Biochimica et Biophysica Acta 1850 (2015) 1072–1090

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Molecular dynamic simulations of protein/RNA complexes: CRISPR/Csy4 endoribonuclease ☆

Carolina Estarellas [a], Michal Otyepka [b], Jaroslav Koča [a], Pavel Banáš [b], Miroslav Krepl [c], Jiří Šponer [a,c,*]

[a] CEITEC — Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 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] Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 26 June 2014
Received in revised form 15 October 2014
Accepted 20 October 2014
Available online 24 October 2014

Keywords:
Cas6 superfamily
Endoribonuclease
RNA cleavage
Protein/RNA complex
Molecular dynamic simulation
Force field

ABSTRACT

Background: Many prokaryotic genomes comprise Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) offering defense against foreign nucleic acids. These immune systems are conditioned by the production of small CRISPR-derived RNAs matured from long RNA precursors. This often requires a Csy4 endoribonuclease cleaving the RNA 3′-end.
Methods: We report extended explicit solvent molecular dynamic (MD) simulations of Csy4/RNA complex in precursor and product states, based on X-ray structures of product and inactivated precursor (55 simulations; ~3.7 μs in total).
Results: The simulations identify double-protonated His29 and deprotonated terminal phosphate as the likely dominant protonation states consistent with the product structure. We revealed potential substates consistent with Ser148 and His29 acting as the general base and acid, respectively. The Ser148 could be straightforwardly deprotonated through solvent and could without further structural rearrangements deprotonate the nucleophile, contrasting similar studies investigating the general base role of nucleobases in ribozymes. We could not locate geometries consistent with His29 acting as general base. However, we caution that the X-ray structures do not always capture the catalytically active geometries and then the reactive structures may be unreachable by the simulation technique.
Conclusions: We identified potential catalytic arrangement of the Csy4/RNA complex but we also report limitations of the simulation technique. Even for the dominant protonation state we could not achieve full agreement between the simulations and the structural data.
General significance: Potential catalytic arrangement of the Csy4/RNA complex is found. Further, we provide unique insights into limitations of simulations of protein/RNA complexes, namely, the influence of the starting experimental structures and force field limitations. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

In many bacteria and archaea, fragments of foreign nucleic acids are integrated into *C*lustered *R*egularly *I*nterspaced *S*hort *P*alindromic *R*epeat (CRISPR) loci providing an adaptive immune system to protect the prokaryote cells against invading phages or plasmids. These foreign genetic sequences are placed between copies of a repeated host RNA sequence [1–4]. The CRISPR transcripts (long RNAs precursor, pre-crRNAs) are cleaved after the transcription within each repeat sequence to generate short mature CRISPR-derived RNAs (crRNAs). The crRNAs assemble with CRISPR-associated (Cas) proteins into large ribonucleoprotein (RNP) complexes (Cas/crRNA complexes) [5,6]. The targeting Cas/crRNA complex may base pair with the invading DNA and then triggers its degradation.

In type I and type III CRISPR systems [7], a specific Cas6 endoribonuclease splits the pre-crRNAs in a sequence-specific way to generate ~60-nucleotide (nt) crRNA products in which segments of the repeat sequence flank the spacer (the target "foreign" nucleic acid sequence) [5,8–14]. In contrast, crRNA biogenesis in type II systems requires RNase III [15]. The inactivation of the Cas proteins results in a total loss of the immune mechanism function [5,15–18]. Doudna and coworkers [10] studied the pre-crRNA processing by means of Csy4 enzyme (recently reclassified as Cas6f) [7] in CRISPR subtype I–F. They reported multiple X-ray structures of complex of Csy4 both with precursor pre-crRNA and product crRNA [10,19]. The precursor was inactivated by deoxy-mutation of the G20 nucleotide just

☆ This article is part of a Special Issue entitled Recent developments of molecular dynamics.

* Corresponding author at: Institute of Biophysics, Kralovopolska 135, 612 65 Brno, Czech Republic. Tel.: +420 541517 133.
E-mail address: sponer@ncbr.muni.cz (J. Šponer).

http://dx.doi.org/10.1016/j.bbagen.2014.10.021
0304-4165/

upstream of the scissile phosphate (PDB code of the best resolved precursor structure is 2XLK, Fig. S1). The electron density of C21 (downstream of the scissile phosphate) is disordered except of the scissile phosphate moiety. The Csy4 protein consists of an N-terminal ferredoxin-like domain and a C-terminal domain. This later domain forms most of the recognition interactions with the RNA (Fig. 1A). The RNA adopts a stem-loop structure [20] with five base pairs in A-form helical stem capped by GUAUA pentaloop containing a sheared G11-A15 base pair and a bulged nucleotide U14. The pentaloop structurally belongs to family of GNRA tetraloops with insertion, i.e., GNR(N)A family (see Fig. 1B) [21].

The Csy4 binds to both the pre-crRNA precursor and the crRNA product with high affinity (K_d ~50 pM, Fig. S2) [14]. Product and precursor co-crystal structures of Csy4 bound to crRNA and pre-crRNA, respectively, revealed three regions mediating the protein/RNA recognition (Fig. S3) [10,19]. The catalytic center region includes interactions between the 3′-end of RNA and the β6–β7 hairpin loop, approaching the RNA from the minor groove side. The RNA spans the β-hairpin loop with the stem terminal C6-G20 base pair. The G20 stacks with the aromatic side chain of Phe155 (see Fig. S3D). The second region includes base-specific interactions of Arg102 and Gln104 (linker segment connecting the body of Csy4 to the Arginine-rich helix) with the major groove faces of A19 and G20 (Fig. S3B) [14]. The third region includes interactions of the Arg-rich helix α3 with the RNA pentaloop and part of the adjacent major groove (Fig. S3C). All these interactions ensure specificity of Csy4 binding to the pre-crRNA repeat and its site-specific cleavage. Table S1 in Supporting information lists the protein/RNA interactions.

The endoribonuclease cleaves the pre-crRNAs at the 3′-end of RNA between G20 and C21 nucleotides, producing crRNAs with only one spacer. The spacer is confined between 8th- and 19th-nucleotides of the 5′- and 3′-ends of RNA, respectively (see Fig. 1B) [10]. The scissile phosphate forms hydrogen bond interactions with the Gln149 backbone nitrogen (β-hairpin loop) and the side chain of His29 (α1 helix) [10,19]. In addition, three residues near the active site were identified: Ser148, His29, and Tyr176 (Fig. S3A). Production of crRNA was strongly inhibited by S148C and H29A mutations, while Y176F mutation exhibited near wild-type activity [10,19]. These mutational assays suggested a crucial role of Ser148 and His29 in the catalytic mechanism. In addition, R102A and P155A mutations seriously reduced the crRNA processing, whereas Q104A mutation did not compromise the activity. The binding of the RNA substrate to Csy4 is unaffected by the mentioned mutations [10,19].

The cleavage reaction occurs in a metal ion-independent way and two reaction mechanisms were suggested. The first mechanism assumes activation of the 2′-hydroxyl nucleophile by Ser148 while His29 acts as a general acid protonating the leaving O5′ group (Fig. 2A) [10]. The second mechanism suggests that Ser148 rearranges the G20 ribose into the C2′-endo conformation and holds the 2′-hydroxyl nucleophile in place, positioning His29 to act as a general base deprotonating the 2′-hydroxyl nucleophile (Fig. 2B) [19]. Consensus about catalytic roles of Ser148 and His29 still has not been reached.

We use classical molecular dynamic (MD) simulations in explicit solvent to study the basic structural dynamics of the Csy4/RNA complex from *Pseudomonas aeruginosa* (Fig. 1A). MD simulations allow to obtain structural dynamic insights, investigate different protonation states of the catalytically important residues, and eventually prepare suitable structures for theoretical investigation of the catalytic mechanism via hybrid QM/MM methods [22–29]. Thus, MD simulations can complement the experimental data [30,31] and were successfully used to study RNA enzymes (ribozymes) catalyzing the sugar-phosphate backbone self-cleavage, i.e., the same reaction as catalyzed by the Csy4/RNA complex [32–37]. However, attention should always be paid to possible force field inaccuracies, which may affect the simulations [38]. Obviously, studies of catalytic RNP complexes are even more challenging than studies of ribozymes. In the present study we use a combination of *ff*99bsc0χ_{OL3} RNA force field [39–42] with *ff*99SB protein force field [43]. While reporting relevant information on the potential catalytic mechanism, we also discuss the uncertainties and limitations of the simulation methodology.

The X-ray structures undoubtedly bring irreplaceable insights into studies of RNA and RNP enzymes. Nevertheless, the X-ray experiment provides static and averaged structures, which in addition can be to a certain extent affected by the crystal packing. However, structural dynamics and flexibility may often be important for the biochemical processes, justifying simulation studies. Another uncertainty in experimental structures stems from the fact that the hydrogen atoms are not directly observable. This can create ambiguity in directionality of catalytically important hydroxyl groups and in determination of protonation states. Further, for example the X-ray structures of Csy4/RNA complexes were obtained at relatively low pH, which also could affect the protonation states. In case of enzyme precursors, the X-ray structures can be affected by the inactivating modifications and some key residues often remain unresolved. In addition, the experimental structures of RNA and RNP enzymes may capture inactive ground-state arrangements which require further rearrangements, including changes of the protonation states, when progressing toward the chemical reaction. It is important to point out (see also below) that the chemical reaction may utilize rarely populated (minor) conformations with high reactivity, while the dominant structures (captured by the experiments)

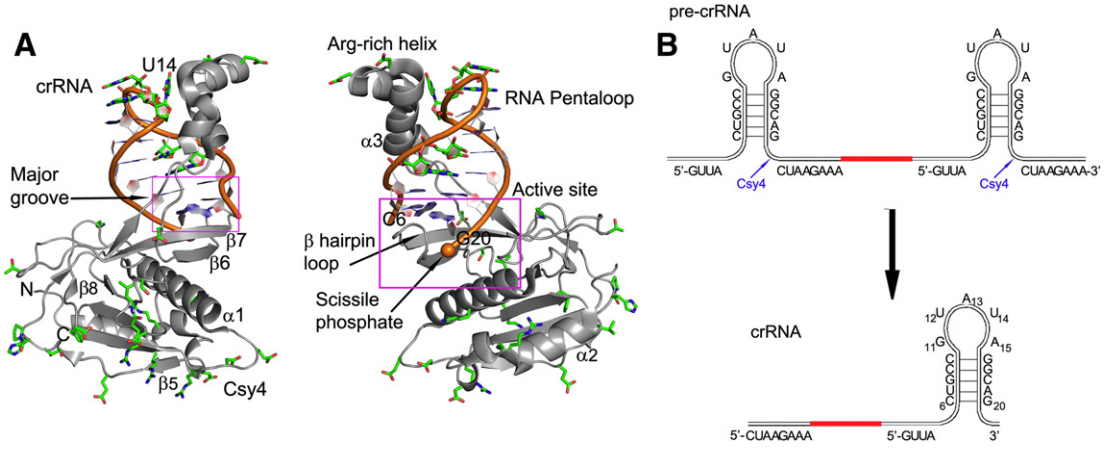


Fig. 1. (A) Back (left) and front (right) views of the crystal structure of Csy4/RNA product complex at 2.0 Å resolution (PDB ID: 4AL5) [19]. Csy4 is colored in gray and crRNA backbone is colored in orange. The major groove of crRNA and active site moieties are depicted in magenta boxes. The amino acid residues highlighted in green are involved in crystal contacts. (B) Schematic representation of pre-crRNA cleaved by Csy4 endoribonuclease to generate crRNA. The spacer sequence is shown in red and cleavage sites are indicated with blue arrows.

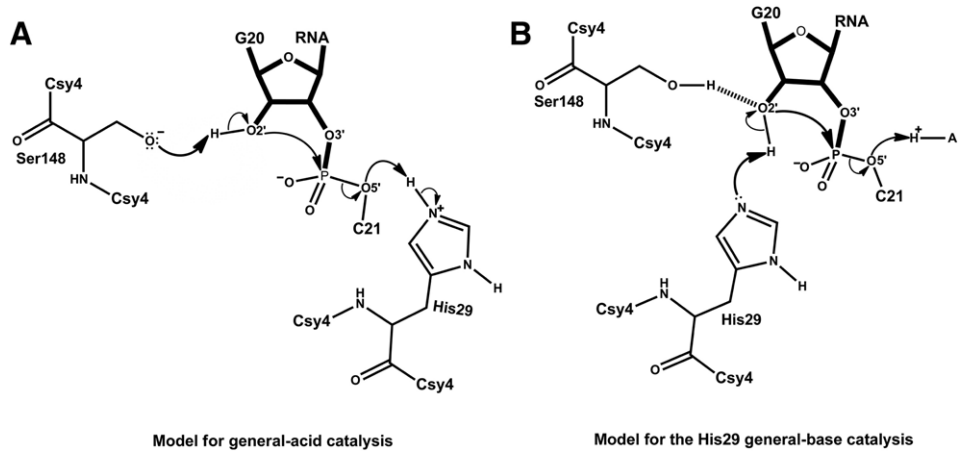


Fig. 2. Two catalytic strategies proposed for the phosphodiester cleavage by the Csy4 endoribonuclease. (A) Mechanism in which the double-protonated His29 (His^+) acts as the general acid protonating the leaving nt21(O5′) alcoholate [10]. The Ser148 can either structurally stabilize the in-line attack geometry or (as shown in the figure) it could act as the general base after being deprotonated. (B) Mechanism with δ-protonated His29 (His^0) acting as the general base activating the G20 (2′-OH) nucleophile. The leaving nt21(O5′) alcoholate is protonated by an unknown proton-donor group, marked as "A" [19]. The cleavage reaction occurs in a metal ion-independent way.

may be unreactive. As suggested recently for HDV ribozyme using combined experimental and MD simulation investigations, its available X-ray structures likely represent static snapshots on an overall rugged RNA folding free-energy landscape [36].

MD simulations can help to resolve all the above-noted issues. Simulations can probe different protonation states of the studied complexes, follow the H-bond orientations and remove the inactivating modifications. However, in order to properly understand the results, limitations of the MD technique should be taken into consideration, which is an issue ignored in many simulation studies. Besides the quality of the force fields, MD simulations of complex RNP systems with dense network of molecular interactions solved at medium resolution may be also affected by properties of the experimental starting structure [44,30,31]. The relation between the starting experimental structures and simulations is complex. In general, we would like to see that the MD technique is reproducing as many of the observed interactions as possible. However, genuine flexibility of some parts of the biomolecules such as protein loops and linkers may lead to deviations of the simulated structures from the single conformations captured by the experiments. Some additional differences may be caused by potential uncertainties stemming from the resolution limits (ranging from uncertain RNA backbone conformations up to entirely unresolved residues). The crystal packing effects can distort the biomolecule in comparison to its natural, fully solvated state, which can also lead to structural uncertainties that are transferred to the simulations. There could be occasionally improperly assigned side-chain rotamers of Asn, Gln, and His amino acids which contain atoms that are difficult to distinguish inside the electron density maps [45]. Similarly, in RNA, *anti*/*syn* bias may occasionally occur for the positions of the bases. All such problems lead to deviations between the experimental and simulated structures. In principle, converged simulations with correct force field should be ultimately capable to eliminate these uncertainties. However, the affordable lengths of the simulations are often insufficient to relax the structures [31]. Finally, even a structure which is in all aspects entirely correct from the experimental point of view at the limits of the resolution may initially introduce instabilities into MD simulations. The X-ray structures taken from the database are obviously not fully relaxed with respect to the simulation force field. In other words, the starting structure does not fully correspond to the potential energy minimum of the simulation force field, so that the unprocessed experimental coordinate file is always high in force field potential energy. Therefore, production simulations are always preceded by careful equilibration protocols. However, we suggest that in contrast to small systems it is not guaranteed that the standard equilibration protocols are robust enough for intricate RNP molecular complex. The elimination of the remaining excess energy may then randomly affect the production simulations even after careful equilibration [31]. All the above issues need to be considered when interpreting the simulations presented in our study, as well as in all other simulation studies on complicated RNP complexes.

In the present work we investigate product form of the Csy4/RNA complex and precursor-like structures with modeled methyl group in C5 position of nucleotide 21 and with different protonation states of the active site residues. Our study is primarily based on 35 simulations with total simulation time ~2.5 μs. Additional ~1.2 μs of simulation data is discussed in the Supporting information to understand the accuracy limits of our study.

2. Methods

2.1. Preparation of studied systems

The starting geometries of Csy4/RNA complexes were based on the X-ray structure of Csy4/RNA product complex with cleaved crRNA (PDB ID: 4AL5; resolution 2.0 Å) [19] and Csy4/RNA precursor complex (with pre-crRNA; inhibited by the deoxy-mutation of G20 upstream of the scissile phosphate; PDB ID: 2XLK; resolution 1.8 Å; two independent structures in the crystal lattice) [10].

The product X-ray structure shares similar overall folds, architectures of the catalytic pocket and hydrogen bond interaction patterns with the precursor (RMSd 0.4 Å over 811 atoms; Supporting information Fig. S2) [19]. Supporting information Table S1 lists the protein/RNA interactions. Note that the second structure of the 2XLK file has a different orientation of the Thr151 side chain compared to the first structure and the product structure. This locally affects H-bonding pattern at the catalytic center (Supporting information Table S1).

2.2. Protonation states of histidines

The protonation states of histidines were determined according to the structural context to maximize the hydrogen-bond network. His3 was double protonated, His161 and His82 δ-protonated and His61, His93, and His154 ε-protonated. All other protonation states would be inconsistent with the structural data. The protonation state of His120 was suggested to be double protonated using MD simulations testing all three possible protonation states. Only the double protonation state did not show deviations from the starting X-ray structure (not shown). Regarding the catalytic His29, its neutral ε-protonated form can be excluded due to a clear structural H-bond involving the δ position as donor in all X-ray structures. We extensively simulated its

double-protonated and δ-protonated states (see below). The approximate calculations using PROPKA 3.1 [46] indicate tentative pKa value of 5.1 for the His29 in the 4AL5 X-ray structure, however, pKa calculations are always highly uncertain.

Despite the pH shifts in the crystalline conditions away from neutrality (4.6–5.0), these pH shifts should not change the dominant protonation states of any other amino acid or nucleobases.

2.3. Basic sets of simulations using the 4AL5 X-ray structure (4al5 and $4al5_{Met}$ simulation sets)

We performed two different sets of simulations based on the 4AL5 product structure: (i) simulations of *product form* (corresponding exactly to the X-ray structure; these simulations are abbreviated as "4al5") with the scissile phosphate being the terminal group, and (ii) *precursor* simulations with methyl group modeled and bonded to the O5′ of the scissile phosphate, mimicking part of the nt C21 in the precursor (hereafter termed "$4al5_{Met}$" systems), where the subscripted "Met" indicates the presence of the 3′-end methoxy-phosphate (MP3). Numerous attempts to model the complete C21 nucleotide resulted into rather unstable simulations (Supporting information). Thus, we decided to base our study on simulations with the smallest possible modeled group sufficient to study the cleavage reaction. We have no idea why it was not possible to prepare stable simulations with the full C21. Problems with the force field or the product experimental structure (or their combination) are possible. Supporting information Table S2 explains all abbreviations and labels used in this study.

2.3.1. His29 protonation states

Two possible protonation states of His29 (δ-protonated, His^0 or double-protonated, His^+) were considered, resulting in four subsets of simulations: 4al5(His^+), 4al5(His^0), $4al5_{Met}$(His^+) and $4al5_{Met}$(His^0). When we occasionally use abbreviations 4al5(His) and $4al5_{Met}$(His) not specifying the His charge we refer to all 4al5 or $4al5_{Met}$ simulations regardless of the protonation state of His29 (see Supporting information Fig. S4).

2.3.2. Orientations and protonation states of the scissile phosphate

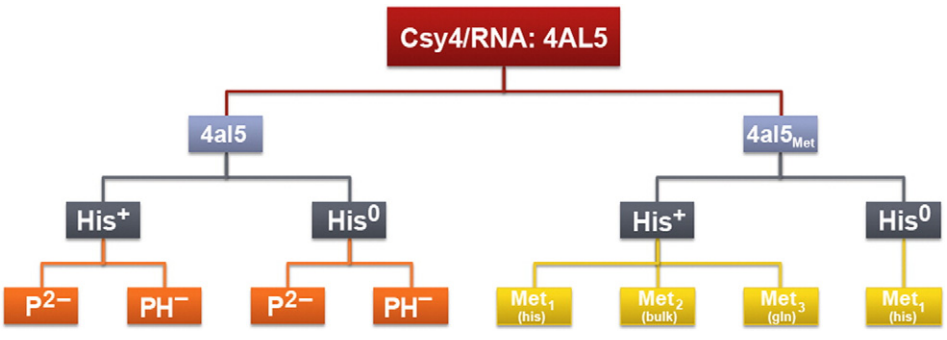
In the precursor $4al5_{Met}$ simulations, we considered three different initial orientations of the methoxy group with its O5′ oxygen: (i) oriented toward N_ε of His29, (ii) being oriented between these two residues and exposed to the bulk solvent, or (iii) oriented toward Gln149 backbone amido group. These three initial methoxy group orientations are marked as $4al5_{Met1}$, $4al5_{Met2}$ and $4al5_{Met3}$, respectively (see Supporting information Fig. S5 explaining details of the initial methoxy group placement). However, the methoxy group in simulations can easily rotate and its initial orientation does not seem to have any systematic impact on the simulations; all simulations irrespective of the starting geometry prefer the methoxy group oriented toward the His29. Therefore, only this starting orientation was used in His^0 simulations.

For the product 4al5 simulations, we considered both deprotonated double-charged phosphate (labeled as 4al5(His^0, P^{2-}) or 4al5(His^+, P^{2-})) and mono-protonated single-charged phosphate (4al5(His^0, PH^-) or 4al5(His^+, PH^-)). Protonation of all three phosphate oxygen atoms in the initial structure was tested (see Supporting information Fig. S6).

Thus, twelve different starting structures were used based on the product Csy4/crRNA complex 4AL5 (Scheme 1). Each simulation from the $4al5_{Met}$ set was verified by additional independent simulations while several independent simulations were additionally performed for the 4al5(His^+, P^{2-}) system (see Supporting information Scheme S1, Table S2 and Table S3 for more details about all simulations performed and the nomenclature used).

2.4. Basic sets of simulations using the 2XLK X-ray structure ($2xlk_A$ and $2xlk_B$)

We performed 6 simulations for each crystallographically independent biological unit (BU) of the 2XLK precursor X-ray structure (simulations using the first and second BU as start are marked as $2xlk_A$ and $2xlk_B$, respectively). The first set of simulations contained the inactivating G20 deoxy sugar, as observed in the 2XLK X-ray structure. In the second set we modified the X-ray structure to contain the native G20 ribose. In addition, despite that the crystallographic experiment included the whole terminal C21 nucleotide at the 3′-end RNA, only the scissile phosphate was resolved due to C21 disorder. Attempts to model the complete C21 nucleotide in the 2XLK X-ray structure again resulted into rather unstable trajectories (Figs. S7 and S8 in Supporting information). Thus, as in the case of $4al5_{Met}$ simulations, we added a methyl group to the O5′of the scissile phosphate, considering three starting positions (Fig. S5). For reasons in detail explained in the Results, we considered only double-protonated state for His29 residue (His^+, see Fig. S4). The preparation of the starting structure for the $2xlk_B$ simulations had an additional difficulty since its amino acid residues 13–15 and 134–138 were entirely disordered and invisible. We modeled them employing equivalent fragments from the first BU by superimposition of both biological units (RMSd 0.43 Å). See Scheme S1 in Supporting information for overview of simulations based on the 2XLK X-ray structure.



Scheme 1. Overview of simulations based on the 4AL5 X-ray structure. Blue boxes mark two different starting sets of structures. Product 4al5 structures are terminated by PO_3^{2-} or HPO_3^-. Precursor $4al5_{Met}$ structures have modeled methyl group bonded at nt21(O5′). Gray boxes mark protonation states of His29 (double-protonated, His^+ and δ-protonated, His^0). Orange boxes represent deprotonated double-negatively charged (P^{2-}) and mono-protonated single-negatively charged (PH^-) scissile phosphates in product simulations. For the 4al5(His^+, PH^-) and 4al5(His^0, PH^-) systems we used three different starting structures with hydrogens attached to different terminal phosphate oxygens, corresponding to atoms formally marked as OP1, OP2 and O5′ in the PDB file (obviously all these oxygen atoms are equivalent). This results into 8 basic 4al5 starting structures. Yellow boxes mark different orientations of the methyl group in the $4al5_{Met}$ starting structures (oriented toward His29, bulk or Gln149). However, the phosphate group is free to rotate, its initial orientation does not seem to be important and the $4al5_{Met1}$ orientation is ultimately preferred in all simulations independently of the start. Thus only one (the $4al5_{Met1}$) orientation was used at the start in simulations with His^0. Note also that when more than one simulation is done for the same starting structure, we mark them consecutively as Sim1, Sim2, … etc.

2.5. Simulation protocol

We combined simulation protocols tested for proteins and RNA, respectively [47,48]. Missing hydrogen atoms were added by LEaP module of the AMBER package [49]. The system was immersed in a rectangular SPC/E [50,51] water box with a 15 Å thick layer of water molecules around the solute, leading to box size ~130 × 85 × 75 Å^3. The systems were first neutralized with 8–10 Na^+ cations depending on the protonation state of His29 and scissile phosphate (Na^+ radius 1.212 Å and well depth 0.3526 kcal/mol) [52]. The cations were iteratively placed into the minima of the electrostatic potential calculated on a grid with spacing 1 Å using the program LEaP of AMBER 12.0 [53]. To provide more realistic ion background, excess of sodium chloride salt (0.3 mol/L, obtained by addition of 119 cations and anions. Cl^- has radius 2.711 Å and well depth 0.0128 kcal/mol) has then been added; cf. ref [54] for a discussion of ion treatment in simulations. The complete structures contained ~55,500 atoms. Parameters for all nonstandard residues, deprotonated and monoprotonated phosphate, as well as methyl group were developed according to RESP procedure by Cornell et al. [39,55] (see Supporting information for more details, together with parameters and prep files). The RNP-solvent system was minimized prior to the simulation as follows. Minimization of the solute hydrogen atoms was followed by minimization of counterions and water molecules. Subsequently, the protein/RNA system was frozen and solvent molecules with counterions were allowed to move during 10 ps long MD run in order to relax the density of the system. After that, the system was relaxed in several minimization runs with decreasing force constant applied to the RNA and protein backbone. After full relaxation, the system was slowly heated to 298.15 K over 50 ps using 2 fs time steps and NpT conditions using a weak-coupling scheme with a coupling time of 1 ps [56]. The production phases of MD simulations were carried out under periodic boundary conditions (PBC) in the NpT ensemble (298.15 K, 1 atm) with 2 fs time steps. The SHAKE algorithm was applied to constrain all bonds containing hydrogen atoms. The PMEMD module of AMBER 12.0 [53] was used for simulations.

We use the modern *ff99bsc0*χ_{OL3} variant of the RNA Cornell et al. force field [39,40] with bsc0 [41] and χ_{OL3} [42] corrections to the α/γ and χ torsions, respectively, combined with the *ff*99SB protein force field [39,43]. Two additional simulations were performed with the latest AMBER [53] protein force field *ff12SB* which has extensively modified side chain torsions. This force field suddenly appeared in the AMBER12 code as the suggested default, but unfortunately has not yet been documented in the literature and no tests are available. While this work was under review, ff12SB protein force field has been replaced by ff14SB which also has not yet been documented. The cumulative simulation time of all simulations was 3.7 μs (most simulations 50 ns).

The ab initio calculations required for the parameterization of all three nonstandard residues were carried out using Gaussian 09 [57] at the HF/6-31G(d) level of theory. Analyses of trajectories were performed by Ptraj (from AmberTools package) program. The VMD [58] and PyMOL [59] programs were used for visualizations.

2.6. Restrained simulations

Although our paper is primarily based on unrestrained simulations, for reasons explained in the Results section we also performed several restrained simulations, with restraints applied to several dihedral angles of the S148, S150 and T151 amino acids and the pucker of G20 nucleotide. The restraints are applied using the standard flat-well approach. Specifically, pucker was restrained via τ_2 torsion, with no penalty below −20°, quadratic restraining potential in the −20° to 0 region with force constant 10 kcal/mol · rad^2 and continuing linear penalty above 0°. The restrained simulations are summarized in Supporting information (Table S4).

2.7. H-bond network analyses

The individual H-bond interactions were primarily analyzed by measuring the interatomic distance between the heavy atoms (3.2 Å was used as the criterion). We have also always simultaneously monitored the H-bond angles, although in the paper we report them only when necessary, e.g., in detailed analyses of the catalytic center. Note that the distance analysis would be in principle sufficient for comparison with the experimental structures, since experiments do not reveal hydrogen positions.

3. Results

In the following text we first describe the product simulations, which allow to assess the dominant protonation states of 4AL5 X-ray structure and to obtain insights into accuracy limits (length of simulations, properties of the starting structure and force field) of the MD simulation technique. Then, based on the precursor simulations, we propose a plausible structure, which is in all aspects consistent with the Ser148 general base mechanism.

3.1. The overall simulation behavior

Overall structural dynamics of the Csy4/RNA complex were similar in all simulations. The ferredoxin-like domain of the protein and the stem region of the RNA were rather rigid. There were two flexible regions at the Csy4/RNA interface, namely, (i) the Arg-rich helix α3 together with the RNA pentaloop in the upper part of the complex and (ii) the β6–β7 hairpin loop together with the 3′-end of RNA at the active site (Fig. 1A). The individual simulations often visibly differed in these regions. In the following analyses, we primarily concentrate on the second region due to its direct catalytic relevance. In addition, there were other regions, e.g., the linkers between α2 and β5 that also fluctuated (Fig. 3). The fluctuations and rearrangements of protein and RNA were coupled, i.e., movements of the protein affected the RNA and vice versa.

Fig. 4 shows the RMSd data for six product 4al5 and precursor-like $4al5_{Met}$ simulations. In general, all simulations show similar overall RMSd developments (see Supporting information Fig. S9).

3.2. Simulations of the unmodified product 4AL5 structure

Simulations of the Csy4/crRNA product form (the 4al5 simulation set) with different protonation states of the His29 and terminal phosphate (Scheme 1) were made in order to determine dominant protonation states corresponding to the 4AL5 X-ray structure. We considered δ-protonated (His^0) or double-protonated (His^+) His29 residue, and mono-protonated (PH^-) or deprotonated (P^{2-}) 3′-end phosphate. Different initial positions of the hydroxyl group of the protonated phosphate were tested (Fig. S6), resulting into altogether eight distinct initial structures (Scheme 1). These simulations were also used to assess the performance of the simulation methodology, since this is the only set of simulations simulating the experimental structure without any modification. We assumed that the right protonation state should lead to the smallest deviations from the experimental structure. The active site contains two important interaction networks, namely, interactions between the His29 and the 3′-end of RNA, and hydrogen bond interactions involving triad residues of the β-hairpin loop (Ser148, Ser150 and Thr151) and nt G20 (Fig. 5). Direction of majority of the H-bonds in the experimental structure can be determined quite unambiguously. Thr151 appears to act as a donor to Ser148 which in turn acts as a donor to G20(O2′). Ser150 acts as a donor to G20(O3′). The interactions of Ser148 and Ser150 with the ribose likely fix the mutual position of the serine side-chains. This interaction network thus implies that G20(O2′) acts as a donor to one of the oxygen atoms of the nt21 phosphate. Reversal of this H-bond network would require deprotonation of the Ser148 which is unlikely in case of the product structure. It would also be

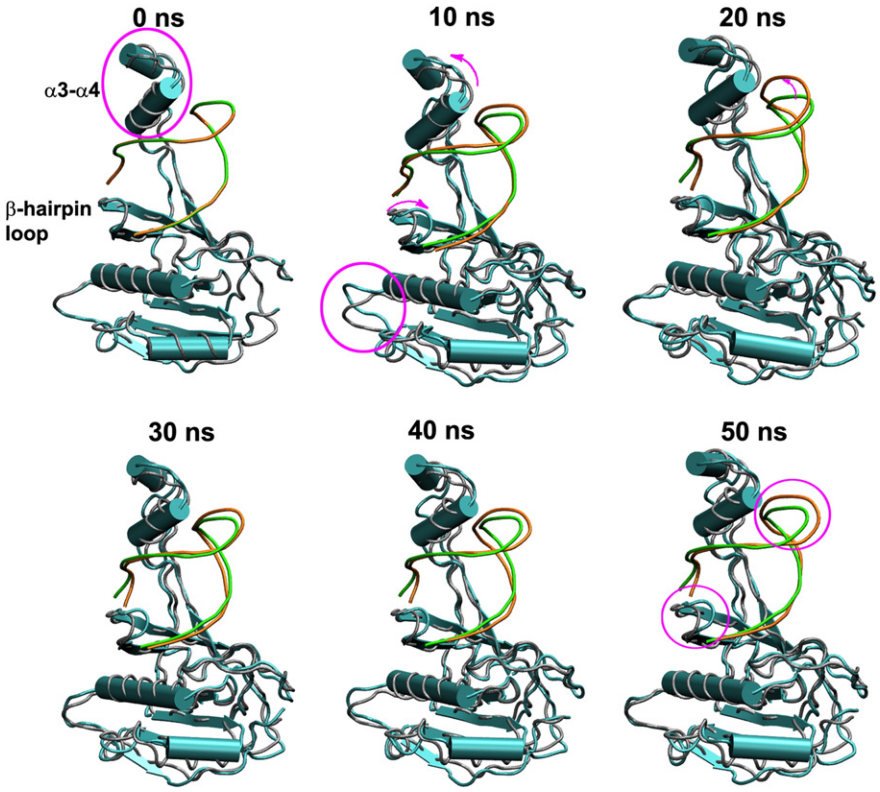


Fig. 3. Overlay of the 4AL5 X-ray structure (gray and green tubes) and 4al5(His^+, P^{2-}) MD structure (cyan cartoon and orange tube) at different simulation times illustrating (i) vertical fluctuations of the α3–α4 segment and (ii) fluctuations of the β-hairpin loop. The magenta arrows show the direction of the dominant dynamics while the magenta circles highlight the most flexible areas.

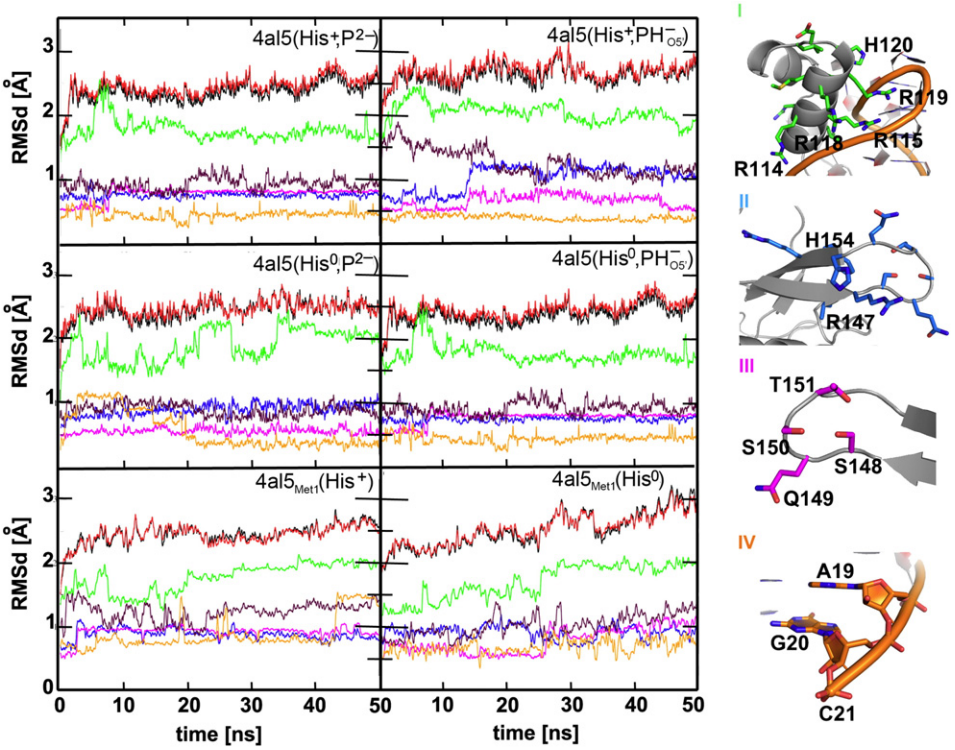


Fig. 4. RMSd profiles for selected simulations with respect to the starting 4AL5 X-ray structure: the whole Csy4/RNA complex (black line), the Csy4 protein (red line), the RNA (maroon) and four small regions at the active site, namely regions I (green), II (blue), III (magenta) and IV (orange). All residues depicted in stick representation (right) were used for the RMSd calculations of the regions I–IV. The simulation names are given in the upper right corners; the O5′ subscript indicates to which atom of the PDB file the terminal hydrogen has been added. The RMSD was calculated considering all heavy atoms of the selected residues.

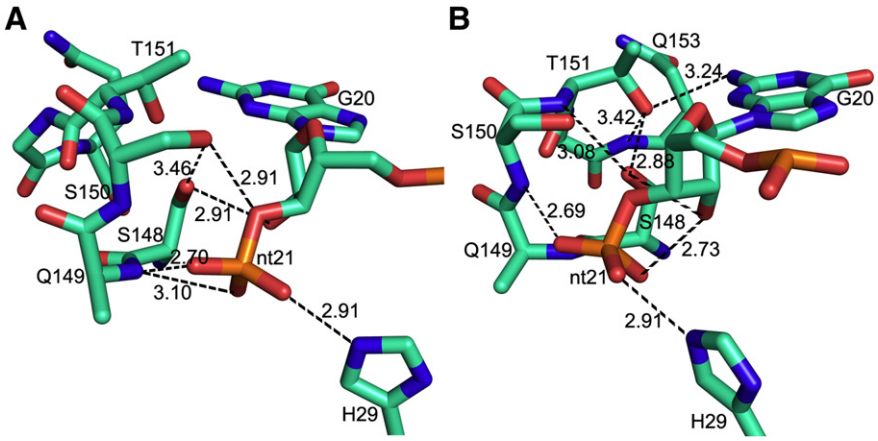


Fig. 5. Front view (A) and side view (B) of the active site of 4AL5 X-ray structure showing the intricate network of interactions between His29, the 3′-end of RNA and the residues of the β-hairpin loop.

incompatible with the observed interaction between G20(O2′) and the nt21 phosphate.

3.2.1. The 4al5(His^+, P^{2-}) Sim-1 simulation achieved reasonable but not full agreement with the 4AL5 X-ray structure

As explained below, the 4al5(His^+, P^{2-}) simulations with double-protonated His29 and deprotonated scissile phosphate were the most consistent with the X-ray data. Let us first discuss one of the 4al5 (His^+, P^{2-}) simulations (marked as Sim-1) which revealed the best agreement with the X-ray structure and which was ultimately extended to 300 ns. We examined key interactions in three crucial segments, namely, (i) interactions at the active site (Figs. S3D and E and 5), (ii) interactions between the protein and the major groove of RNA (Fig. S3B) and (iii) interactions of the recognition region (Arg-rich helix α3/RNA pentaloop, Fig. S3C). In the main text we report structural dynamics of the active site while segments (ii) and (iii) are described in Supporting information, Table S5, Figs. S10 and S11 and the accompanying text.

There are potential contacts between His29 amino acid and the O2′ and O3′ atoms of G20. However, in the product crystal structure these distances are too long to form hydrogen bond interactions. In the 4al5 (His^+, P^{2-}) Sim-1 simulation these His29–G20 distances shortened but H-bonds were still not established (see Fig. 6A, middle and 7). However, the His29 came even closer to the terminal phosphate compared to the X-ray structure. In fact, due to adjustment of the phosphate position, the simulation established double (bifurcated) H-bond between His29(N_ε) and the oxygen atoms of nt21 phosphate while only one H-bond was suggested by the X-ray data (Fig. 6B).

Figs. 6–8 illustrate that the simulated system dramatically departed from the starting structure immediately after the simulation start, with some interatomic distances (namely between the β-hairpin loop and the RNA) deviating as much as 5 Å from their starting values. However, the system returned close to the starting structure after ~20 ns, although the catalytic center was somewhat rearranged compared to the experimental structure. Then, despite some fluctuations, the simulated system remained essentially unchanged and local differences of the β-hairpin loop interactions compared to the X-ray structure persisted till the end (Figs. 6A, 7A–B, C). The Ser148(O_γ)–Ser150(O_γ) distance increased from 3.5 Å to 4–5 Å (Fig. 6A, right and Table S6). This was related to change of interactions between the β-hairpin loop and G20, namely, weakening of Ser148(O_γ)–G20(O2′) and loss of Ser150(O_γ)–G20(O3′) interactions (Fig. 6A, left). Water density analysis of the Sim-1 simulation revealed a hydration site between the 3′-end of RNA and the β-hairpin loop which was not seen in the 4AL5 X-ray structure. Interestingly, similar hydration site was found in the 2XLK inactivated precursor crystal structure having less dense network of H-bonds due to the absence of the 2′-hydroxyl group of G20.

Apparently, the distortion of the β-hairpin loop was initiated by rearrangements of the triad residues, which subsequently propagated to G20. To achieve a stable hydrogen bond network, the hydroxyl groups of the β-hairpin loop triad should be oriented toward the catalytic pocket. Experimental orientations of the hydrogens (the χ_2 angles, C_α–C_β–O_γ–H_γ) are obviously not known. Nevertheless, the proper disposition of hydrogen bond network would require χ_2 *gauche*$^+$, *gauche*$^-$ and *gauche*$^-$ conformations of Ser148, Ser150 and Thr151, respectively (see Supporting information and Fig. S12 for further details). Fig. 8A shows that after ~20 ns the χ_2 angles reached the expected *gauche*$^+$ and *gauche*$^-$ conformation for Ser148 and Thr151, respectively. Also the χ_1 dihedral angles for Ser148 and Thr151 are maintained after the first 20 ns at the experimental values. However, the χ_1 dihedral angle of Ser150 flipped from the X-ray *gauche*$^+$ value to *gauche*$^-$ after few picoseconds and remained in this orientation till the end of the simulation.

The Ser150 χ_2 dihedral angle oscillated between *gauche*$^+$, *trans* and *gauche*$^-$ orientation during the simulation, although the *gauche*$^+$ orientation was the most populated (Fig. 8A, middle), allowing formation of an alternative arrangement of the catalytic center. The χ_1 flip of Ser150 was related to the elongation of the Ser148(O_γ)–Ser150(O_γ) distance and loss of the Ser148(O_γ)–G20(O2′) interaction, leading to the re-conformation of the β-hairpin loop and modified local H-bond network compared to the experimental structure.

Since the Sim-1 simulation started with the *trans* χ_2 angles of the triad, we run three 5 ns simulations where all χ_1 and χ_2 dihedral angles of the triad initially had the correct (expected) values, i.e., χ_1 Ser148 (180°), χ_2 Ser148 (g^+), χ_1 Ser150 (g^+), χ_2 Ser150(g^-), χ_1 Thr151(g^+) and χ_2 Thr151(g^-). The flip of χ_1 of Ser150 occurred again and the starting orientation of the hydrogens did not affect the simulations.

The initial distortion of the β-hairpin loop was correlated with orientation and movement of the 2′-hydroxyl group of G20. Specifically, the most visible movement of 2′-OH coincided with C2′-endo to C3′-endo re-puckering of G20 during the first 20 ns. This may facilitate the initial loss of Ser148(O_γ)–G20(O2′) and Ser150(O_γ)–G20(O3′) interactions. However, after 20 ns of the simulation the G20 pucker returned to C2′-endo conformation. However, since the χ_1 angle of Ser150 remained flipped to *gauche*$^-$, the catalytic region remained somewhat rearranged (Figs. 7A–C and 8B) compared to the experimental structure.

The local re-conformation of the β-hairpin loop did not involve any significant flip in the backbone torsion angles of amino acids and did not alter their position on the Ramachandran plot. The overall impression is that the simulation at the beginning eliminated some energy imbalance (potentially some conflict between the interaction network in the starting structure and its force field description) and then it relaxed to similar but not identical geometry as seen in the experiment. Thus,

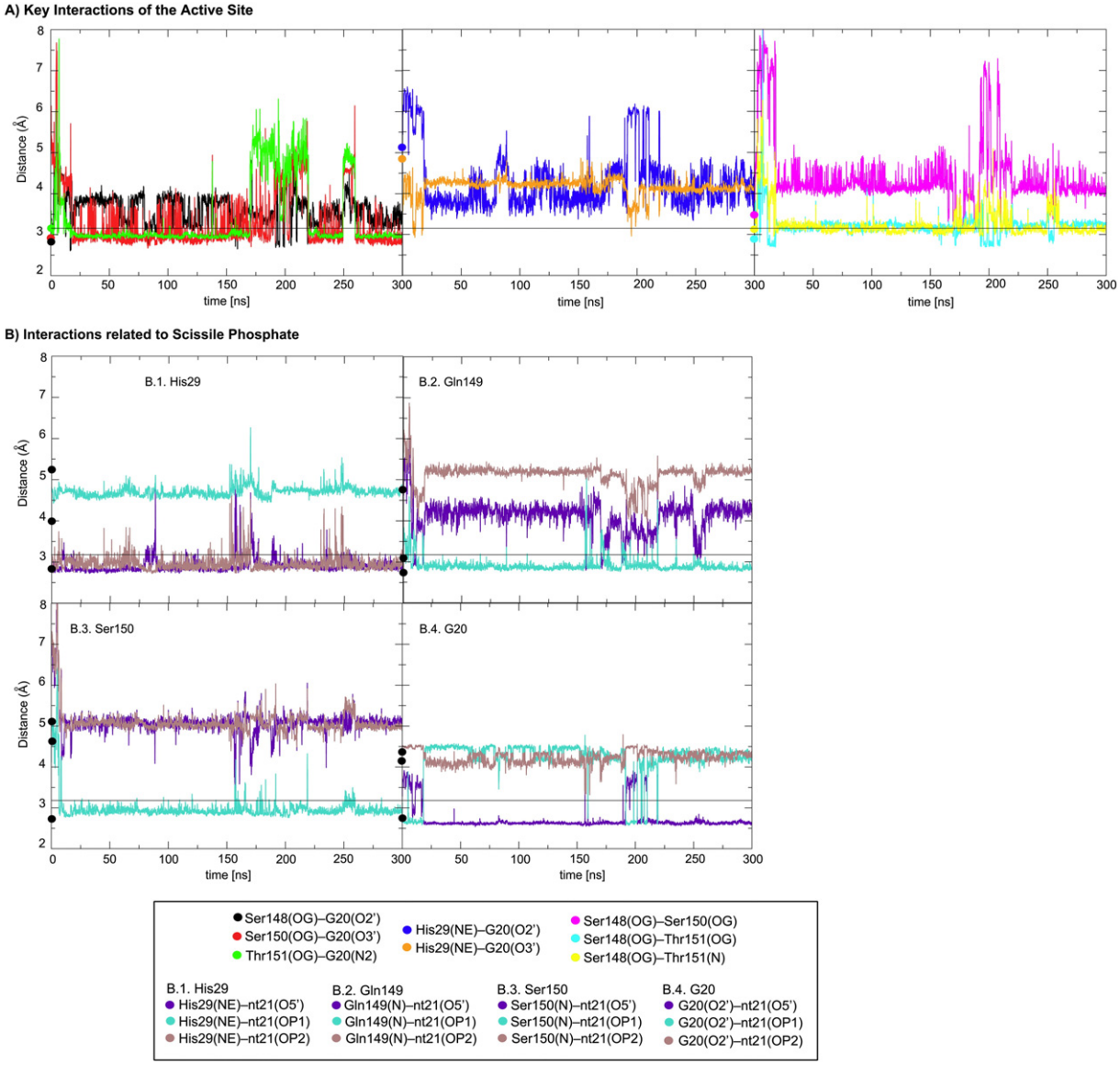


Fig. 6. Time evolution of the most significant distances in the 4al5(His^+, P^{2-}) Sim-1 simulation. The simulated system dramatically deviated from the starting structure during the first ~20 ns but then it partly returned back. The experimental values are given by the dots located on the y-axes while the colors are explained at the bottom of the figure. The thin horizontal lines mark the 3.2 Å distance. (A) The key interactions of the active site. Left: interactions between the β-hairpin loop and the 3′-end of RNA. Middle: interactions between the His29 and the 3′-end of RNA. Right: distances between the triad residues Ser148, Ser150 and Thr151. (B) The interactions of the three oxygen atoms of the scissile phosphate at nt21 with His29 (B.1), Gln149 (B.2), Ser150 (B.3) and G20 (B.4). Note that the three oxygen atoms of the phosphate group are chemically equivalent. In the simulation description we maintain their formal numbering (O5′, OP1 and OP2) as given in the PDB ID: 4AL5 file, and we use different colors for their interactions to visualize rotation of the phosphate group in the simulations. However, the three experimental values are marked by black dots to reflect that the oxygen atoms are equivalent and thus indistinguishable.

the 4al5(His^+, P^{2-}) Sim-1 simulation achieved qualitative but not quantitative agreement with the X-ray structure.

3.2.2. Additional 4al5(His^+, P^{2-}) simulations

To check the reproducibility of the 300 ns Sim-1 4al5(His^+, P^{2-}) simulation, we performed four additional 50 ns simulations from the same starting structure (simulations Sim-2 to Sim-5, see Figs. S12–S21 and Table S4 in Supporting information for the meaning of the nomenclature).

The Sim-3 simulation showed essentially the same behavior as the Sim-1 simulation. There was a large initial perturbation of the structure compared to the X-ray data. However, after 20–25 ns most of the distances returned back or close to the experimental values (Fig. S14 in Supporting information). The χ_1/χ_2 *gauche*$^-$/*gauche*$^+$ conformation of Ser150 again did not allow full restoration of the Ser150 contacts (see details below and Fig. S15 in Supporting information).

The Sim-2 and Sim-5 simulations deviated from the X-ray structure during the whole 50 ns trajectories (Figs. S16–S19 in Supporting information), with a progressive increase of the distances between the β-hairpin loop and the 3′-end of RNA (Figs. S16 and S18). The G20 sugar pucker was in C3′-endo conformation for most of the time, and χ_1 and χ_2 dihedral angles for all residues of the triad were unstable (Figs. S17 and S19). Finally, the Sim-4 simulation showed visibly smaller deviation from the X-ray data compared to the Sim-2 and Sim-5 simulations, but was not as close to the X-ray data as the Sim-1 and Sim-3 simulations (Figs. S20 and S21).

We suggest the following interpretation of the data. After initial large perturbation seen in all simulations two simulations (Sim-1 and Sim-3) relaxed almost back to the 4AL5 crystal structure arrangement. Two simulations (Sim-2 and Sim-5) deviated progressively from the experimental structure at the 50 ns time scale. The remaining simulation showed intermediate behavior. Obviously, the results were far from being converged. Based on recent simulations of small RNAs it is clear that there was no chance to obtain converged results for this large and exceptionally intricate system [60,61]. Nevertheless, we suggest that

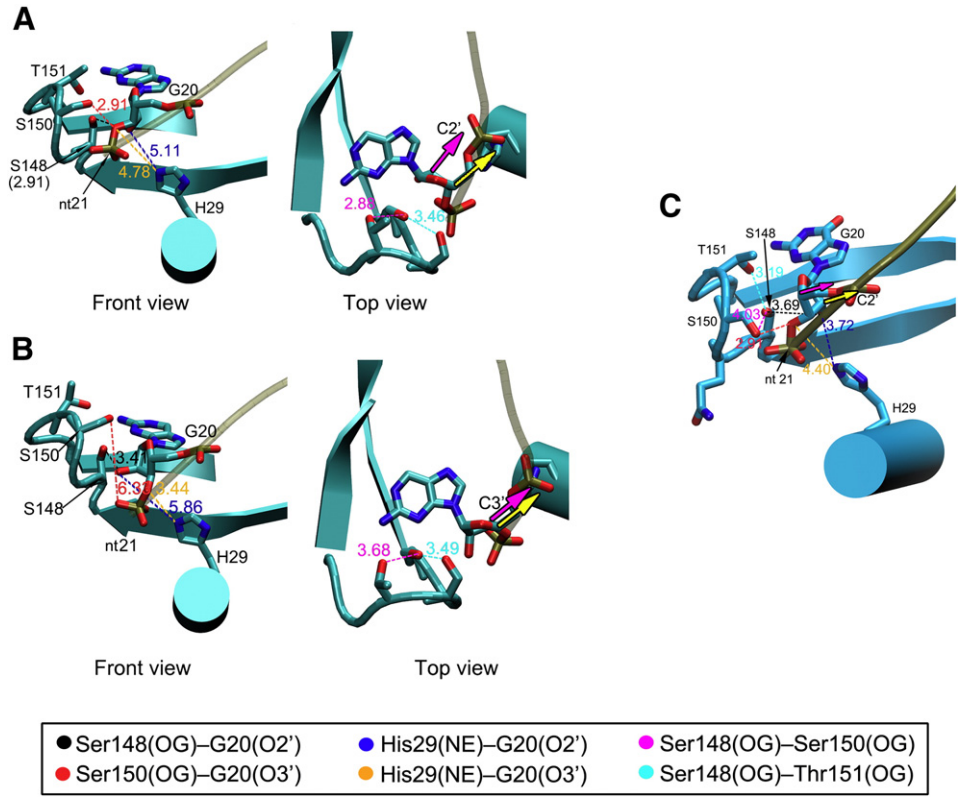


Fig. 7. The enlarged views of the active site show the evolution of some distances at different times of the Sim-1 simulation. See Table S6 in Supporting information for further details. (A) Interactions of the active site in the 4AL5 X-ray structure. (B) Interactions of the active site after 150 ps of the simulation. (C) Interactions of the active site after 50 ns. See the legend at the bottom of the figure for the colors used. The magenta and yellow arrows depict the change of the ribose pucker.

the behavior of simulations Sim-1 and Sim-3 was the most representative one, because the capability of the simulated structure to come close to the starting geometry after a profound geometry perturbation is a considerably more significant observation than mere stability of the starting structure [31]. Based on all simulations we suggest that the flip of χ_1 and χ_2 dihedral angles of Ser150, when the remaining dihedral angles were at the experimental values, was related to local imperfectness of the simulated structure (Supporting information Figs. S12–S21 and Table S4).

3.2.3. Double-protonated His29 and deprotonated scissile phosphate represent the dominant protonation state consistent with the 4AL5 X-ray structure

We suggest that the (His^+, P^{2-}) protonation state is the dominant protonation state consistent with the product X-ray structure because all simulations performed with different protonation states showed unstable behavior. None of them was comparable with the best behaving 4al5(His^+, P^{2-}) simulations.

The mono-protonated His29 decisively moved away from the nt21 phosphate in all simulations, losing all interactions with the 3′-end RNA. This created a large cleft between the N-ferredoxin-like domain (where His29 is located) and the RNA stem/C-terminal domain segment directly participating in the active site (see Fig. 9 and Supporting information Figs. S22 and S23 and Table S6 for further information). This was observed even in all three 4al5(His^0, PH^-) simulations, where the His29 side-chain positioned near the phosphate hydroxyl group could in principle accept hydrogen bond from this hydroxyl group. Thus, singly protonated His29 is not consistent with the 4AL5 structure.

In contrast, in 4al5(His^+) simulations the double-protonated His29 remains in proximity of the scissile phosphate (Fig. 9B, middle and right). However, there were differences between the 4al5(His^+, P^{2-}) and 4al5(His^+, PH^-) simulations in the β-hairpin loop and in the Arg-rich helix α3/RNA pentaloop segment (Fig. 9A, middle and right). All 4al5(His^+, PH^-) simulations showed large deviations of the β-hairpin loop from the X-ray geometry. Specific example is shown in Figs. S24 and S25 in Supporting information. In the X-ray structure the phosphate oxygen atoms of P21 formed five tight H-bond interactions with key residues (His 29, Gln149 and G20) which were seen also in well-behaving 4al5(His^+, P^{2-}) simulations (Fig. 6) but became unstable after the P21 protonation (Fig. S25).

3.2.4. Summary of β-hairpin loop deviations

As discussed above, majority of 4al5(His^0) and all 4al5(His^+, PH^-) simulations revealed significant deviations of the β-hairpin loop with respect to the X-ray structure. Let us now briefly describe hallmarks of this deviation. The β-hairpin loop rearrangements created more space for free movements of the scissile phosphate, which could then adopt two distinct positions, being either close to His29 residue or close to the β-hairpin loop. The second position always severely distorted the active site (see Fig. 9B, left). We usually observed that the movement of the scissile phosphate was also accompanied by the rearrangement of Gln149 residue which severely affected the β-hairpin loop, and thus the arrangement of the active site. In the experimental structures (both product and precursor) Gln149 acted as an anchoring point due to a hydrogen bond between its backbone nitrogen and the 3′-end phosphate (Fig. S3E). If this H-bond formed in simulations, the conformation of the β-hairpin loop was closer to the X-ray structure albeit, as explained above, never in perfect agreement (see Fig. 9B, middle). However, if side chain hydrogen bond interaction between Gln149(N_ε) and the scissile phosphate was established instead, the β-hairpin loop was shifted away from the 3′-end of RNA. The hydrogen bond interactions between the β-hairpin loop triad residues and the 3′-end RNA were lost and the β-hairpin loop was distorted (see Fig. 9A, left and right panels). For more details, see also below the $4al5_{Met}$ simulation section. None of the above-noted movements

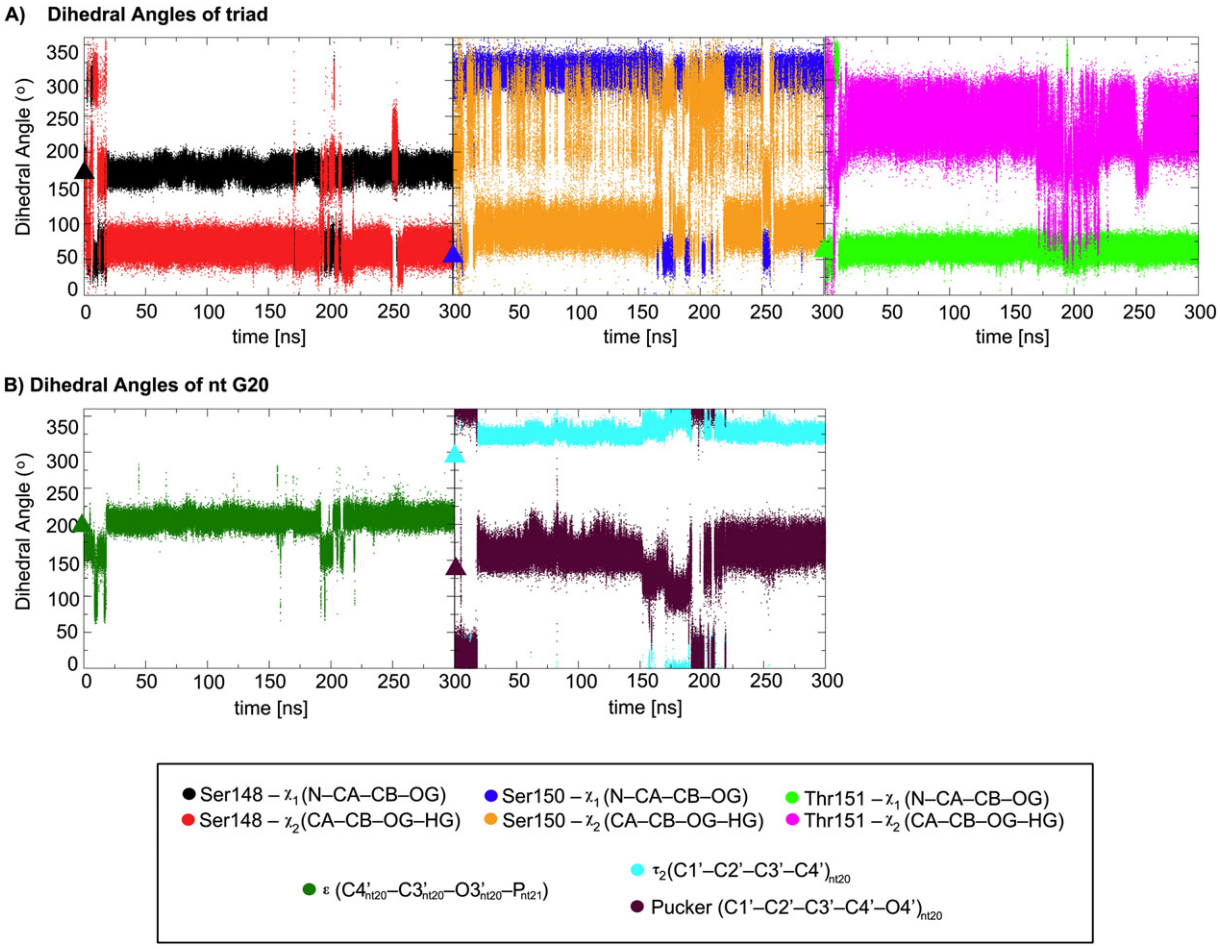


Fig. 8. Time evolution of the most significant dihedral angles in the 4al5(His$^+$, P^{2-}) Sim-1 simulation. (A) The χ_1 and χ_2 dihedral angles for triad residues of the β-hairpin loop. Left: the χ_1 and χ_2 of Ser148. Middle: the χ_1 and χ_2 of Ser150. Right: the χ_1 and χ_2 of Thr151. (B) Key dihedral angles to define the sugar pucker of nt G20. Left: ε dihedral angle of G20. Right: τ_2 dihedral angle and sugar pucker of G20. The triangles located on the y axes of the graphs show the 4AL5 experimental values where relevant.

occurred in the two stable 4al5(His$^+$, P^{2-}) simulations Sim-1 and Sim-3 (Fig. 9B, middle).

3.2.5. Restrained simulations of the 4al5(His$^+$, P^{2-}) complex

The most stable 4al5(His$^+$, P^{2-}) simulations Sim-1 and Sim-3 provided the following picture: (i) unfavorable orientation of the side-chains of the triad (mainly χ_1/χ_2 dihedrals of Ser150) preventing full restoration of the hydrogen bond network with the RNA suggested by the X-ray structure; (ii) re-puckering of G20 ribose from C2′-endo to C3′-endo conformation at the beginning of the simulations and its return back, and (iii) the accompanying notable fluctuations of the position of the 2′-hydroxyl group of the G20 ribose in the first part of the simulations. These factors significantly affected the interactions between both serines and G20. These interactions were difficult to be established simultaneously.

We conducted a set of additional simulations in which we attempted to stabilize the scissile phosphate H-bond network using combination of various restraints. Rather than restraining the individual H-bonds, we tried to stabilize the initial conformations of several dihedral angles, whose changes were coupled with the disruptions of the H-bond network. While these simulations provided some insights, we were unable to stabilize all of the interactions. The results are discussed in details in the Supporting information (Figs. S26–S33 and text description on pages S44–S46). We tested restraints of the χ_2 dihedral angle (C_α–C_β–O_γ–H_γ) to *gauche*$^+$, *gauche*$^-$ and *gauche*$^-$ orientation for Ser148, Ser150 and Thr151, respectively, restraint of the τ_2 dihedral angle of G20 ribose (negative values of τ_2 correspond to the C2′-endo pucker) and restraint of η_{2}' dihedral angle (C1′–C2′–O2′–HO2′). The inability of the dihedral restraints to stabilize the observed arrangement indicated that the simulation behavior should not be attributed to some straightforward errors in the force field dihedral terms.

The simulations confirmed the role of sugar pucker of nt G20. If the pucker changes to C3′-endo conformation and was maintained in this state, the key active site interactions were completely lost. The C2′-endo conformation stabilizes some but not all key interactions. It was difficult to simultaneously obtain the right C2′-endo pucker of G20 ribose and gauche$^+$ χ_1 conformation of Ser150. We observed the following combinations of the χ_1/χ_2 dihedrals of Ser150 during the simulations: (i) *gauche*$^+$/*trans*, (ii) *gauche*$^-$/*trans*, (iii) *gauche*$^+$/*gauche*$^-$ (the expected native arrangement), (iv) *gauche*$^-$/100°, (v) *gauche*$^-$/*gauche*$^+$ and (vi) *gauche*$^-$/*gauche*$^-$. The *gauche*$^+$/*gauche*$^+$ combination was not populated. Obviously, if χ_1 flips, χ_2 must follow to establish a similar (but not identical) H-bond network. Ser150 χ_1/χ_2 angle combinations (iii), (iv) and (v) were those that more favorably affected the monitored interactions. When these χ_1/χ_2 combinations occurred, the Ser148(O_γ)–G20(O2′) and Ser150(O_γ)–G20(O3′) interactions showed proper distances (Figs. 6 and 8). Fig. S33 shows the native (but essentially not supported by simulations) *gauche*$^+$/*gauche*$^-$ combination (iii) allowing fully correct arrangement of the active site. The combinations (iv) and (v) could be considered as conformations allowing viable alternative organization of key active site interactions, although they are not fully consistent with the X-ray data.

3.2.6. 4al5(His$^+$, P^{2-}) simulations with the ff12SB protein force field

We have also performed two simulations with the *ff12SB* protein force field (Tables S4 and S7 and Figs. S34–S36, see Methods) [53].

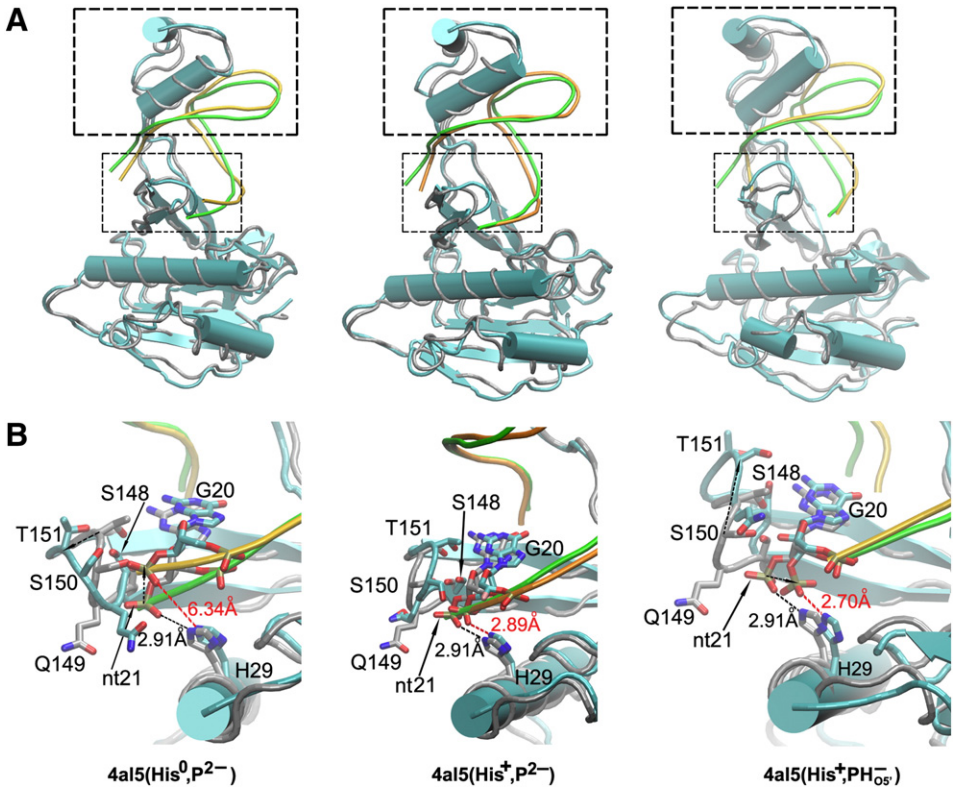


Fig. 9. (A) Average structures from the last nanosecond (bluish cartoons and orange tube) superimposed over the starting structure (gray and green tubes) for the 4al5(His^0, P^{2-}) (left), 4al5(His^+, P^{2-}) Sim-1 (middle) and 4al5(His^+, $PH^-_{O5'}$) (right) simulations. Key deviations are highlighted by dashed black rectangles. The best agreement with the crystallographic active site is seen in the 4al5(His^+, P^{2-}) simulation (middle). (B) Enlarged views of the active site in the indicated simulations. Note the cleft in the 4al5(His^0, P^{2-}) simulation visualized by the N_ε(H29)–O5′(nt21) distance (red dashed line — the simulation, black dashed line — the X-ray value) and the large deviation of the β-hairpin loop in the 4al5(His^+, $PH^-_{O5'}$) simulation with respect to the starting structure, which occurs in all simulations with singly-protonated phosphate (PH^-). Large movements at the active site are indicated by dashed black arrows.

The first simulation resulted in a complete distortion of the active site with the characteristic rearrangement of the Gln149–P21 interactions described earlier in this paper (Fig. S34A). The second simulation, which we extended to 300 ns, was also worse than the best simulations utilizing the *ff99SB* force field (Figs. S35 and S36). The χ_2 dihedral angle term of Thr151 in the *ff12SB* protein force field disfavors the *gauche*$^-$ orientation by ~1.5 kcal/mol with respect to the *ff99SB* version. This might explain why the χ_2 dihedral angle of Thr151 sampled *trans* orientation during this simulation. Thus, the *ff12SB* force field refinement was not sufficient to improve the description of the catalytic center, which again indicated that the deviations from the 4AL5 structure should not be attributed to trivial errors in the force field dihedral terms. Our unpublished data for some other RNP systems indicate that the *ff12SB* shows somewhat improved behavior over the *ff99SB*, although, in general we do not see large differences between these two protein force fields in simulations of RNP systems.

3.2.7. Summary of the 4al5 simulation set

To summarize the results, the best agreement with the 4AL5 crystal structure was obtained in some 4al5(His^+, P^{2-}) simulations. We suggested that these simulations identified the dominant protonation state of the product X-ray structure. However, complete agreement with the active site of the 4AL5 structure was not achieved. We suggest that it may indicate some force field imbalance. However, the overall behavior of all simulations indicates that such potential force field imbalance should not be straightforwardly attributed to some very few specific force field terms. We tentatively suggest that the simulations may be affected also by the intermolecular terms and cannot be corrected by refinements of the intramolecular dihedral terms (see Discussion). In addition, we cannot rule out that the simulations were affected by properties of the starting structure.

3.3. Potential catalytic geometry with Ser148/His29 acting as the general base/acid identified in $4al5_{Met}$ precursor simulations

Our first model of the precursor structure was based on the 4AL5 product X-ray structure. The advantage of using the 4AL5 structure is the native G20 ribose, which is important to establish the H-bond network in the catalytic pocket. However, the catalytic relevance of the 4AL5 structure can be compromised by the absence of the C21 residue. We mimicked the C21 residue by modeling a methyl group into the 4AL5 X-ray structure, yielding the 3′-methoxyphosphate terminus, abbreviated as $4al5_{Met}$ (see Methods).

We carried out two simulations (100 ns and 50 ns) for each starting structure (Fig. 10). $4al5_{Met}$ simulations with double-protonated His29 (His^+) were run with initially three different orientations of the methoxy group (see Methods and Fig. S5), however, all these simulations clearly preferred orientation of the methoxy group toward His29 suggesting that protonated His29 might interact in pre-cleavage state with the leaving O5′ oxygen. For the $4al5_{Met}$(His^0) simulations we considered only this starting orientation (see Figs. S5B and S37–S56 in Supporting information).

One of our six $4al5_{Met}$(His^+) simulations spontaneously adopted a stable substate which looks very relevant for catalysis. In the last ~10 ns of the 50 ns $4al5_{Met2}$(His^+) simulation, stable simultaneous short contacts between Ser148(O_γ)–G20(O2′), Ser150(O_γ)–G20(O3′) and Ser148(O_γ)–Ser150(O_γ) were established (Figs. 11 and S45 and S46). This was the only simulation in the whole our study where these short interatomic contacts formed simultaneously. Phe155 remained stacked below G20 and Tyr176 stacked with His29. Thr151 interaction with G20(N2) was lost. This, however, was consistent with the inactivated precursor X-ray structures. Most importantly, the Ser148 hydroxyl group was reoriented away from the G20(O2′) nucleophile

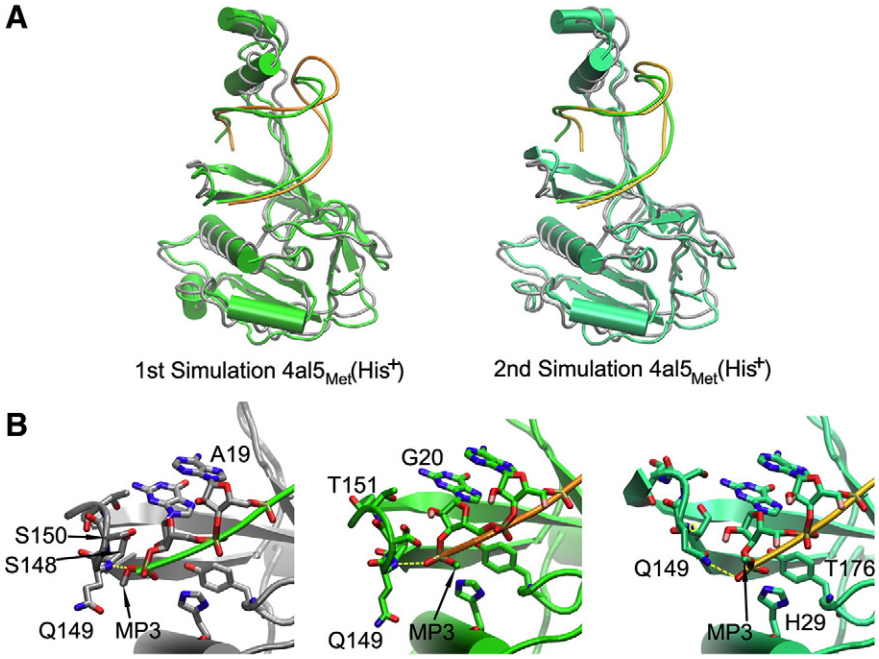


Fig. 10. (A) Green/orange (protein/RNA) cartoon diagrams showing the average structure of the last nanosecond of the two independent $4al5_{Met3}(His^{+})$ simulations superimposed over the initial structure (gray and green tubes). (B) Left; arrangement of the active site based on the 4AL5 X-ray crystal structure with modeled C5′-methyl group (MP3) of the nt21. Middle; the active site at the end of the 100 ns $4al5_{Met3}(His^{+})$ simulation with the side chain of Gln149 oriented outwards the catalytic pocket. The native hydrogen bond interaction established between Gln149 backbone and the scissile phosphate is highlighted in yellow. Right; the active site at the end of the 50 ns $4al5_{Met3}(His^{+})$ simulation with the side chain of Gln149 oriented toward the catalytic pocket. The interaction highlighted in yellow is the non-native side-chain Gln149(N_{ε})–nt21(Ph) H-bond, which formed after the departure of the β-hairpin loop. The catalytic pocket was disrupted. Altogether, of the six $4al5_{Met}(His^{+})$ simulations (cf. Scheme 1 in Methods), three ended up with rather stable catalytic core (as in B, middle) and three with disrupted catalytic core (as in B, right).

and fluctuated between donating a H-bond either to the Thr151 side chain or to a water molecule (Fig. 12). Despite the loss of the Ser148–G20 ribose H-bonding, the Ser148 hydroxyl and the G20(O2′) nucleophile amazingly remained in a close contact, due to all the other interactions cementing the catalytic center (Fig. 12). Upon H-bonding of the Ser148 hydroxyl group to Thr151, Thr151 side chain acted as H-bond donor to water molecules and sometimes also to Gln153($O_{\varepsilon 1}$).

In this arrangement, Ser148 could be straightforwardly deprotonated either by a water molecule (Fig. 12) or perhaps via Thr151 side chain deprotonated by a water molecule (not shown). Then, Ser148 could act as the general base and deprotonate the G20(O2′) nucleophile. Further, the RNA backbone was sampling near in-line attack configurations (Fig. S56 in Supporting information) while the double-protonated His29 established a hydrogen bond with the nt21(O5′) oxygen, consistent with the complementary His29 general acid role. We suggest that this geometry achieved all the necessary features to promote the Ser148 general base His29 general acid mechanism with no requirement for any further rearrangements. The fact that we identified this structure only in one simulation does not weaken the result. The catalytically active geometry may differ from the dominantly populated ground state precursor geometry, i.e., structural rearrangements to rare but catalytically active conformations may precede the chemical reaction [36].

3.3.1. The remaining precursor $4al5_{Met}$ simulations

The behavior of the remaining $4al5_{Met}$ simulations was quite variable. The simulations revealed similar global conformational behavior as observed in 4al5 simulations reported above (see Figs. S37–S56 and Table S8 in Supporting information). Their hallmark was again difficulty to establish simultaneously short contacts (below 3.5 Å) between Ser148(O_{γ})–G20(O2′), Ser150(O_{γ})–G20(O3′) and Ser148(O_{γ})–Ser150(O_{γ}). Some simulations showed large scale rearrangements of the catalytic center (Fig. 9).

In all $4al5_{Met}(His^{+})$ simulations, the methoxy group independently of its initial location always flipped toward His29, which established a hydrogen bond interaction with the nt21(O5′) oxygen. Such arrangement could support the mechanism where protonated His29 acts as the general acid (Figs. 2A and S38). Note, however, that eventual rearrangement supporting the second catalytic scenario ($His29^{0}$ acting as the general base) could be difficult to reach by our simulations from the available starting structures because it could require concerted rearrangement of several residues. Indeed, no interaction between neutral His29 and the G20(O2′) was established in our simulations. The $His29^{0}$ in the 100 ns $4al5_{Met}(His^{0})$ simulation was modestly shifted away from the 3′-end phosphate (no direct interaction with the phosphate was established) but it did not result into any large distortion of the active site, in contrast to formation of the large cleft seen in comparable product $4al5(His^{0})$ simulations. In the 50 ns $4al5_{Met}(His^{0})$ simulation, His29 and Gln149 decisively departed from the 3′-end RNA.

In contrast to the product simulations we did not have an unambiguous reference experimental structure to assess quality of the precursor simulations. Due to the added methoxy group, we could not use the 4AL5 structure as the reference. The 2XLK precursor structures could not be used, because they were obtained with the inactivating G20 deoxyribose. The absence of the key native 2′-hydroxyl group of G20 affects the network of H-bonds between the β-hairpin loop and the RNA 3′-end. For instance, there exists an H-bond interaction of 2.7 Å between Ser148(O_{γ}) and Ser150(O_{γ}) in the first unit of the 2XLK structure, which is absent in the product 4AL5 X-ray structure where Ser148 interacts with G20(O2′). Additionally, the inactivated precursor crystal structure shows a water molecule involved in the hydrogen bond network, which is not seen in the product structure. In general, the β-hairpin loop and His29 sampled in the $4al5_{Met}$ simulations similar substates as in the 4al5 set of simulations, including large mobility of the 3′-end RNA and different substates of Gln149 (Fig. 10).

3.4. Simulations of the Csy4/pre-crRNA precursor form based on the 2XLK inactivated precursor X-ray structure ($2xlk_A$ and $2xlk_B$ simulation sets)

Our additional two starting models of the precursor structure were based on the 2XLK inactivated precursor X-ray structure (see Methods).

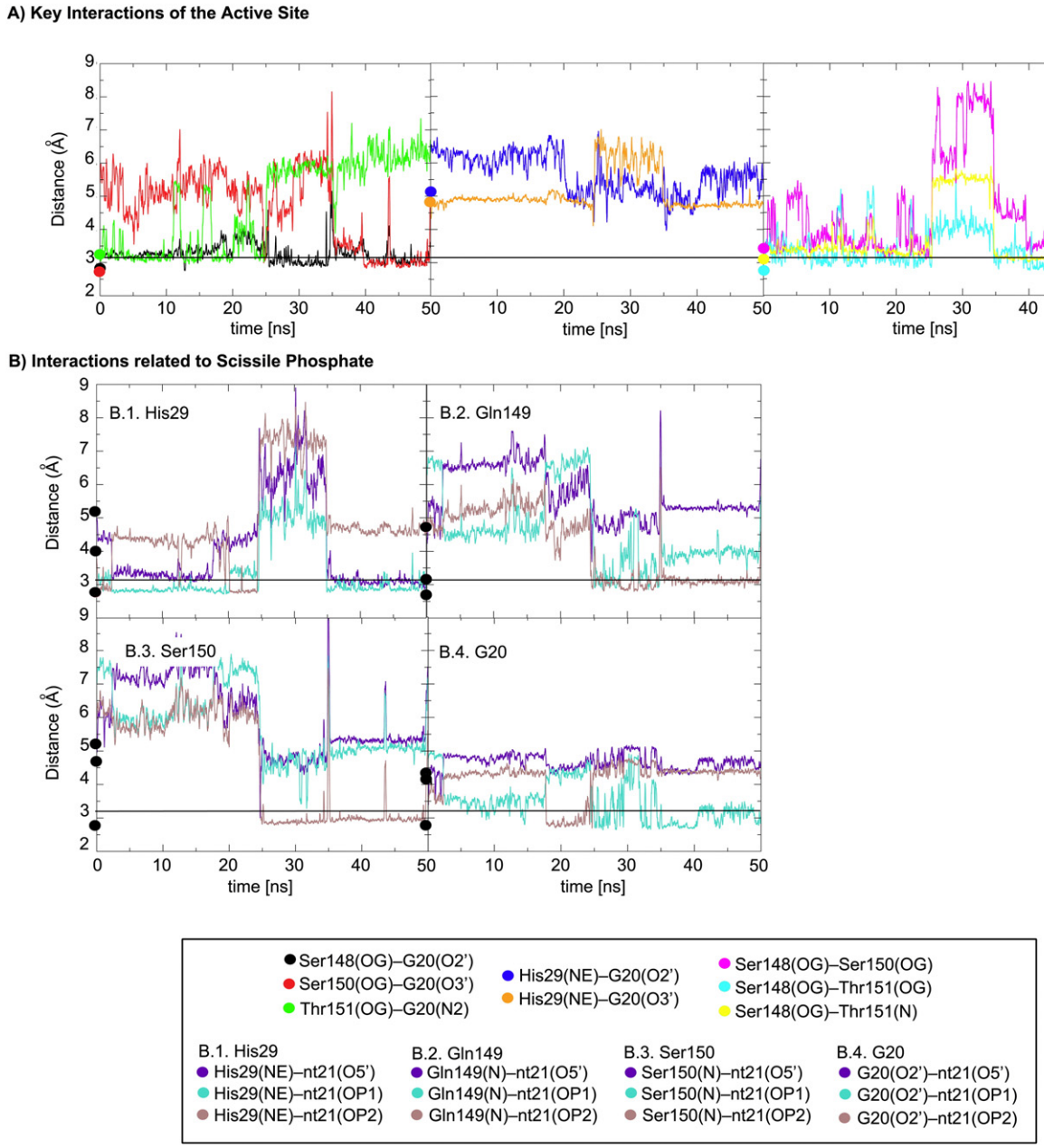


Fig. 11. Time evolution of the most significant distances in the 50 ns 4al5$_{Met2}$(His^{+}) precursor simulation. The experimental values marked by dots are taken from the product 4AL5 structure. The thin horizontal lines mark the 3.2 Å distance. (A) The key interactions of the active site. Left: interactions between the β-hairpin loop and the 3′-end of RNA. Middle: interactions between the His29 and the 3′-end of RNA. Right: distances between the triad residues Ser148, Ser150 and Thr151. (B) The interactions of the three oxygen atoms of the scissile phosphate with His29 (B.1), Gln149 (B.2), Ser150 (B.3) and G20 (B.4). Note that the three oxygen atoms of the phosphate group are chemically equivalent in the 4AL5 structure, so that their reference experimental distances cannot be distinguished and are marked by black dots. In the simulation, the methyl group has been added to the O5′ atom (see Methods for further details). The last 10 ns of this simulation sampled arrangements that are entirely consistent (at the level of classical molecular mechanics description) with the general base/general acid Ser148/His29 mechanism.

These simulations were rather consistent with the 4al5$_{Met}$ set of precursor simulations, though they more often resulted in larger structural perturbations. The double-protonated His29 again revealed a clear tendency to H-bond to the nt21(O5′). In contrast to the 4al5$_{Met}$ simulation set, we did not identify any simulation which would simultaneously stabilize the Ser148(O_γ)–G20(O2′), Ser150(O_γ)–G20(O3′) and Ser148(O_γ)–Ser150(O_γ) contacts. The 2xlk$_A$ and 2xlk$_B$ simulations are in more detail described in the Supporting information, Figs. S57–S65 and the accompanying text. Our 4al5$_{Met}$ precursor simulations based on the product X-ray structure were more insightful than the 2xlk$_A$/2xlk$_B$ precursor simulations based on the inactivated precursor X-ray structure.

4. Discussion

We employed explicit solvent MD simulations of Csy4/RNA endonuclease with the aim to investigate its structural dynamics, the arrangement of the active site and different protonation states. MD, when used with enough attention paid to its limitations, can complement the experimental data [30,54,62,63]. We have carried out 55 simulations with aggregate length of ~3.7 μs. We have extensively investigated the product 4AL5 X-ray structure to assess the dominant protonation states of the catalytic residues consistent with this experimental structure (see Figs. S3 and S5 in Supporting information). Further, we have used the 4AL5 structure, which has been simulated without any modification,

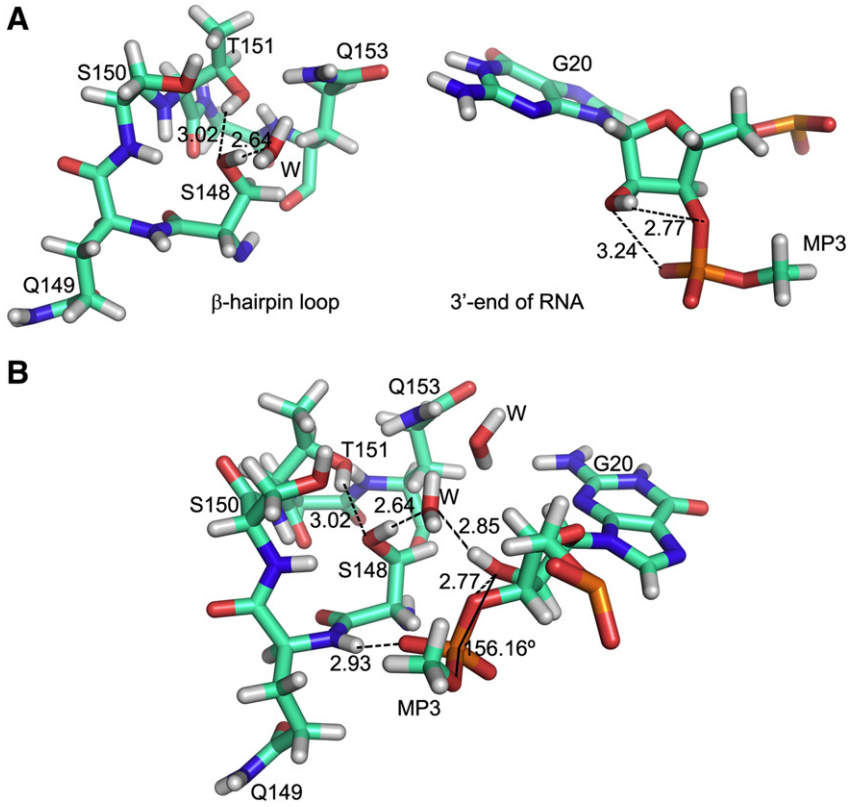


Fig. 12. (A) Front views of the β-hairpin loop (left) and the 3′-end of RNA (right), snapshot of a typical arrangements populated during the last 10 ns of the 4al5$_{Met2}$(His^{+}) 50 ns simulation. The most important residues and distances between the heavy atoms are indicated. (B) The same snapshot: hydrogen bond network between the β-hairpin loop and the 3′-end of RNA. The in-line attack O2′–P–O5′ angle is indicated. The Ser148 side chain is exposed to the solvent as H-bond donor and could thus be deprotonated. Then it could immediately deprotonate the G20(O2′) nucleophile because the Ser148(O_γ)–G20(O2′) distance fluctuates around 3.2 Å in this part of simulation.

to obtain insights into capability of simulations to properly describe this RNP complex. Then, we have carried out MD simulations of precursor structures built up using the available Csy4/RNA product and precursor X-ray structures. The catalytic residue has been modeled by the 3′-end methoxy-phosphate (MP3) rather than the complete C21 nucleotide (Fig. S5), because we were unable to model full C21 stably into our simulated systems.

4.1. Precursor simulation predicts a substate that could allow straightforward deprotonation of Ser148 to act as a general base

We have carried out twelve precursor simulations with double protonated His29. These simulations support the general acid role of His29 due to its tendency to bind to the scissile phosphate. Further, one of the simulations resulted into stable arrangement which appeared to have all necessary hallmarks of a catalytically potent geometry. This structure was in all aspects (at the level of molecular mechanical description) consistent with a catalytic scenario with double protonated His29 acting as a general acid and Ser148 acting (after deprotonation) as a general base (Fig. 12). It is important to point out that a catalytically active geometry may be represented by a rarely populated reactive conformation and may thus differ from the dominant ground state arrangement [33,36,64]. Therefore, observation of a suitable geometry in only one simulation does not weaken the result. Indeed, rarely populated but reactive enough protonation states of the active sites have been suggested to achieve general acid/base catalysis in RNA enzymes catalyzing the same RNA backbone self-cleavage reaction [65,66].

The simulation established stable simultaneous short contacts between Ser148(O_γ)–G20(O2′), Ser150(O_γ)–G20(O3′) and Ser148(O_γ)–Ser150(O_γ) (Fig. 11). The Ser148 hydroxyl group was reoriented away from the G20(O2′) nucleophile to donate H-bond to either the Thr151 side chain or to a water molecule. However, close contact between the Ser148 hydroxyl and the G20(O2′) nucleophile still persisted due to the additional H-bond interactions in the catalytic center (Fig. 12). The hydroxyl group of Thr151 was acting as H-bond donor to water molecules or Gln153. In this geometry, Ser148 could be straightforwardly deprotonated either directly by a water molecule or via proton shuttle involving Thr151 side chain deprotonated by solvent. Then, Ser148 could act as the general base and deprotonate the 2′-hydroxyl nucleophile. Further, the RNA backbone was extensively sampling near in-line attack configurations (see Fig. S56 in Supporting information) and the double-protonated His29 was H-bonded with the nt21 O5′ leaving group, so it could act as the general acid.

In our preceding MD simulations of glmS and hairpin ribozymes, we extensively simulated analogous scenarios when deprotonated catalytic guanine (i.e., in its rare protonation state) was assumed to act as a general base deprotonating the 2′-hydroxyl nucleophile [33,38]. However, we could not obtain trajectories that would suggest a mechanism of guanine deprotonation and its subsequent return to the catalytic center to establish the catalytic interaction with the nucleophile. In contrast, our potential Cys4/RNA catalytic geometry is in all aspects structurally consistent with the catalytic scenario having Ser148 acting as a general base and His29 as a general acid. The Ser148 residue can be deprotonated and then immediately or concomitantly deprotonate the nucleophile without any relocation.

4.2. Protonation state and the proposed functional role of His29 based on the precursor simulations

Experimental studies have identified two amino acid residues essential for the catalysis, Ser148 and His29 [10,19]. Two plausible reaction mechanisms were proposed (Fig. 2). Initially, double-protonated His29 (His^{+}) was suggested to act as a general acid, donating its proton to the leaving alcoholate nt21(O5′) [10]. In this mechanism, Ser148

would either play a structural role or activate the G20(2′-OH) by accepting its proton (it would have to be deprotonated before the reaction). The subsequent study was interpreted in favor of the second mechanism, with His29 δ-protonated (His^0) and acting as a general base accepting the proton from the G20(2′-OH) group [19]. The Ser148 would play a structural role and the leaving alcoholate nt21(O5′) could be protonated by solvent.

To our opinion, the available experimental techniques do not provide sufficient information to unambiguously distinguish which mechanism is present or to rule out co-existence of several competing microscopic mechanisms. The His29 general acid role was suggested, among other reasons, based on geometries of the active site of inactivated precursor and product low-pH crystal structures [10]. The subsequent suggestion of general base role of His29 was based on the pH-profile and His29Lys mutation assay [19]. The pH-rate profiles of both wild type and His29Lys mutant clearly supported the active participation of the His29 in proton transfer during the self-cleavage and linked it with the ascending part of the pH-rate profile. However, the reaction rate pH-profiles are notoriously ambiguous [67]. When the self-cleavage of the RNA sugar-phosphate backbone is catalyzed via general acid/base mechanism, it should lead to bell-shaped pH-rate profile. However, low pKa general base and high pKa general acid may lead to the same pH dependence of the reaction as low pKa general acid and high pKa general base. In other words, the ascending parts of bell-shaped pH-rate profiles (at more acidic pHs) do not necessarily correspond to a general base and in turn may also reflect pKa of a general acid, when the general base has higher pKa than the general acid [64,68,69]. In the case of Cys4 catalysis, however, the observed pH-rate profile has rather sigmoidal shape. Nevertheless, when the apparent pKa of a general base or general acid is outside the measurable pH range, the bell-shaped profile might appear as sigmoidal. This would almost surely be the case of the Ser148 participation. Thus, in this particular case, all biochemical data might be also consistent with general acid role of His29 combined with activation of the 2′-OH nucleophile by a general base with pKa above ~11 (the upper pH value of the observed His29Lys pH-rate profile). Notably, the Ser148 might indeed play such a role as the (unperturbed) pKa of serine hydroxyl group is ~13 [70]. In such mechanism, the reactive form of the active site would correspond to the rare protonation states of Ser148 and His29 and would be reactive enough to compensate the thermodynamic penalty due to the rare protonation states. Textbook example of such scenario is the HDV ribozyme, where the general acid (rare protonated state of C75 with pKa around 6) is assumed to be possibly complemented by Mg^{2+}-bound OH^- with pKa ~11.5. In fact, deprotonated serine participates in catalysis of serine proteases. However, we would like to underline that we in no case claim that our simulations prove such a mechanism for the Cys4 endonuclease. We merely suggest that this mechanism should not be dismissed based on the available data and that we found a very plausible atomistic arrangement for such a scenario. On the other hand, our simulations are also consistent with the more conservative role of Ser148 structurally supporting the C2′-endo conformation of the catalytic ribose and the in-line attack conformation, while the 2′-OH nucleophile could be deprotonated (activated) by solvent via specific base mechanism.

As our simulations confirm, the arrangement of the active site in both X-ray structures does not support the general base role of neutral His29. Obviously, we need to consider the low pH of ~5 of these structures, which may support the His29 protonation and cause subsequent local re-conformation of the catalytic center. However, we again re-emphasize that the chemical reaction may advance via rare but chemically reactive conformations and protonation states, as a result of the inherently rugged conformational surface of the enzyme. Minor conformational states cannot be easily captured by X-ray crystallography, as recently discussed for small ribozymes [36]. Then, for example, the low pH X-ray structures may in fact capture the catalytically relevant rare (at pH = 7) protonation states. Nevertheless, when assessing the present simulations, we certainly need to keep in mind that the simulation outcome is affected by the starting structures [30]. The general acid role of His29 is clearly consistent with our precursor simulations, which reveal that the terminal methoxy group prefers to be oriented toward His29, forming a clear hydrogen bond between the protonated His29(N_ε) and nt21(O5′) (Fig. S38). On the other hand, the simulations do not indicate any reactive conformations suitable for His^0 acting as a general base. However, this particular result should not be overinterpreted. Direct proton abstraction of G20(2′-OH) by means of $His29^0$ would require the $His29^0$ side-chain positioned near to the 2′-hydroxyl of G20 (Fig. 2B). However, in the available crystal structures, the G20(O2′) is too far from His29 to establish any H-bonds (Fig. S3E). Simulations explicitly considering neutral δ-protonated $His29^0$ reveal a further increase of the distance between His29 and G20(2′-OH). However, the simulation time scale may be insufficient to complete the rearrangements of the catalytic center for the His29 general base mechanism, preventing us, in the absence of the corresponding X-ray data, to observe a suitable catalytic geometry for the neutral His29. Similar story can be followed in experimental and subsequent simulation studies of the HDV ribozyme [32,34,35,71,72]. In general, for small self-cleaving ribozymes, it is not always easy to obtain an unambiguous agreement between the X-ray, MD simulation and biochemical data. Specifically, the dominant protonation states suggested by simulations (based on the available X-ray structures) may be different (i.e., opposite) than those suggested by the mechanistic data [33,38,64]. Textbook example is the G40 of glmS ribozyme, which could be deprotonated based on the biochemical data, but so far could not be stabilized as $G40^-$ in appropriate geometry in the simulations [33]. Similar complex relation between X-ray, biochemical and simulation data for potentially deprotonated guanine has been discussed also for the hairpin ribozyme [38]. The X-ray structures of ribozymes suggest that the potentially catalytic guanines can be readily bound to the 2′-hydroxyl nucleophile in their canonical state. However, before the reaction the guanine would have to be deprotonated. This is unlikely to take place when the guanine is H-bonded to the 2′-hydroxyl nucleophile as H-bond donor. The guanine would have to be deprotonated in a geometry in which its N1 is accessible to the solvent and then return back, which is not seen in simulations carried out so far for the ribozymes [33,38,64].

In summary, our present study identifies a structure where His29 can act as a general acid and Ser148 is optimally oriented for deprotonation through solvent to be able to act (besides its structural role) as a general base. We will further investigate this potential atomistic reaction mechanism via QM/MM computations [73]. However, for reasons explained above, we in no case exclude the His29 general base role and more efforts to locate suitable geometries for this catalytic scenario need to be done before either ultimately excluding it or finding it.

4.3. Simulations with doubly-protonated His29 and deprotonated doubly-charged phosphate best agree with the 4AL5 product structure

We identified the dominant protonation state of key residues corresponding to the 4AL5 structure based on evaluation of the agreement between the simulated and experimental 4AL5 X-ray structures. The 4al5(His^+, P^{2-}) system having double-protonated His29 residue and deprotonated double-charged scissile phosphate was the most consistent with the product crystal structure (Fig. 7). Simulations done with single-negatively charged (mono-protonated) scissile phosphate resulted in large perturbations at the active site (see Figs. 9, right and S24 and S25 in the Supporting information). The extensive H-bond network involving the oxygen atoms of the P21 phosphate seen in the 4AL5 structure was always lost. In turn, simulations performed with neutral His29 led to formation of a deep cleft between the N-terminal and C-terminal domains, which substantially distorted the active site by shifting the His29 residue away from the scissile phosphate and G20 nucleotide (Fig. 9, left).

4.4. Full agreement between the 4AL5 structure and the simulations could not be obtained

Even in the five independent 4al5(His^+, P^{2-}) simulations considering the dominant protonation state we observed a visible flexibility in the β-hairpin loop. In all five simulations, the structure initially very quickly (often within picoseconds) lost many observed interactions, with some interatomic distances changed by several Å (see Figs. 6 and 8). This is a rather unusual simulation behavior (not common, for example, in simulations of medium-sized RNAs). It indicates some energy imbalance in the starting structure, which could stem from the resolution limits of the experimental structure, its overall complexity and some force field inaccuracies. Obviously, the initial local instabilities in the simulations may be a consequence of removal of crystal packing contacts, since they occur in intrinsically flexible regions of the RNP complex (see description of the crystal contacts in the Supporting information). We would like to underline (see Introduction) that we in no case suggest that the experimental structure is incorrect. For large systems the starting structure may cause initial instabilities in simulations even when being basically correct from the experimental point of view, since the unprocessed experimental coordinate file is always high in the molecular mechanical potential energy [31]. It is possible that the standard equilibration protocols that are sufficient to prepare simulations of simple systems may be less robust for intricate molecular complexes such as the Csy4/RNA complex. Full relaxation for systems with dense networks of molecular interactions may be difficult, at least compared to small RNAs.

The role of the starting structure is supported by the fact that two 4al5(His^+, P^{2-}) simulations return remarkably close to the starting structure after ~20 ns. It indicates good performance of the force field (see Figs. 6 and 8 for Sim-1 and S14 and S15 in Supporting information for Sim-3). Capability of the system to return back after a large initial perturbation is convincing evidence that these simulations capture the dominant protonation state consistent with the 4AL5 structure. The fact that three other simulations are not repaired on the 50 ns timescale is not surprising, considering the magnitude of the initial perturbation (Figs. S14–S21 in Supporting information). Nevertheless, the final local H-bonding network established in the catalytic center even in the best behaving 4al5(His^+, P^{2-}) simulations differs from the experimental geometry. Thus, we achieve qualitative but not quantitative agreement with the experimental structure. This may indicate some limitation of the force field description.

4.5. Potential force field issues in description of the β-hairpin loop and its interactions with the 3′-end of RNA

Deviations between the simulated and X-ray structure were seen in the β-hairpin loop, specifically the Ser148, Ser150 and Thr151 triad (Fig. S3F) [10,19]. Ser150 residue typically flipped its χ_1 dihedral to a non-native value leading to further local rearrangements including Ser150 χ_2 (Fig. 8A, middle). It seemed that the unstable side-chain behavior of Ser150 residue destabilized the arrangement of the active site. This could at first sight indicate errors in the torsion potentials of the triad residues. However, based on some subsequent computations we suggested that the behavior was not caused by the parametrization of the dihedral terms (Fig. S65). We attempted to stabilize the simulations by restraining the χ_2 dihedrals of all three triad residues since it should indirectly stabilize (via supporting the native H-bonds) also the Ser150 χ_1 angle. These restraints did not solve the problem. The data indicated that (i) when we restrained the χ_2 dihedral angles, the G20 ribose was unable to maintain the initial (native) C2′-endo conformation and (ii) the optimal combination of the χ_2 dihedral angles of the triad located in the β-hairpin loop caused imbalance of the 3′-end of the RNA (Figs. S26–S31). This suggested an overall imbalance in the target region which cannot be resolved just by dihedral parameter tuning (Fig. S5 in Supporting information). We also tried the *ff12SB* protein force field with refined side chain dihedrals. However, it also did not improve the simulations (Figs. S34–S36 in Supporting information). In other words, it does not seem that tuning the side chain dihedrals of the β-hairpin loop triad can decisively stabilize the simulations (see also Supporting information Fig. S65 for a comparison of MM and QM serine χ_1 potential energy profiles).

Besides the β-hairpin loop triad, an important player was the pucker conformation of G20 ribose. The electron density of the 4AL5 crystal structure was inconsistent with a C3′-endo conformation but agreed well with a C2′-endo form [19]. The hydrogen bond interactions between the 2′-hydroxyl group of G20 and the side chains of Ser148 and Tyr176 may enforce the observed C2′-endo sugar pucker of G20 (Fig. S3D) [19]. The C2′-endo conformation may support the in-line attack conformation requiring a locally extended RNA backbone (Fig. S3A) [19]. Sugar-pucker is a difficult parameter in nucleic acids force fields which together with the glycosidic χ torsion has dramatic impact on simulations [21,40,48]. During the simulations, we usually first observed rapid re-puckering of the G20 ribose from C2′-endo to C3′-endo, which then led to swift major perturbation of the active site arrangement. Fortunately, in many simulations the sugar pucker returned spontaneously back to the native C2′-endo conformation after the initial relaxations. Thus, we also experimented with restraining the C2′-endo conformation but it also did not lead to full agreement with the 4AL5 structure (see the Results section for more details and Figs. S29 and S30 in Supporting information).

It is to be noted that our inability to decisively stabilize the simulations via dihedral angle restraints is not surprising. As discussed elsewhere [54,74], the intramolecular torsional terms are often used for the final fine-tuning of the force field performance. They effectively compensate for inaccuracies and deficiencies of the other force field terms, including the intermolecular terms [75]. Obviously, capability of torsional potentials to eliminate all force field imbalances and transferability of the optimized torsional potentials are not unlimited [31,76]. So, we suggest that rather than primarily originating in the parameterization of the torsional terms, our problems to describe the intricate details of the catalytic center (Fig. 5) may be due to combination of imperfectness of many different force field terms. The terminal highly polarizable doubly-charged anionic phosphate group is inherently difficult to describe by non-polarizable pair-additive force fields with fixed atomic charges [74]. This entity in addition interacts with the positively charged His29, creating a complex network of interactions (Fig. 5). The balance of forces is further complicated by extensive interactions between the terminal P^{2-} group and monovalent ions (not shown) for which the non-polarizable force field may also introduce imperfectness. Thus, this region is almost surely affected by substantial geometry-dependent electronic structure redistribution effects that are not fully describable by the simple fixed-parameters force field form, irrespective of its parameters [77,78]. This is the likely reason why our data indicate an overall difficulty to balance all interactions simultaneously, including the struggle between the Ser150 orientation and the G20 sugar pucker. Considering the electronic structure complexity of the interactions in the core of the 4AL5 structure the performance of the force field is actually quite good and its eventual further improvement may be tricky. Polarizable force fields may in future provide insights into the nature and magnitude of polarization effects in the catalytic core of the Csy4/RNA RNP complex [79,80]. Obviously, we still cannot rule out that the initial perturbations are also due to properties of the starting structure (see above) and we as well cannot rule out that our simulations were too short to allow full relaxation of the catalytic center. In addition, we should keep in mind that we compare MD simulations with one static medium resolution X-ray structure, which could be affected by averaging over several local substates and by crystal packing.

Irrespective of the origin of the above-noted simulation behavior, our data indicate that accurate simulations of complex RNP systems may be more difficult than simulations of small RNA and protein

molecules alone. The literature experience with simulations of protein/RNA interactions is so quite fragmented [81–98] with essentially no discussion of the limits of the simulation methodology. Some of the studies are based on simulations apparently progressively deviating from the starting structures. As we noted in the Introduction, standard unbiased simulations using a realistic force field and initiated from an appropriately prepared starting structure should, considering the simulation time scale, remain close to the starting structures. Although the increased dynamics may sometimes be praised as desirable to obtain biochemically interesting results, one has to be very careful with increased dynamics that is caused by the force fields. Increased flexibility of molecules can be achieved for example by enhanced sampling methods or external denaturating conditions such as elevated temperature, and absence of salt [31]. Although many of these approaches do not have an equivalent experimental method and are prone to certain over-interpretations, they are in principle entirely valid [31]. They bias the dynamics by applying external sampling-enhancing factors to presumably correct models of the studied molecule. In contrast, when the sampling is "enhanced" due to force field imbalances, then the correctness of the core description of the studied molecule itself is compromised, which may severely invalidate all the results. Therefore, we suggest that performance of contemporary force fields in simulations of different types of protein–RNA complexes should be systematically tested in future, since it may differ from simulation behavior of isolated RNA and protein molecules. Equally important will be to understand the role of starting structures and their equilibration on stability of simulations of protein–RNA complexes.

5. Conclusions

We have carried out series of explicit solvent MD simulations of the Csy4/RNA RNP complex, which participates in adaptive immune system of prokaryote cells. We used the available X-ray structures of product and inactivated precursor as the starting geometries. The simulations were executed with the bsc0χ_{OL3} (RNA) and ff99SB (protein) variants of the AMBER Cornell et al. force field. With ~3.7 μs of total simulation data, our study is one of the most extended MD studies of RNP systems currently available in the literature.

Our precursor simulations reveal that double protonated His29 could easily bind to the O5′ leaving oxygen of the scissile phosphate to act as a general acid. Further, we report sufficiently stable arrangement consistent with catalytic scenario utilizing double-protonated His29 as a general acid and deprotonated Ser148 as a general base. The Ser148 side chain is in a close contact with the nucleophile 2′-hydroxyl group but donates H-bond either to Thr151 or to a water molecule. The protonated His29 is bound to the O5′ leaving group and the backbone samples in-line attack geometries. In this geometry, the Ser148 side chain can be deprotonated either by a water molecule or perhaps by the proton shuttle via the Thr151 side chain. The deprotonated Ser148$^-$ would then be capable to immediately act as the general base with essentially no further structural rearrangement. Obviously, MD simulations alone cannot prove catalytic mechanisms, but we suggest that if the Ser148/His29 general base/general acid catalysis occurs, it could proceed via the structure shown in our work. Although this catalytic scenario would involve rare protonation forms of Ser148 and His29, it could be reactive enough to compensate the thermodynamic penalty coming from the rare protonation state, as common for small ribozymes.

Our simulations do not reveal any tendency of the δ-protonated neutral His29 to bind to the nucleophile. Rather, monoprotonated His29 facilitates disruption of the catalytic pocket. However, as explained in the Discussion section, this particular result should not be over-interpreted since it may be affected by the available experimental structures used as the starting geometries in connection with the limited time scale of the simulations. Nevertheless, the simulations show that if His29 acts as a general base, the active site arrangement would have to significantly differ from the available X-ray structures and would be separated from them by significant rearrangements not observable on the simulation time scale. In summary, although we undoubtedly derive a suitable atomistic geometry for the general acid catalysis by His29, we in no case rule out its general base role.

The simulations suggest that double-protonated His29 and deprotonated doubly-charged phosphate are the dominant protonation states corresponding to the product X-ray structure.

Our simulations give also insights into the performance of the simulation technique. At the beginning of all simulations, we observe rapid losses of multiple interactions in the catalytic center at the protein/RNA interface, which indicates stress in the initial structures. Such behavior is uncommon in simulations of RNA molecules based on X-ray starting structures and suggests that simulations of RNP complexes may be more complicated. However, when simulating the product structure with the right protonation states, some simulations are capable to remarkably restore majority of the native interactions at the catalytic center after dozens of ns. We suggest that the initial period of the production simulations is needed for a proper relaxation of the starting structure which is not completed by the standard equilibration protocol. However, the simulations arrive at alternative rather than the native H-bonding pattern, achieving qualitative but not quantitative agreement with the experimental structure of the catalytic center. Hallmark of this behavior is a struggle between the correct geometry of the β-hairpin loop triad (Ser148, Ser150 and Thr151) and the RNA 3′-end, especially the G20 ribose. The simulations tend to flip the Ser150 χ_1 angle to a non-native orientation and it is difficult to obtain stable simultaneous short contacts between Ser148(O_γ)–G20(O2′), Ser150(O_γ)–G20(O3′) and Ser148(O_γ)–Ser150(O_γ) atom pairs observed in the X-ray structure. It is also difficult to simultaneously stabilize the correct β-hairpin loop triad geometry and the native C2′-endo pucker of G20. We suggest that the simulation results may reflect some limitations of the force field potential energy description and may be related to non-polarizable nature of the force field.

Our work confirms that MD simulation technique [22,23,30,54,63,81,99,100] is a powerful tool to complement experimental data on RNA and RNP enzymes. However, MD simulations of intricate RNP complexes may be much more challenging than apparent from the presently available literature.

Acknowledgement

This work was funded by the grant P208/12/1878, Grant Agency of the Czech Republic. This work was further supported by the project "CEITEC — Central European Institute of Technology" CZ.1.05/1.1.00/02.0068 from European Regional Development Fund (J.S.), the Operational Program Research and Development for Innovations — European Regional Development Fund (project CZ.1.05/2.1.00/03.0058), and the Operational Program Education for Competitiveness — European Social Fund (CZ.1.07/2.3.00/20.0017) of the Ministry of Education, Youth and Sports of the Czech Republic, by the EU Seventh Framework Program under the "Capacities" specific program (contract no. 286154) and by project Employment of Best Young Scientists for International Cooperation Empowerment, reg. number CZ.1.07/2.3.00/30.0037 co-financed by the European Social Fund and the state budget of the Czech Republic. We would like to thank Martin Jinek for many stimulating discussions.

Appendix A. Supplementary data

The Supporting information file contains Supporting Methods and Results, Figs. S1–S65, and Tables S1–S8. In Supporting Methods are also listed the AMBER prep-files of the 3 nonstandard residues. Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2014.10.021.

References

[1] A. Bolotin, B. Ouinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology 151 (2005) 2551–2561, http://dx.doi.org/10.1099/mic.0.28048-0-10097.

[2] F.J.M. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, J. Mol. Evol. 60 (2005) 174–182, http://dx.doi.org/10.1007/s00239-004-0046-3.

[3] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiology 151 (2005) 653–663, http://dx.doi.org/10.1099/mic.0.27437-0.

[4] R. Sorek, V. Kunin, P. Hugenholtz, CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea, Nat. Rev. Microbiol. 6 (2008) 181–186, http://dx.doi.org/10.1038/nrmicro1793.

[5] S.J.J. Brouns, M.M. Jore, M. Lundgren, E.R. Westra, R.J.H. Slijkhuis, A.P.L. Snijders, M.J. Dickman, K.S. Makarova, E.V. Koonin, J. van der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes, Science 321 (2008) 960–964, http://dx.doi.org/10.1126/science.1159689.

[6] B. Wiedenheft, E. van Duijn, J. Bultema, S. Waghmare, K.H. Zhou, A. Barendregt, W. Westphal, A. Heck, E. Boekema, M. Dickman, J.A. Doudna, RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10092–10097, http://dx.doi.org/10.1073/pnas.1102716108.

[7] K.S. Makarova, D.H. Haft, R. Barrangou, S.J.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J.M. Mojica, Y.I. Wolf, A.F. Yakunin, J. van der Oost, E.V. Koonin, Evolution and classification of the CRISPR-Cas systems, Nat. Rev. Microbiol. 9 (2011) 467–477, http://dx.doi.org/10.1038/nrmicro2577.

[8] J. Carte, R.Y. Wang, H. Li, R.M. Terns, M.P. Terns, Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes, Genes Dev. 22 (2008) 3489–3496, http://dx.doi.org/10.1101/gad.1742908.

[9] J. Carte, N.T. Pfister, M.M. Compton, R.M. Terns, M.P. Terns, Binding and cleavage of CRISPR RNA by Cas6, RNA 16 (2010) 2181–2188, http://dx.doi.org/10.1261/rna.2230110.

[10] R.E. Haurwitz, M. Jinek, B. Wiedenheft, K.H. Zhou, J.A. Doudna, Sequence- and structure-specific RNA processing by a CRISPR endonuclease, Science 329 (2010) 1355–1358, http://dx.doi.org/10.1126/science.1192272.

[11] E.M. Gesner, M.J. Schellenberg, E.L. Garside, M.M. George, A.M. MacMillan, Recognition and maturation of effector RNAs in a CRISPR interference pathway, Nat. Struct. Mol. Biol. 18 (2011) 688–692, http://dx.doi.org/10.1038/nsmb.2042.

[12] N.G. Lintner, M. Kerou, S.K. Brumfield, S. Graham, H.T. Liu, J.H. Naismith, M. Sdano, N. Peng, Q.X. She, V. Copie, M.J. Young, M.F. White, C.M. Lawrence, Structural and functional characterization of an archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE), J. Biol. Chem. 286 (2011) 21643–21656, http://dx.doi.org/10.1074/jbc.M111.238485.

[13] D.G. Sashital, M. Jinek, J.A. Doudna, An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3, Nat. Struct. Mol. Biol. 18 (2011) 680–687, http://dx.doi.org/10.1038/nsmb.2043.

[14] S.H. Sternberg, R.E. Haurwitz, J.A. Doudna, Mechanism of substrate selection by a highly specific CRISPR endoribonuclease, RNA 18 (2012) 661–672, http://dx.doi.org/10.1261/rna.030882.111.

[15] E. Deltcheva, K. Chylinski, C.M. Sharma, K. Gonzales, Y.J. Chao, Z.A. Pirzada, M.R. Eckert, J. Vogel, E. Charpentier, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III, Nature 471 (2011) 602–607, http://dx.doi.org/10.1038/nature09886.

[16] L.A. Marraffini, E.J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA, Science 322 (2008) 1843–1845, http://dx.doi.org/10.1126/science.1165771.

[17] M.P. Terns, R.M. Terns, CRISPR-based adaptive immune systems, Curr. Opin. Microbiol. 14 (2011) 321–327, http://dx.doi.org/10.1016/j.mib.2011.03.005.

[18] R. Sapranauskas, G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, V. Siksnys, The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*, Nucleic Acids Res. 39 (2011) 9275–9282, http://dx.doi.org/10.1093/nar/gkr606.

[19] R.E. Haurwitz, S.H. Sternberg, J.A. Doudna, Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA, EMBO J. 31 (2012) 2824–2832, http://dx.doi.org/10.1038/emboj.2012.107.

[20] V. Kunin, R. Sorek, P. Hugenholtz, Evolutionary conservation of sequence and secondary structures in CRISPR repeats, Genome Biol. 8 (2007) R61, http://dx.doi.org/10.1186/gb-2007-8-4-r61.

[21] C. Hsiao, S. Mohan, E. Hershkovitz, A. Tannenbaum, L.D. Williams, Single nucleotide RNA choreography, Nucleic Acids Res. 34 (2006) 1481–1491, http://dx.doi.org/10.1093/nar/gkj500.

[22] P. Banas, P. Jurecka, N.G. Walter, J. Sponer, M. Otyepka, Theoretical studies of RNA catalysis: hybrid QM/MM methods and their comparison with MD and QM, Methods 49 (2009) 202–216, http://dx.doi.org/10.1016/j.ymeth.2009.04.007.

[23] P. Carloni, U. Rothlisberger, M. Parrinello, The role and perspective of ab initio molecular dynamics in the study of biological systems, Acc. Chem. Res. 35 (2002) 455–464, http://dx.doi.org/10.1021/ar010018u.

[24] J. Garrec, C. Patel, U. Rothlisberger, E. Dumont, Insights into intrastrand cross-link lesions of DNA from QM/MM molecular dynamics simulations, J. Am. Chem. Soc. 134 (2012) 2111–2119, http://dx.doi.org/10.1021/ja2084042.

[25] J. Sgrignani, A. Magistrato, The structural role of $Mg2+$ ions in a class I RNA polymerase ribozyme: a molecular simulation study, J. Phys. Chem. B 116 (2012) 2259–2268, http://dx.doi.org/10.1021/jp206475d.

[26] A. Warshel, Computer simulations of enzyme catalysis: methods, progress, and insights, Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 425–443, http://dx.doi.org/10.1146/annurev.biophys.32.110601.141807.

[27] J. Aqvist, C. Lind, J. Sund, G. Wallin, Bridging the gap between ribosome structure and biochemistry by mechanistic computations, Curr. Opin. Struct. Biol. 22 (2012) 815–823, http://dx.doi.org/10.1016/j.sbi.2012.07.008.

[28] K. Nam, J. Gao, D.M. York, Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis, J. Am. Chem. Soc. 130 (2008) 4680–4691, http://dx.doi.org/10.1021/ja0759141.

[29] A. Ganguly, P.C. Bevilacqua, S. Hammes-Schiffer, Quantum mechanical/molecular mechanical study of the HDV ribozyme: impact of the catalytic metal ion on the mechanism, J. Phys. Chem. Lett. 2 (2011) 2906–2911, http://dx.doi.org/10.1021/jz2013215.

[30] M.A. Ditzler, M. Otyepka, J. Sponer, N.G. Walter, Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in, Acc. Chem. Res. 43 (2010) 40–47, http://dx.doi.org/10.1021/ar900093g.

[31] J. Šponer, P. Banáš, P. Jurečka, M. Zgarbová, P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer, M. Otyepka, Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome, J. Phys. Chem. Lett. 5 (2014) 1771–1782, http://dx.doi.org/10.1021/jz500557y.

[32] M.V. Krasovska, J. Sefcikova, K. Reblova, B. Schneider, N.G. Walter, J. Sponer, Cations and hydration in catalytic RNA: molecular dynamics of the hepatitis delta virus ribozyme, Biophys. J. 91 (2006) 626–638, http://dx.doi.org/10.1529/biophysj.105.079368.

[33] P. Banas, N.G. Walter, J. Sponer, M. Otyepka, Protonation states of the key active site residues and structural dynamics of the glmS riboswitch as revealed by molecular dynamics, J. Phys. Chem. B 114 (2010) 8701–8712, http://dx.doi.org/10.1021/jp9109699.

[34] T.S. Lee, G.M. Giambasu, M.E. Harris, D.M. York, Characterization of the structure and dynamics of the HDV ribozyme at different stages along the reaction path, J. Phys. Chem. Lett. 2 (2011) 2538–2543, http://dx.doi.org/10.1021/jz201106y.

[35] N. Veeraraghavan, A. Ganguly, J.H. Chen, P.C. Bevilacqua, S. Hammes-Schiffer, B.L. Golden, Metal binding motif in the active site of the HDV ribozyme binds divalent and monovalent ions, Biochemistry 50 (2011) 2672–2682, http://dx.doi.org/10.1021/bi2000164.

[36] K.N. Sripathi, W.W. Tay, P. Banáš, M. Otyepka, J. Šponer, N.G. Walter, Disparate HDV ribozyme crystal structures represent intermediates on a rugged free-energy landscape, RNA 20 (2014) 1112–1128, http://dx.doi.org/10.1261/rna.044982.114.

[37] Y. Xin, D. Hamelberg, Deciphering the role of glucosamine-6-phosphate in the riboswitch action of glmS ribozyme, RNA 16 (2010) 2455–2463, http://dx.doi.org/10.1261/rna.2334110.

[38] V. Mlynsky, P. Banas, D. Hollas, K. Reblova, N.G. Walter, J. Sponer, M. Otyepka, Extensive molecular dynamics simulations show that canonical G8 and protonated A38H+ forms are most consistent with crystal structures of hairpin ribozyme, J. Phys. Chem. B 114 (2010) 6642–6652, http://dx.doi.org/10.1021/jp1001258.

[39] 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, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197, http://dx.doi.org/10.1021/ja00124a002.

[40] J.M. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074, http://dx.doi.org/10.1002/1096-987X(200009)21.

[41] A. Perez, I. Marchan, D. Svozil, J. Sponer, T.E. Cheatham III, C.A. Laughton, M. Orozco, Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers, Biophys. J. 92 (2007) 3817–3829, http://dx.doi.org/10.1529/biophysj.106.097782.

[42] M. Zgarbova, M. Otyepka, J. Sponer, A. Mladek, P. Banas, T.E. Cheatham III, P. Jurecka, Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles, J. Chem. Theory Comput. 7 (2011) 2886–2902, http://dx.doi.org/10.1021/ct200162x.

[43] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins Struct. Funct. Bioinf. 65 (2006) 712–725, http://dx.doi.org/10.1002/prot.21123.

[44] Y. Hashem, P. Auffinger, A short guide for molecular dynamics simulations of RNA systems, Methods 47 (2009) 187–197, http://dx.doi.org/10.1016/j.ymeth.2008.09.020.

[45] C.X. Weichenberger, M.J. Sippl, NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures, Nucleic Acids Res. 35 (2007) W403–W406, http://dx.doi.org/10.1093/nar/gkm263.

[46] M. Rostkowski, M.H.M. Olsson, C.R. Søndergaard, J.H. Jensen, Graphical analysis of pH-dependent properties of proteins predicted using PROPKA, BMC Struct. Biol. 11 (2011) 6, http://dx.doi.org/10.1186/1472-6807-11-6.

[47] P. Florova, P. Sklenovsky, P. Banas, M. Otyepka, Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact, J. Chem. Theory Comput. 6 (2010) 3569–3579, http://dx.doi.org/10.1021/ct1003687.

[48] P. Banas, D. Hollas, M. Zgarbova, P. Jurecka, M. Orozco, T.E. Cheatham, J. Sponer, M. Otyepka, Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins, J. Chem. Theory Comput. 6 (2010) 3836–3849, http://dx.doi.org/10.1021/ct100481h.

[49] D.A. Pearlman, P.A. Kollman, P. Connelly, Abstr. Pap. Am. Chem. Soc. 209 (1995) 91.

[50] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271, http://dx.doi.org/10.1021/j100308a038.

[51] B. Guillot, A reappraisal of what we have learnt during three decades of computer simulations on water, J. Mol. Liq. 101 (2002) 219–260, http://dx.doi.org/10.1016/S0167-7322(02)00094-6.
[52] I.S. Joung, T.E. Cheatham III, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, J. Phys. Chem. B 112 (2008) 9020–9041, http://dx.doi.org/10.1021/jp8001614.
[53] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman, AMBER 12, University of California, San Francisco, 2012.
[54] J. Sponer, X.H. Cang, T.E. Cheatham III, Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures, Methods 57 (2012) 25–39, http://dx.doi.org/10.1016/j.ymeth.2012.04.005.
[55] W.D. Cornell, P. Cieplak, C.I. Bayly, P.A. Kollman, Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation, J. Am. Chem. Soc. 115 (1993) 9620–9631, http://dx.doi.org/10.1021/ja00074a030.
[56] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684, http://dx.doi.org/10.1063/1.448118.
[57] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven Jr., J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision A.1, Gaussian, Inc., Wallingford CT, 2009.
[58] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. Model. 14 (1996) 33–38.
[59] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific LLC, Palo Alto, CA, 2008.
[60] P. Kuhrova, P. Banas, R.B. Best, J. Sponer, M. Otyepka, Computer folding of rna tetraloops? Are we there yet? J. Chem. Theory Comput. 9 (2013) 2115–2125, http://dx.doi.org/10.1021/ct301086z.
[61] T.E. Cheatham III, D.A. Case, Twenty-five years of nucleic acid simulations, Biopolymers 99 (2013) 969–977, http://dx.doi.org/10.1002/bip.22331.
[62] J. Feng, N.G. Walter, C.L. Brooks, Cooperative and directional folding of the preQ1 riboswitch aptamer domain, J. Am. Chem. Soc. 133 (2011) 4196–4199, http://dx.doi.org/10.1021/ja110411m.
[63] T. Schlick, R. Collepardo-Guevara, L.A. Halvorsen, S. Jung, X.Q. Xiao, Biomolecular-modeling and simulation: a field coming of age, Rev. Biophys. 44 (2011) 191–228, http://dx.doi.org/10.1017/S0033583510000284.
[64] J.L. Wilcox, A.K. Ahluwalia, P.C. Bevilacqua, Charged nucleobases and their potential for RNA catalysis, Acc. Chem. Res. 44 (2011) 1270–1279, http://dx.doi.org/10.1021/ar2000452.
[65] D.M.J. Lilley, Structure, folding and mechanisms of ribozymes, Curr. Opin. Struct. Biol. 15 (2005) 313–323, http://dx.doi.org/10.1016/j.sbi.2005.05.002.
[66] D.M.J. Lilley, Catalysis by the nucleolytic ribozymes, Biochem. Soc. Trans. 39 (2011) 641–646, http://dx.doi.org/10.1042/BST0390641.
[67] P.C. Bevilacqua, T.S. Brown, S. Nakano, R. Yajima, Catalytic roles for proton transfer and protonation in ribozymes, Biopolymers 73 (2004) 90–109, http://dx.doi.org/10.1002/bip.10519.
[68] T.J. Wilson, D.M.J. Lilley, Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid–base catalysis? A critical assessment of available experimental data, RNA 17 (2011) 213–221, http://dx.doi.org/10.1261/rna.2473711.
[69] S. Kath-Schorr, T.J. Wilson, N.S. Li, J. Lu, J.A. Piccirilli, D.M.J. Lilley, General acid–base catalysis mediated by nucleobases in the hairpin ribozyme, J. Am. Chem. Soc. 134 (2012) 16717–16724, http://dx.doi.org/10.1021/ja3067429.
[70] A. Kessel, N. Ben-Tal, Introduction to Proteins: Structure, Function and Motion, Chapman & Hall, CRC Press, 2010.
[71] A.L. Ke, K.H. Zhou, F. Ding, J.H.D. Cate, J.A. Doudna, A conformational switch controls hepatitis delta virus ribozyme catalysis, Nature 429 (2004) 201–205, http://dx.doi.org/10.1038/nature02522.
[72] J.H. Chen, R. Yajima, D.M. Chadalavada, E. Chase, P.C. Bevilacqua, B.L. Golden, A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage, Biochemistry 49 (2010) 6508–6518, http://dx.doi.org/10.1021/bi100670p.
[73] V. Mlynsky, P. Banas, J. Sponer, M.W. van der Kamp, A.J. Mulholland, M. Otyepka, Comparison of ab initio, DFT, and semiempirical QM/MM approaches for description of catalytic mechanism of hairpin ribozyme, J. Chem. Theory Comput. 10 (2014) 1608–1622, http://dx.doi.org/10.1021/ct401015e.
[74] J. Sponer, A. Mladek, J.E. Sponer, D. Svozil, M. Zgarbova, P. Banas, P. Jurecka, M. Otyepka, The DNA and RNA sugar-phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies, Phys. Chem. Chem. Phys. 14 (2012) 15257–15277, http://dx.doi.org/10.1039/c2cp41987d.
[75] M. Zgarbova, F.J. Luque, J. Sponer, M. Otyepka, P. Jurecka, A novel approach for deriving force field torsion angle parameters accounting for conformation-dependent solvation effects, J. Chem. Theory Comput. 8 (2012) 3232–3242, http://dx.doi.org/10.1021/ct3001987.
[76] M. Krepl, M. Zgarbova, P. Stadlbauer, M. Otyepka, P. Banas, J. Koca, T.E. Cheatham III, P. Jurecka, J. Sponer, Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA, J. Chem. Theory Comput. 8 (2012) 2506–2520, http://dx.doi.org/10.1021/ct300275s.
[77] A. Mladek, M. Krepl, D. Svozil, P. Cech, M. Otyepka, P. Banas, M. Zgarbova, P. Jurecka, J. Sponer, Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar-phosphate backbone and their comparison with modern density functional theory, Phys. Chem. Chem. Phys. 15 (2013) 7295–7310, http://dx.doi.org/10.1039/c3cp44383c.
[78] K. Gkionis, H. Kruse, J.A. Platts, A. Mládek, J. Koča, J. Šponer, Ion binding to quadruplex DNA stems. Comparison of MM and QM descriptions reveals sizable polarization effects not included in contemporary simulations, J. Chem. Theory Comput. 10 (2014) 1326–1340, http://dx.doi.org/10.1021/ct4009969.
[79] A. Savelyev, A.D. MacKerell Jr., All-atom polarizable force field for DNA based on the classical drude oscillator model, J. Comput. Chem. 35 (2014) 1219–1239, http://dx.doi.org/10.1002/jcc.23611.
[80] P.E.M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, A.D. MacKerell Jr., Polarizable force field for peptides and proteins based on the classical drude oscillator, J. Chem. Theory Comput. 9 (2013) 5430–5449, http://dx.doi.org/10.1021/ct400781b.
[81] A.D. MacKerell Jr., L. Nilsson, Molecular dynamics simulations of nucleic acid–protein complexes, Curr. Opin. Struct. Biol. 18 (2008) 194–199, http://dx.doi.org/10.1016/j.sbi. 2007.12.012.
[82] Y. Zhao, B.L. Kormos, D.L. Beveridge, A.M. Baranger, Molecular dynamics simulation studies of a protein–RNA complex with a selectively modified binding interface, Biopolymers 81 (2006) 256–269, http://dx.doi.org/10.1002/bip.20408.
[83] D.M. Blakaj, K.J. McConnell, D.L. Beveridge, A.M. Baranger, Molecular dynamics and thermodynamics of protein–RNA interactions: mutation of a conserved aromatic residue modifies stacking interactions and structural adaptation in the U1A-stem loop 2 RNA complex, J. Am. Chem. Soc. 123 (2001) 2548–2551, http://dx.doi.org/10.1021/ja005538j.
[84] R.P. Bahadur, S. Kannan, M. Zacharias, Thermodynamic calculations for biochemical transport and reaction processes in metabolic networks, Biophys. J. 97 (2009) 3139–3144, http://dx.doi.org/10.1016/j.bpj.2010.09.043.
[85] P.M. Petrone, C.D. Snow, D. Lucent, V.S. Pande, Side-chain recognition and gating in the ribosome exit tunnel, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 16549–16554, http://dx.doi.org/10.1073/pnas.0801795105.
[86] B.L. Kormos, Y. Benitex, A.M. Baranger, D.L. Beveridge, Affinity and specificity of protein U1A–RNA complex formation based on an additive component free energy model, J. Mol. Biol. 371 (2007) 1405–1419, http://dx.doi.org/10.1016/j.jmb.2007.06.003.
[87] N. Spackova, K. Réblová, J. Sponer, Structural dynamics of the box C/D RNA kink-turn and its complex with proteins: the role of the A-minor 0 interaction, long-residency water bridges, and structural ion-binding sites revealed by molecular simulations, J. Phys. Chem. B 114 (2010) 10581–10593, http://dx.doi.org/10.1021/jp102572k.
[88] K. Reblova, N. Spackova, J. Koca, N.B. Leontis, J. Sponer, Long-residency hydration, cation binding, and dynamics of loop E/Helix IV rRNA-L25 protein complex, Biophys. J. 87 (2004) 3397–3412, http://dx.doi.org/10.1529/biophysj.104.047126.
[89] L.A. Michael, J.A. Chenault, B.R. Miller III, A.M. Knolhoff, M.C. Nagan, Water, shape recognition, salt bridges, and cation–Pi interactions differentiate peptide recognition of the HIV Rev-responsive element, J. Mol. Biol. 392 (2009) 774–786, http://dx.doi.org/10.1016/j.jmb.2009.07.047.
[90] V. Cojocaru, S. Nottrott, R. Klement, T.M. Jovin, The snRNP 15.5 K protein folds its cognate K-turn RNA: a combined theoretical and biochemical study, RNA 11 (2005) 197–209, http://dx.doi.org/10.1261/rna.7149605.
[91] S. Fulle, H.J. Gohlke, Molecular recognition of RNA: challenges for modelling interactions and plasticity, Mol. Recognit. 23 (2010) 220–231, http://dx.doi.org/10.1002/jmr.1000.
[92] T.N. Do, P. Carloni, G. Varani, G. Bussi, RNA/peptide binding driven by electrostatics—insight from bidirectional pulling simulations, J. Chem. Theory Comput. 9 (2013) 1720–1730, http://dx.doi.org/10.1021/ct3009914.
[93] T.N. Do, E. Ippolotti, P. Carloni, G. Varani, M. Parrinello, Counterion redistribution upon binding of a Tat-protein mimic to HIV-1 TAR RNA, J. Chem. Theory Comput. 8 (2012) 688–694, http://dx.doi.org/10.1021/ct2005769.
[94] W. Li, B. Ma, B.A. Shapiro, Binding interactions between the core central domain of 16S rRNA and the ribosomal protein S15 determined by molecular dynamics simulations, Nucleic Acids Res. 31 (2002) 629–638, http://dx.doi.org/10.1093/nar/gkg149.
[95] K. Sanbonmatsu, Computational studies of molecular machines: the ribosome, Curr. Opin. Struct. Biol. 22 (2012) 168–174, http://dx.doi.org/10.1016/j.sbi.2012.01.008.
[96] M. Kim, H.R. Kim, S.Y. Chae, R.G. Larson, H. Lee, J.C. Park, Effect of arginine-rich peptide length on the structure and binding strength of siRNA–peptide complexes, J. Phys. Chem. B 117 (2013) 6917–6926, http://dx.doi.org/10.1021/jp402868g.
[97] Y. Mu, G. Stock, Conformational dynamics of RNA-peptide binding: a molecular dynamics simulation study, Biophys. J. 90 (2006) 391–399, http://dx.doi.org/10.1529/biophysj.105.069559.
[98] W. Ye, J. Yang, Q. Yu, W. Wang, J. Hancy, R. Luo, H.-F. Chen, Kink turn sRNA folding upon L7Ae binding using molecular dynamics simulations, Phys. Chem. Chem. Phys. 15 (2013) 18510–18522, http://dx.doi.org/10.1039/c3cp53145g.
[99] T. Schlick, Molecular Modeling: An Interdisciplinary Guide, Second edition Springer-Verlag, NY, 2010.
[100] T. Schlick, Innovations in biomolecular modeling and simulations, Volume 2, ed., The Royal Society of Chemistry, 2012.