysis of MD simulations can be very insightful.

Evaluation of RNA backbone conformations is difficult for both computational and experimental approaches. The flexibility and polarizability of the backbone are challenging for non-polarizable force fields that are based on constant point charges. In addition, while phosphates and bases are rather well defined by electron densities, sugar atoms and thus details of the backbone are often difficult to determine by X-ray crystallography. Obviously, if the electron density corresponds to a mixture of sub-states, then the resulting "average" experimental geometry may be meaningless. However, we usually observe surprisingly good agreement between experiments and RNA simulations due to compensating errors, base pairing and stacking constraints, and the accuracy of the starting structures. For example, the dihedral backbone angles around an S-turn motif in a 2.05 Å resolution structure of the hairpin ribozyme with a single-nucleotide U39C mutation[95] differ from those of lower resolution (2.65 Å) structures carrying wild-type U39.[96] This difference could either be due to the distinct crystallographic constructs used in the two studies or an artifact of the more limited resolution of the second structure. MD simulations resolved this ambiguity.[12] Starting from the lower resolution crystal structure, the backbone dihedrals switched to (and in fact predicted[9]) those observed in the higher resolution structure. Structural bioinformatics using a recently developed backbone nomenclature assisted in the rapid evaluation of the backbone behavior.[12,97]

The importance of the starting structure for MD simulations is well illustrated by simulation studies of HIV-1 dimerization initiation site (DIS) kissing-loop complexes (Figure 6.7). When initially simulating the RNA kissing complexes, we observed substantial rearrangement of the bulged-out bases upon starting from the crystal structure available at the time, where the bulged-out bases were involved in packing interactions.[16] The four-purine stack predicted by MD simulations was subsequently confirmed by new X-ray crystallography studies as the preferred topology.[98] However, the bulged-out position of the unpaired nucleotides in the HIV-I DIS complexes consistently seen in X-ray and simulation studies does not agree with three independent solution experiments that predict bulged-in geometries of these bases.[52] Subsequent prolonged simulations still predicted the bulged-out geometry as the

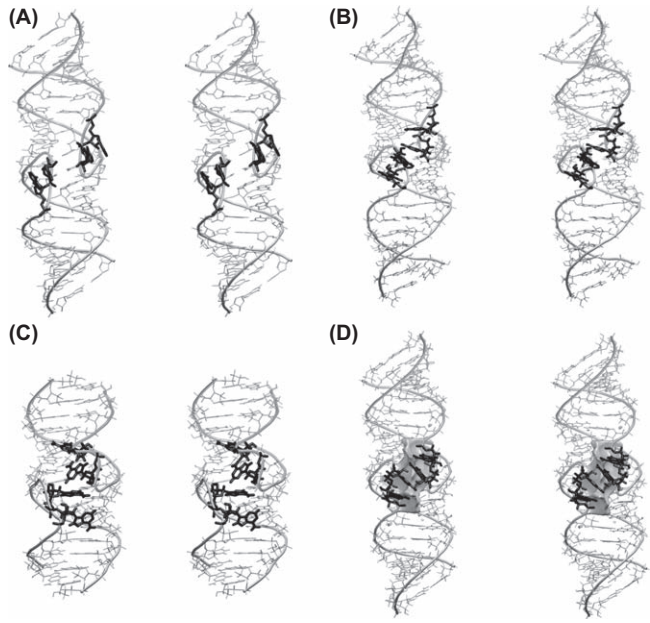

Figure 6.7 (A) Stereo view of the X-ray structure of the HIV DIS kissing loop complex (PDB ID 2B8R) with four bulged adenines (in black). These bases are in bulged-out geometry and are involved in crystal packing (not shown), so that there are two adenine-stacks separated by a gap. (B) Stereo view of the MD structure of the HIV DIS kissing loop complex. The four bulged adenines (in black) form a continuous four-adenine stack.[16,52] This conformation was confirmed by a new crystal structure (PDB ID 1ZCI). (C) Stereo view of the NMR structure of the HIV DIS kissing loop complex (PDB ID 2D19) where bulged bases (in black) are in a bulged-in geometry. (D) Stereo view of a new crystal structure of the HIV DIS kissing loop complex (PDB ID 1ZCI) showing a continuous four-adenine stack (in black) with the electrostatic potential map found inside the complex. This wide and partially open electrostatic pocket is a binding site for cations. While some X-ray crystallographic experiments show Mg^{2+} ion binding, other structures show at first sight an empty pocket. MD simulations show that the pocket is always occupied by ions, with a competition between monovalent and divalent ions. The simulations show a smooth exchange of the ions between the bulk and the pocket on a scale of few ns. Since the monovalent ions alternate between many distinct positions (they are delocalized in the pocket) they are difficult to detect experimentally.

most likely one, however, the simulations also revealed a tendency to sample bulge-in geometries, which thus appear to be competitive with the bulged out geometries.[52] It should be noted that if two distinct structures have similar free energies, it is virtually impossible to determine their relative population by MD. MD can predict that they are similar in energy but MD does not have the accuracy to safely predict the relative populations, since an error of $\sim$1.5 kcal/mol (*cf.* with the approximate nature of the force field) would change

the relative population of two competing sub-states by an order of magnitude. Regarding the RNA kissing complexes, it should be noted that the available NMR structures are in mutual disagreement concerning the specific base positions, while the simulations clearly indicate substantial local inaccuracies in the NMR structures.[52] Thus it is difficult to assess significance and quality of the NMR structures (*i.e.*, which of them, if any, is the correct one), as the simulations starting from these structures did not converge in simulation behavior (*i.e.*, the simulations were not able to overcome the impact of the starting structures).[52] It is obviously possible that the crystal packing shifts the equilibrium in favor of the bulged-out bases in the X-ray structures, but it is also not clear if the NMR experiment would even detect the bulged-out structures if present. This example shows that sometimes it is not straightforward to fully resolve the question of which starting structures is most accurate. As well, it shows that many RNA NMR structures may be over-interpreted and not sufficiently accurate to initiate explicit solvent MD simulations.[52]

6.5 What Specific Questions Can MD Simulations of RNA Currently Address?

6.5.1 Resolving Experimental Artifacts

Although simulations cannot predict the overall folding of an RNA, they can locally resolve regions of limited resolution in known experimental structures and reveal structural defects due to crystal packing. For example, a local region of lower resolution was observed near the conformationally dynamic active site in precursor crystal structures of the HDV ribozyme. An unusual set of backbone dihedral angles at the active site became more canonical during MD simulations, ultimately adopting a common U-turn motif.[11] Crystal structures of the HDV ribozyme also showed an extruded guanine (G76) that participates in crystal packing.[99] Multiple simulations revealed the rapid loss of this particular conformation and predicted a possible role of G76 in promoting catalysis through novel hydrogen bonding interactions with stem P1.[10] Similarly, inactivating 2′-deoxy or 2′-O-methyl backbone modifications or base mutations used to trap ribozymes in their precatalytic structures may distort the active site. Multiple MD simulations of the hairpin ribozyme consistently resulted in a change in the A-1 sugar pucker in the absence of the 2′-O-methyl modification present in the experimental structure, leading to significant repositioning of the catalytically important nucleotides G8 and A38.[12]

It should, however, be noted that the power of MD is not unlimited. In our studies of the HDV ribozyme, computations aimed to predict the catalytic mechanism based on structural sampling were initiated from an inactivated precursor structure available at the time, with the catalytic C75 replaced by U.[100] This crystal structure (upon substituting C back to the static structure) is in fact quite consistent with a general base role of C75 during catalysis, which predicts that C75 deprotonates the U-1 2′-hydroxyl group. The computations

readily suggested structural details consistent with such a catalytic mechanism. However, a number of relevant mechanistic studies are rather consistent with C75 acting as general acid, which would mean that C75 is protonated before the reaction.[101] We initiated extensive simulations with a protonated C75 but could not derive structures compatible with this mechanism. A new X-ray structure of HDV ribozyme now appears to be consistent with the C75 indeed acting as the general acid, although the key region is not resolved in the electron density map.[102] We conclude that MD simulations were not able to overcome a substantial energy barrier associated with sufficiently re-configuring the catalytic core of the earlier crystal structure (where the geometry of the catalytic center is presumably affected by the C75U substitution) to reach a conformation compatible with the general acid mechanism. Of course, it is also possible that the general base mechanism is an alternative catalytic pathway in HDV ribozyme chemistry. Regardless, computational studies provide an important overall picture of the dynamic properties of the HDV ribozyme.

6.5.2 Flexibility of RNA Building Blocks

Stochastic flexibility is a key functional feature of RNA that is difficult to derive from experiment. MD fills this gap by achieving a qualitative, atomistic understanding of the stochastic dynamics and flexibility of RNA building blocks.[13,14] For example, simulations have revealed striking intrinsic dynamics of RNA kink-turns, which can act as anisotropic molecular "elbows" to facilitate functional dynamics of the ribosome (Figure 6.8).[13] The idea of kink-turns contributing to functional dynamics of the ribosome was investigated by MD for two flexible regions of the large ribosomal subunit with strategically located kink-turns. Comparison of MD with cryo-EM data suggested that kink-turn 38 at the base of Helix 38 (A-site finger) may indeed allow fast penalty-free relocations of the tip of Helix 38 during translocation.[74] Interestingly, the archeal kink-turn 38 is not conserved in other kingdoms. The simulations, however, revealed that the at first sight unrelated (considering their distinct sequences and 2D structures) corresponding RNA regions in bacteria can in fact act as structural and dynamic analogs of the kink-turn.[74] We also suggested that the universally conserved kink-turn in the center of Helix 42 may facilitate the functional dynamics of the RNA in the GTP-ase associated center (GAC), which is found in one of the most conserved regions of the large subunit that interacts with the elongation factors and incoming tRNAs. However, subsequent simulation analysis, including the L10 protein directly and considering the Helix 42–Helix 97 tertiary interaction indirectly indicated that the static sub-states of the GAC visualized by available X-ray crystallography can be best explained by another anisotropic hinge-like RNA building block, that of the three-way junction between Helix 42 and the GAC.[73]

Note that, when the RNA molecule is substantially flexible with anisotropic and large fluctuations originating from local interactions, this flexibility is not captured by coarse-grained normal mode analysis.[14,73] In addition, we have

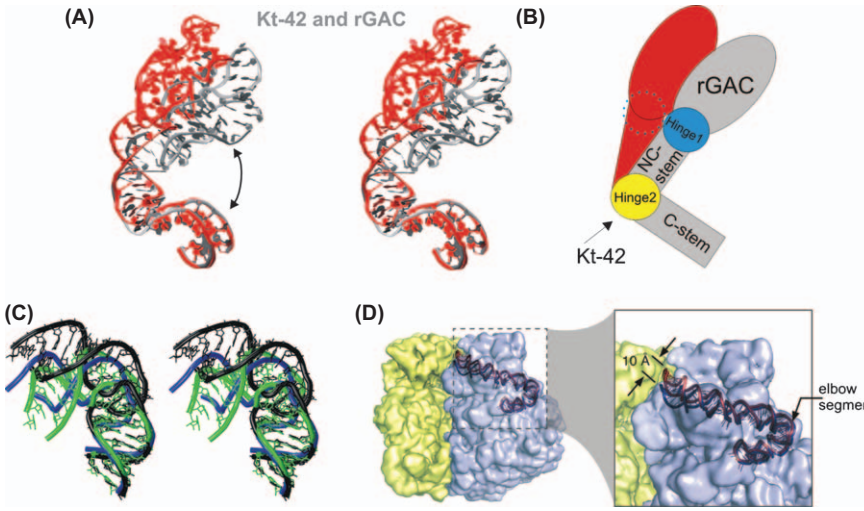

Figure 6.8 Flexible RNA segments. (A) Spontaneous fluctuations of the 23S rRNA kink-turn (Kt) 42 of Helix 42 and the GTPase associated center RNA (GAC, Helices 43 and 44) are captured by MD on a timescale of 50 ns (stereo view). The main dynamics occur at Kt-42 and the three-way junction between Helix 42 and the GAC RNA. (B) The Helix 42–44 segment of the large ribosomal subunit has evolved as a sophisticated anisotropic non-harmonic RNA double-elbow nanoarm (scheme). The C and NC stems flanking the flexible Kt-42 (marked as hinge 2) are relatively stiff. Kt-42 and the three-way junction (hinge 1) are flexible anisotropic elbow-like RNA building blocks. It was initially suggested that the flexibility of Kt-42 can explain the positional variability of this RNA segment in the available ribosomal X-ray structures. More recent analysis, however, suggests that the experimentally observed positional variability originates from hinge 1. Kt-42 is restricted in the ribosome by a tertiary contact between the NC-stem and Helix 97 (not shown).[73] (C) The *Escherichia coli* 23S rRNA elbow segment of Helix 38, with the range of spontaneous fluctuations captured by MD. The fluctuations sampled spontaneously in unrestrained simulations define the low-energy region of conformations and the intrinsic flexibility of the simulated molecule. (D) Cryo-electron microscopy map of the ribosome (at ~10 Å resolution) with the A-site finger highlighted (*i.e.*, Helix 38, in ribbon representation, fitted by flexible-fitting refinement). The A-site finger leads from the backside of the large subunit (in blue) towards the small subunit (in yellow), making a dynamic inter-subunit bridge. The A-site finger is, among carrying other functions, an attenuator of translocation. It also likely contributes to communication of structural changes in the ribosome during translocation. The cryo-EM experiment shows a movement on the order of ~10 Å of the tip of the A-site finger over the course of the translation cycle, during ribosome ratcheting. Although the resolution of the experiment does permit an unambiguous analysis, the movement may originate from the flexible elbow segment and may propagate through the entire ~100-Å-long Helix 38 structure towards the tip.[74]

shown that the popular Essential Dynamics Analysis (EDA or principle component analysis, PCA) may be similarly misleading. Blind application of this tool is prone to over-interpretation, due do the considerable approximations inherent to PCA.[73] Although tools such as EDA may help to understand or visualize the dynamics, none of these methods can substitute for directly identifying the movements in full simulation trajectories.

6.5.3 Revealing Solvent and Ion Dynamics

MD simulations can detect long-residency water molecules that occupy prominent hydration sites and remain bound for many nanoseconds, contrasting with common water binding events of only ~50–500 ps duration. Long-residency water molecules can serve structural, functional and possibly catalytic roles.[9,13,16,103] A structurally relevant long-residency hydration site was detected in the A-minor I tertiary interactions of kink-turns 38 and 42 in 23S ribosomal RNA. Their *cis* Sugar-Edge/Sugar-Edge A:C base pairs dynamically oscillate between direct and water-mediated hydrogen bonds whose interconversion significantly contributes to the elbow-like flexibility of the kink-turns.[13] The static crystal structures show both geometries, where the A:C interaction of kink-turn 38 is water-mediated and that of kink-turn 42 is direct.[13] The presence of inter-domain long residency water molecules was also predicted in the simulations based on lower-resolution crystal structures of the hairpin ribozyme. These water molecules were ultimately verified by the emergence of higher resolution structures.[9,104]

MD simulations can qualitatively characterize major binding sites of monovalent ions that are primarily determined by electrostatic interactions. Simulations in monovalent ions alone have revealed ion densities in known multivalent ion binding sites.[9,15–17,30,52,105] For example, simulations of the HDV ribozyme predict monovalent cation binding at the cleavage site in a crystallographically resolved divalent metal ion binding site proximal to the 5′-O leaving group. Two Na^+ ions and their accompanying first hydration spheres fill the catalytic pocket and may contribute to catalysis in the absence of divalent ions.[17] This MD prediction was later verified by crystal structures solved in the presence of Tl^+, which reveal two Tl^+ ions at the active site.[99] Simulations further predict a competition between ion binding and protonation of C75,[17] a feature not evident from the crystal structures but supported by mechanistic studies.[106] An additional binding site was predicted near the 2′-OH nucleophile. This site is again verified by crystal structures, however, the exact coordination geometry differs between experiment and simulation, likely reflecting a combination of differences between the ions used, force field approximations, and crystallographic ambiguities.[17] Ion binding sites may also elude experimental detection, either due to low resolution or ion delocalization in the pocket, as observed for 5SrRNA loop E and the HIV-DIS kissing complex.[16,52,105] To better understand the substantial, but not always acknowledged, limitations in experimental determination of the ions bound to RNA, we suggest the following article for further reading.[107]

6.5.4 Probing the Structural Effects of Base Substitutions and Ionizations

MD can assess the effects of base substitutions at atomic resolution to complement experimental mutagenesis studies. Thus, several base substitutions were modeled into the experimentally determined crystal structures of the hairpin ribozyme,[9] with good agreement between the MD predicted and experimentally determined stability of the tertiary structure.[9] The same simulations also revealed the importance of coupled networks of hydrogen bonds involving long-residency water molecules for tertiary structure stability, whereby mutations exert significant long-range effects.[9] In simulations of the HDV ribozyme, each of the four standard nucleobases was separately modeled into the −1 position immediately 5′ of the cleavage site. The wild-type U-1 was found to have the most tightly folded catalytic core, consistent with experimental footprinting data.[11] The same simulations revealed that a hydrogen bond characteristic of U-turn motifs from U-1 to the phosphate of C3 is only transiently sampled, reflecting a local flexibility at the cleavage site that correlates with increased catalytic activity. Simulations of ribozymes thus reveal additional structural and functional features that expand on experimental structures.

MD simulations of the HDV and hairpin ribozymes and *glmS* riboswitch were used to assess the impact of protonation states on catalytically relevant structures. In simulations of the HDV ribozyme in which C75 was in its neutral (unprotonated) form, a geometry was adopted that is suitable for general base catalysis,[15,18] while simulations with a protonated $C75H^+$ did not predict a reasonable geometry for C75 to act as a general acid.[15] (However, see the above section concerning limitations caused by the starting structure.) For hairpin ribozyme catalysis, compelling mechanistic evidence likewise suggests a direct catalytic role for A38. Similarly to the HDV ribozyme, MD simulations of the hairpin ribozyme with an unprotonated A38 lead to a geometry compatible with A38 acting as the general base.[12] Such a role is consistent with the available biochemical evidence but was previously discounted based on heteroatom distances likely influenced by the backbone 2′-O-methyl modification of the cleavage site in crystal structures.[96,104] Unlike the HDV ribozyme, hairpin ribozyme simulations with a protonated $A38H^+$ provide a geometry suitable for general acid catalysis as well.[12] On the other hand, MD simulations with a deprotonated $G8^-$, which was suggested as possible general base, led to active site distortion whereas the G8 tautomer was well tolerated.[77] Similar results were found in the case of the *glmS* catalytic riboswitch, where deprotonation of the catalytically important $G40^-$ was found to distort the active site, while simulations with a neutral canonical G40 closely resembled the crystal architecture.[108] Although these MD studies ultimately identified dominant protonation states of crucial bases under crystalline conditions, it is possible that some rarely populated, but highly reactive protonation states of these bases may be involved in catalysis. The description of these minor populated sub-states usually goes, however, beyond the applicability of MD techniques

and needs support from QM methods. Obviously, assessing a catalytic mechanism is limited in classical MD since bond breaking and making are by definition unobservable, thus warranting the use of QM to further evaluate the feasibility of a specific mechanism suggested by classical MD and biochemical data.[25]

6.6 What General Scope and Limitations Do Quantum Mechanical Calculations Have?

In contrast to the use of force fields from classical mechanics, QM can achieve a physically more rigorous description of chemical systems. *Ab initio* QM methods are free of empirical parameters and offer a systematic (and controllable) tuning of their quality by improving the underlying basis sets of atomic orbitals together with a balanced inclusion of electronic correlation effects. Accurate QM calculations are, however, currently limited to 30–50+ atoms and are carried out in the gas phase.[29] QM allows reliable evaluation of intrinsic (gas phase) interaction energies, defined as differences between the electronic energies of a dimer and its component non-interacting monomers. This direct structure-energy relationship can be accurately calculated for any single geometry of a stacking or base pairing interaction to map the complete potential energy surface.[60–62] Such energies unambiguously reflect direct forces between the interacting partners with no influence of the environment, making QM a genuine reference tool to parameterize and verify other computational approaches, including force fields.[37,61,62] When electron correlation calculations are expanded to complete basis sets of atomic orbitals (abbreviated as CBS) and include corrections for higher-level electron correlation effects (usually *via* the CCSD(T) method), QM calculations reach similar accuracies for both base pairing and stacking, and effectively converge.[60,61] Similar computations are possible also for the conformational analysis of flexible chains such as segments of an NA backbone, although such calculations are in fact trickier than computations of molecular clusters.[66] We refer the reader to the specialized literature cited above. An important advantage of such highest-quality calculations is that they allow for an accurate description of a broad range of systems and chemical reactions so that they serve as a genuine reference method. For some systems, the inclusion of higher-order electron correlation effects is unimportant. Then the second-order MP2 method is sufficient. For other systems, such as base stacking, computationally expensive higher-order electron correlation calculations are crucial.

Currently, standard QM calculations can often be replaced by Density Functional Theory (DFT) approaches that are much faster. There has been a recent explosion in new DFT methods, obscuring the literature for the non-specialist. However, a description of this literature goes beyond the scope of this review. An important point that needs to be considered is the following. DFT methods require certain parameterizations. Although they allow to reach very high accuracy for various chemical problems, none of them appear to be

suitable for all problems simultaneously in a manner comparable to conventional (wave function theory) QM methods. Thus, DFT methods need to always be carefully selected for a specific application, and they have to be tested against standard reference QM computations.

Despite their superb accuracy, it is not straightforward to extend modern QM calculations to biomolecules. NA conformations in particular result from a highly variable mixture of mutually compensating interactions, the balance of which may vary for distinct architectures. In addition, the strong electrostatic forces in NA are substantially modulated by solvent screening effects. Accurate inclusion of solvent effects is beyond the capability of modern QM approaches. Special care is needed when including the NA backbone in QM studies. Isolated small model systems (even as small as a single nucleotide) favor geometries that are biased by gas-phase specific, intramolecular hydrogen bonds, where electrostatic effects of the phosphates dominate the energetics.[109] The problem is not so much the quality of the QM methods, but the incompleteness of the model system. A comprehensive review of the link between QM studies and bioinformatics of RNA has recently been published.[31]

6.7 What Can QM/MM Reveal About the Chemical Change Catalyzed by Ribozymes?

A wide spectrum of both fast and accurate QM approaches have recently emerged, allowing for the inclusion of hundreds of atoms in a QM calculation.[110] Unfortunately, making a QM system larger but still incomplete will only exacerbate the errors resulting from the incompleteness of the system.[109] A system consisting of a few hundred atoms remains still far from a completely solvated biopolymer. However, fast QM methods facilitate applications of QM/MM hybrid methods where a smaller segment of the system is treated quantum chemically while the remainder, including the solvent, is treated classically using force fields.[111] QM is particularly attractive for ribozymes since QM, but not MD, can describe the reactions catalyzed.[25] The main limitations of current QM/MM methods derive from insufficient sampling (including choice of the starting structure), inaccuracies of the QM or MM method, and artifacts due to the obviously unphysical boundary between the QM and MM regions.[25] To enhance QM/MM sampling, semi-empirical (such as AM1, SCC-DFTB) and empirical (EVB) methods are used.[25,111]

Calculations using QM/MM methods have predicted specific roles for nucleobases, divalent ions, and/or electrostatic stabilization in catalyzing self-cleavage by the hairpin and HDV ribozymes.[18–20] QM/MM methods have also been successfully applied to the elucidation of the mechanism of peptide bond formation and translation termination on the ribosome.[22,112] MD simulations of the HDV ribozyme provided a suitable starting geometry for a mechanism in which an unprotonated neutral C75 acts as the general base. QM/MM calculations are consistent with a role of C75 as the general base and Mg^{2+} as the general acid, predicting an energy barrier of $\sim$20 kcal/mol for the catalyzed

reaction.[18] For the QM scans, a region composed of 80 atoms in the active site was treated quantum-mechanically. Multiple starting positions of a specifically bound Mg^{2+} were sampled, establishing a hexacoordinated Mg^{2+} ion with a single innersphere contact to a cleavage site non-bridging oxygen as the most likely configuration, with the Mg^{2+} acting as a Lewis acid in the reaction. Mechanisms in which C75 acts as the general acid instead, suggested by the relevant number of biochemical studies,[101,113] could not be explored due to the paucity of suitable starting geometries (see above). In contrast, MD simulations of the hairpin ribozyme with protonated and unprotonated A38 result in plausible catalytic geometries for A38 acting as general acid and/or general base, respectively.[12] These simulations reveal in part the complex impact of base ionization on the starting ground-state geometry, and may explain the apparent insensitivity to base ionization of an initial QM/MM analysis of the hairpin ribozyme based on a crystal structure of a transition-state analog.[19,20] Large-scale classical simulations may be essential for establishing starting geometries suitable for subsequent QM/MM calculations.

6.8 Conclusion

MD simulations of RNA are a useful tool to expand on experimental structures and biochemical data, providing unique atomistic descriptions of the dynamic roles of nucleobases, the backbone, counter ions, and individual water molecules in imparting biological function to RNA. Experiments benefit from a side-by-side comparison with simulations, where MD can help in refinement, interpretation and better understanding of existing experimental structures. When assessing MD simulations we need to consider that ensemble averaging and error margins of the underlying experimental structures have an impact, and that force field artifacts are pervasive. In some instances, the available force field may not be sufficient to obtain meaningful results, in which case the limitations should be fully acknowledged[75,76] or even be addressed by improving the force field.[37,50] QM calculations, often in the form of hybrid QM/MM approaches, can further build on MD simulations to access reaction chemistry.

Acknowledgments

This work was supported by grants IAA400040802 (J.Š., M.O.) and KJB400040901 (K.R.) from the Grant Agency of the Academy of Sciences of the Czech Republic, Operational Program Research and Development for Innovations – European Social Fund CZ.1.05/2.1.00/03.0058 (M.O, P.B.), Grants 203/09/1476, P208/11/1822 (J.Š.) and P301/11/P558 (P.B.) from the Grant Agency of the Czech Republic, grants LC06030 (J.Š.), AV0Z50040507 (J.Š.) and AV0Z50040702 (J.Š.) from the Ministry of Education of the Czech Republic, and grant GM62357 from the NIH to N.G.W.

References

1. N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 905–920.
2. P. F. Egea, R. M. Stroud and P. Walter, *Curr. Opin. Struct. Biol.*, 2005, **15**, 213–220.
3. J. Liu, *Curr. Opin. Cell Biol.*, 2008, **20**, 214–221.
4. A. Torres-Larios, K. K. Swinger, T. Pan and A. Mondragon, *Curr. Opin. Struct. Biol.*, 2006, **16**, 327–335.
5. S. He, Z. Yang, G. Skogerbo, F. Ren, H. Cui, H. Zhao, R. Chen and Y. Zhao, *Crit. Rev. Microbiol.*, 2008, **34**, 175–188.
6. H. M. Al-Hashimi and N. G. Walter, *Curr. Opin. Struct. Biol.*, 2008, **18**, 321–329.
7. M. A. Ditzler, M. Otyepka, J. Sponer and N. G. Walter, *Acc. Chem. Res.*, 2010, **43**, 40–47.
8. T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung and X. Xiao, *Quart. Rev. Biophys.*, 2011, **44**, 191–228.
9. M. M. Rhodes, K. Reblova, J. Sponer and N. G. Walter, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 13380–13385.
10. J. Sefcikova, M. V. Krasovska, N. Spackova, J. Sponer and N. G. Walter, *Biopolymers*, 2007, **85**, 392–406.
11. J. Sefcikova, M. V. Krasovska, J. Sponer and N. G. Walter, *Nucleic Acids Res.*, 2007, **35**, 1933–1946.
12. M. A. Ditzler, J. Sponer and N. G. Walter, *RNA*, 2009, **15**, 560–575.
13. F. Razga, J. Koca, J. Sponer and N. B. Leontis, *Biophys. J.*, 2005, **88**, 3466–3485.
14. F. Razga, J. Koca, A. Mokdad and J. Sponer, *Nucleic Acids Res.*, 2007, **35**, 4007–4017.
15. M. V. Krasovska, J. Sefcikova, N. Spackova, J. Sponer and N. G. Walter, *J. Mol. Biol.*, 2005, **351**, 731–748.
16. K. Reblova, N. Spackova, J. E. Sponer, J. Koca and J. Sponer, *Nucleic Acids Res.*, 2003, **31**, 6942–6952.
17. M. V. Krasovska, J. Sefcikova, K. Reblova, B. Schneider, N. G. Walter and J. Sponer, *Biophys. J.*, 2006, **91**, 626–638.
18. P. Banas, L. Rulisek, V. Hanosova, D. Svozil, N. G. Walter, J. Sponer and M. Otyepka, *J. Phys. Chem. B*, 2008, **112**, 11177–11187.
19. K. Nam, J. Gao and D. M. York, *J. Am. Chem. Soc.*, 2008, **130**, 4680–4691.
20. K. Nam, J. Gao and D. M. York, *RNA*, 2008, **14**, 1501–1507.
21. S. Trobro and J. Aqvist, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 12395–12400.
22. P. K. Sharma, Y. Xiang, M. Kato and A. Warshel, *Biochemistry*, 2005, **44**, 11307–11314.
23. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

24. *Molecular Modeling and Simulations: An Interdisciplinary Guide*, 2nd edn., ed. T. Schlick, Springer, New York, USA, 2010.
25. P. Banas, P. Jurecka, N. G. Walter, J. Sponer and M. Otyepka, *Methods*, 2009, **49**, 202–216.
26. C. Laing and T. Schlick, *J. Phys. Condens. Matter*, 2010, **22**, 283101.
27. C. Laing and T. Schlick, *Curr. Opin. Struct. Biol.*, 2011, **21**, 1–13.
28. J. Sponer and N. Spackova, *Methods*, 2007, **43**, 278–290.
29. J. Sponer and F. Lankas (eds), *Challenges and Advances in Computational Chemistry and Physics: Computational Studies of RNA and DNA*, Springer, Dordrecht, Netherlands, 2006.
30. S. E. McDowell, N. Spackova, J. Sponer and N. G. Walter, *Biopolymers*, 2007, **85**, 169–184.
31. J. Sponer, J. E. Sponer, A. I. Petrov and N. B. Leontis, *J. Phys. Chem. B*, 2010, **114**, 15723–15741.
32. T. E. Cheatham, *Cur. Opin. Struct. Biol.*, 2004, **14**, 360–367.
33. J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3–31.
34. J. Trylska, *Quart. Rev. Biophys.*, 2009, **42**, 301–316.
35. N. Foloppe and A. D. MacKerell, *J. Comput. Chem.*, 2000, **21**, 86–104.
36. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1357–1377.
37. A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
38. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
39. M. Zgarbova, M. Otyepka, J. Sponer, P. Hobza and P. Jurecka, *Phys. Chem. Chem. Phys.*, 2010, **12**, 10476–10493.
40. P. Cieplak, F. Y. Dupradeau, Y. Duan and J. M. Wang, *J. Phys.-Condens. Mat.*, 2009, 21.
41. A. D. Mackerell, Jr., *J. Comput. Chem.*, 2004, **25**, 1584–1604.
42. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712.
43. R. B. Best, N. V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07–L09.
44. R. B. Best and G. Hummer, *J. Phys. Chem. B*, 2009, **113**, 9004–9015.
45. P. Florova, P. Sklenovsky, P. Banas and M. Otyepka, *J. Chem. Theory Comput.*, 2010, **6**, 3569–3579.
46. J. Vymetal and J. Vondrasek, *J. Phys. Chem. B*, 2010, **114**, 5632–5642.
47. O. Guvench and A. D. J. MacKerell, *Methods Mol. Biol.*, 2008, **443**, 63–88.
48. T. E. Cheatham, 3rd, P. Cieplak and P. A. Kollman, *J. Biomol. Struct. Dyn.*, 1999, **16**, 845–862.
49. J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
50. P. Banas, D. Hollas, M. Zgarbova, P. Jurecka, M. Orozco, T. E. Cheatham, 3rd, J. Sponer and M. Otyepka, *J. Chem. Theory Comput.*, 2010, **6**, 3836–3849.

51. M. Zgarbova, M. Otyepka, J. Sponer, A. Mladek, P. Banas, T. E. Cheatham and P. Jurecka, *J. Chem. Theory Comput.*, 2011, **7**, 2886–2902.
52. K. Reblova, E. Fadrna, J. Sarzynska, T. Kulinski, P. Kulhanek, E. Ennifar, J. Koca and J. Sponer, *Biophys. J.*, 2007, **93**, 3932–3949.
53. A. Perez, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2007, **129**, 14739–14745.
54. A. Perez, F. Lankas, F. J. Luque and M. Orozco, *Nucleic Acids Res.*, 2008, **36**, 2379–2394.
55. N. J. Deng and P. Cieplak, *Biophys. J.*, 2010, **98**, 627–636.
56. I. Faustino, A. Perez and M. Orozco, *Biophys. J.*, 2010, **99**, 1876–1885.
57. E. J. Denning, U. D. Priyakumar, L. Nilsson and A. D. Mackerell, *J. Comput. Chem.*, 2011, **32**, 1929–1943.
58. C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656–1676.
59. C. G. Ricci, A. S. C. de Andrade, M. Mottin and P. A. Netz, *J. Phys. Chem. B*, 2010, **114**, 9882–9893.
60. J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco and P. Hobza, *Chem. Eur. J.*, 2006, **12**, 2854–2865.
61. J. Sponer, P. Jurecka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142–10151.
62. J. E. Sponer, K. Reblova, A. Mokdad, V. Sychrovsky, J. Leszczynski and J. Sponer, *J. Phys. Chem. B*, 2007, **111**, 9153–9164.
63. C. L. Zirbel, J. E. Sponer, J. Sponer, J. Stombaugh and N. B. Leontis, *Nucleic Acids Res.*, 2009, **37**, 4898–4918.
64. C. A. Morgado, P. Jurecka, D. Svozil, P. Hobza and J. Sponer, *J. Chem. Theory Comput.*, 2009, **5**, 1524–1544.
65. Y. X. Sun, D. Spellmeyer, D. A. Pearlman and P. Kollman, *J. Am. Chem. Soc.*, 1992, **114**, 6798–6801.
66. A. Mladek, J. E. Sponer, P. Jurecka, P. Banas, M. Otyepka, D. Svozil and J. Sponer, *J. Chem. Theory Comput.*, 2010, **6**, 3817–3835.
67. J. Sponer, M. Sabat, L. Gorb, J. Leszczynski, B. Lippert and P. Hobza, *J. Phys. Chem. B*, 2000, **104**, 7535–7544.
68. N. Gresh, J. E. Sponer, N. Spackova, J. Leszczynski and J. Sponer, *J. Phys. Chem. B*, 2003, **107**, 8669–8681.
69. J. Aqvist, *J. Phys. Chem.*, 1994, **98**, 8253–8255.
70. A. A. Chen and R. V. Pappu, *J. Phys. Chem. B*, 2007, **111**, 11884–11887.
71. I. S. Joung and T. E. Cheatham, 3rd, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.
72. I. Besseova, M. Otyepka, K. Reblova and J. Sponer, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10701–10711.
73. I. Besseova, K. Reblova, N. B. Leontis and J. Sponer, *Nucleic Acids Res.*, 2010, **38**, 6247–6264.
74. K. Reblova, F. Razga, W. Li, H. X. Gao, J. Frank and J. Sponer, *Nucleic Acids Res.*, 2010, **38**, 1325–1340.

75. E. Fadrna, N. Spackova, J. Sarzynska, J. Koca, M. Orozco, T. E. Cheatham, 3rd, T. Kulinski and J. Sponer, *J. Chem. Theory Comput.*, 2009, **5**, 2514–2530.
76. E. Fadrna, N. Spackova, R. Stefl, J. Koca, T. E. Cheatham, 3rd and J. Sponer, *Biophys. J.*, 2004, **87**, 227–242.
77. V. Mlynsky, P. Banas, D. Hollas, K. Reblova, N. G. Walter, J. Sponer and M. Otyepka, *J. Phys. Chem. B*, 2010, **114**, 6642–6652.
78. T. A. Halgren and W. Damm, *Curr. Opin. Struct. Biol.*, 2001, **11**, 236–242.
79. G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou and T. A. Halgren, *J. Comput. Chem.*, 2002, **23**, 1515–1531.
80. P. Y. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.
81. V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux and A. D. MacKerell, *J. Chem. Theory Comput.*, 2005, **1**, 153–168.
82. A. Warshel, M. Kato and A. V. Pisliakov, *J. Chem. Theory Comput.*, 2007, **3**, 2034–2045.
83. N. Tjandra, S. Tate, A. Ono, M. Kainosho and A. Bax, *J. Am. Chem. Soc.*, 2000, **122**, 6190–6200.
84. P. Varnai and K. Zakrzewska, *Nucleic Acids Res.*, 2004, **32**, 4269–4280.
85. D. L. Beveridge, G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, 3rd, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai and M. A. Young, *Biophys. J.*, 2004, **87**, 3799–3813.
86. K. Reblova, F. Lankas, F. Razga, M. V. Krasovska, J. Koca and J. Sponer, *Biopolymers*, 2006, **82**, 504–520.
87. P. Sklenovsky, P. Florova, P. Banas, K. Reblova, F. Lankas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2011, **7**, 2963–2980.
88. D. A. C. Beck, G. W. N. White and V. Daggett, *J. Struct. Biol.*, 2007, **157**, 514–523.
89. A. Noy, A. Perez, C. A. Laughton and M. Orozco, *Nucleic Acids Res.*, 2007, **35**, 3330–3338.
90. A. E. Garcia and D. Paschek, *J. Am. Chem. Soc.*, 2008, **130**, 815–817.
91. S. Nozinovic, B. Furtig, H. R. A. Jonker, C. Richter and H. Schwalbe, *Nucleic Acids Res.*, 2010, **38**, 683–694.
92. M. Sarver, C. L. Zirbel, J. Stombaugh, A. Mokdad and N. B. Leontis, *J. Math. Biol.*, 2008, **56**, 215–252.
93. P. Auffinger, *in Computational studies of RNA and DNA*, Springer, Dordrecht, Netherlands, 2006, p 283.
94. G. R. Bowman, X. Huang, Y. Yao, J. Sun, G. Carlsson, L. J. Guibas and V. S. Pande, *J. Am. Chem. Soc.*, 2008, **130**, 9676–9678.
95. S. Alam, V. Grum-Tokars, J. Krucinska, M. L. Kundracik and J. E. Wedekind, *Biochemistry*, 2005, **44**, 14396–14408.
96. P. B. Rupert and A. R. Ferre-D'Amare, *Nature*, 2001, **410**, 780–786.
97. J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits,

L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook and H. M. Berman, *RNA*, 2008, **14**, 465–481.
98. E. Ennifar and P. Dumas, *J. Mol. Biol.*, 2006, **356**, 771–782.
99. A. Ke, F. Ding, J. D. Batchelor and J. A. Doudna, *Structure*, 2007, **15**, 281–287.
100. A. L. Ke, K. H. Zhou, F. Ding, J. H. D. Cate and J. A. Doudna, *Nature*, 2004, **429**, 201–205.
101. S. R. Das and J. A. Piccirilli, *Nat. Chem. Biol.*, 2005, **1**, 45–52.
102. J. H. Chen, R. Yajima, D. M. Chadalavada, E. Chase, P. C. Bevilacqua and B. L. Golden, *Biochemistry*, 2010, **49**, 6508–6518.
103. N. G. Walter, *Mol. Cell*, 2007, **28**, 923–929.
104. J. Salter, J. Krucinska, S. Alam, V. Grum-Tokars and J. E. Wedekind, *Biochemistry*, 2006, **45**, 686–700.
105. P. Auffinger, L. Bielecki and E. Westhof, *J. Mol. Biol.*, 2004, **335**, 555–571.
106. S. Nakano, D. M. Chadalavada and P. C. Bevilacqua, *Science*, 2000, **287**, 1493–1497.
107. E. Ennifar, P. Walter and P. Dumas, *Nucleic Acids Res.*, 2003, **31**, 2671–2682.
108. P. Banas, N. G. Walter, J. Sponer and M. Otyepka, *J. Phys. Chem. B*, 2010, **114**, 8701–8712.
109. D. Svozil, J. E. Sponer, I. Marchan, A. Perez, T. E. Cheatham, 3rd, F. Forti, F. J. Luque, M. Orozco and J. Sponer, *J. Phys. Chem. B*, 2008, **112**, 8188–8197.
110. Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.
111. S. C. Kamerlin, M. Haranczyk and A. Warshel, *J. Phys. Chem. B*, 2009, **113**, 1253–1272.
112. S. Trobro and J. Aqvist, *Mol. Cell*, 2007, **27**, 758–766.
113. A. L. Cerrone-Szakal, N. A. Siegfried and P. C. Bevilacqua, *J. Am. Chem. Soc.*, 2008, **130**, 14504–14520.