J Mol Model (2011) 17:1953–1962
DOI 10.1007/s00894-010-0900-8

ORIGINAL PAPER

Nucleotide docking: prediction of reactant state complexes for ribonuclease enzymes

Brigitta Elsässer · Gregor Fels

Received: 18 August 2010 / Accepted: 10 November 2010 / Published online: 1 December 2010

Abstract Ribonuclease enzymes (RNases) play key roles in the maturation and metabolism of all RNA molecules. Computational simulations of the processes involved can help to elucidate the underlying enzymatic mechanism and is often employed in a synergistic approach together with biochemical experiments. Theoretical calculations require atomistic details regarding the starting geometries of the molecules involved, which, in the absence of crystallographic data, can only be achieved from computational docking studies. Fortunately, docking algorithms have improved tremendously in recent years, so that reliable structures of enzyme–ligand complexes can now be successfully obtained from computation. However, most docking programs are not particularly optimized for nucleotide docking. In order to assist our studies on the cleavage of RNA by the two most important ribonuclease enzymes, RNase A and RNase H, we evaluated four docking tools—MOE2009, Glide 5.5, QXP-Flo+0802, and Autodock 4.0—for their ability to simulate complexes between these enzymes and RNA oligomers. To validate our results, we analyzed the docking results with respect to the known key interactions between the protein and the nucleotide. In addition, we compared the predicted complexes with X-ray structures of the mutated enzyme as well as with structures obtained from previous calculations. In this manner, we were able to prepare the desired reaction state complex so that it could be used as the starting structure for further DFT/B3LYP QM/MM reaction mechanism studies.

B. Elsässer (✉) · G. Fels
Department of Chemistry, University of Paderborn,
Warburgerstr. 100,
33098 Paderborn, Germany
e-mail: elsaesse@mail.upb.de

Keywords Ribonuclease · Docking

Introduction

The genetic information associated with living cells is encoded in DNA and translated into proteins via messenger RNA molecules (mRNA) that are copied from the DNA during transcription. Both messenger RNAs, which carry genetic material that is used to make proteins, as well as noncoding RNAs, which function in various cellular processes, are degraded by ribonuclease enzymes (RNase) as part of their life cycles. In addition, active RNA degradation systems represent the first mechanism for defending against RNA viruses, and they provide the underlying machinery for more advanced cellular immune strategies. Therefore, inhibitors of these enzymes could provide new drugs to treat many diseases. Since they are important biological molecules, the enzymes human RNase H and RNase A represent key pharmaceutical targets, and understanding them at the molecular level is essential for the development of the corresponding inhibitory drugs.

RNase A is part of the human immune response. It is responsible for destroying RNA viruses that could infect the body, and it digests RNA in our food. It hydrolyzes the single-stranded RNA behind each cytosyl and uridyl nucleotide and leaves 5′-hydroxy and 3′-phosphorylated products (Fig. 1) [1]. Human RNase H is part of the reverse transcriptase (RT) enzyme, which is absolutely necessary for the proliferation of retroviruses. In contrast to RNase A, RNase H specifically cleaves the RNA strand of a DNA/RNA duplex after transcription at the nucleotides at positions 7–12 from the 3′-DNA/5′-RNA terminus (Fig. 1) [2], producing 5′-phosphorylated and 3′-hydroxy products.

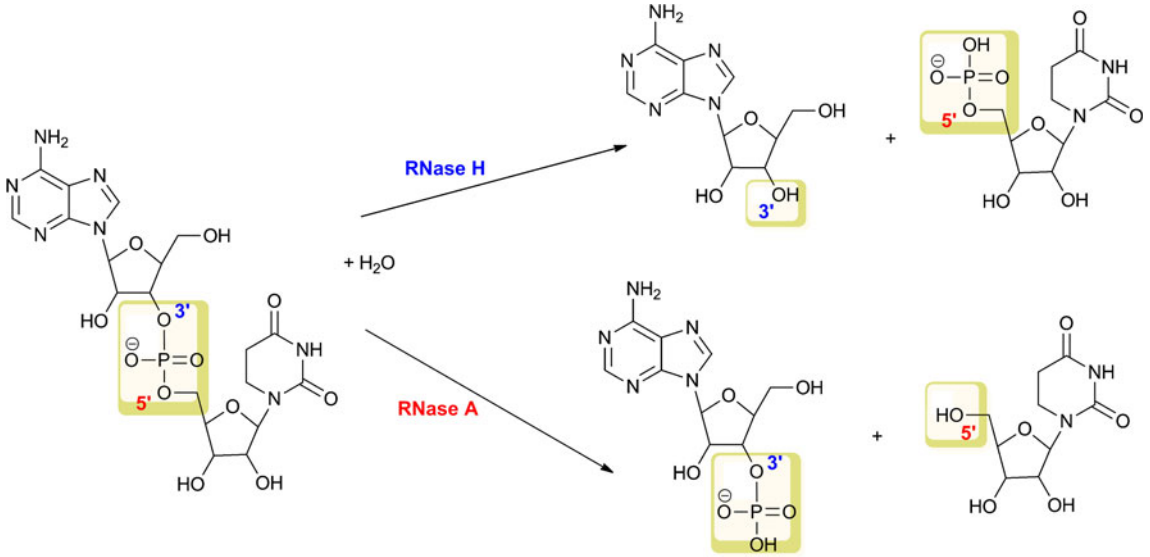

Fig. 1 Cleavage mechanisms of the two ribonuclease enzymes, RNase A and RNase H

An RNase H domain is also present at the C-terminus of retroviral reverse transcriptase, which converts a single-stranded retroviral genomic RNA into a double-stranded DNA for integration into host chromosomes, thus playing a significant role in the HIV reverse transcription process [3]. In vivo studies demonstrated that the inactivation of RNase H results in noninfectious virus particles [4, 5].

In order to investigate the hydrolytic cleavage mechanisms of the two key ribonuclease enzymes human RNase H and RNase A computationally, reliable starting geometries of the two enzyme–ligand complexes are a prerequisite. Here, docking approaches can help by predicting the binding modes of ligands inside the active site of the enzyme (i.e., the orientation and conformation of an inhibitor at the enzyme cleavage site), thereby generating potential structures of enzyme–inhibitor complexes.

In the last few decades, a number of search and placement algorithms have been developed, which differ in whether they treat the ligand in its entirety or build it up from fragment bases inside the binding site [6]. In addition, most programs can even allow for receptor flexibility. To this end, several publications have recently appeared that compare and evaluate the different docking tools for a given enzyme–ligand combination [6–9]. Almost all papers, however, consider the protein-based docking of small ligand molecules or protein–protein docking, with the exception of van Dijk et al., who recently developed a new method of protein–DNA docking [10–12]. Predicting the correct binding between proteins and RNA oligonucleotide chains is a topic of growing interest [13–15], although it is rarely desired in the literature.

Since our main goal is to investigate the hydrolytic mechanism of the above desired enzymes, we need to generate starting structures for our calculations, which, in the absence of crystallographic data, can be generated by molecular docking. In order to find the optimal docking procedure for our purposes, we compared four docking programs (MOE2009 [16], Glide 5.5 [17], QXP-Flo+0802 [18], and Autodock 4.0 [19]). Here, we describe our investigation of possible ribonuclease–ribonucleotide oligomer complexes by molecular docking. Although calculating the RMSD (root mean square deviation) values between docking results and crystal structures is a well-established method of evaluating docking poses, we instead applied the interaction-based accuracy classification (IBAC) method as described by Kroemer et al. [6], as the necessary protein–ligand interactions for each complex are well-described in the literature [6, 20–22]. Therefore, we determined the correct docking poses by comparing the hydrogen bond (H-bond) distances between enzyme and nucleotide. This allowed us to identify the the docking program that is most suitable for our nucleotide docking scenario, as well as to generate possible starting structure geometries for our calculations. However, we did not intend to establish any ranking between the applied docking tools.

Materials and methods

Protein preparation in silico

The crystal structures of both ribonuclease–ribonucleotide complexes were retrieved from the PDB database [23]. The protein preparation for both ribonucleases was carried out using the parallel software package NWChem [24]. After adding hydrogen atoms, the protein was solvated in an

80 Å cubic box of water, and the system was relaxed using stepwise molecular dynamic calculations and optimized by applying high-level B3LYP/DFT QM/MM simulations. Afterwards, the ligand (ribonucleotide) was removed and the resulting enzyme was employed for docking.

The three-dimensional coordinates of the ligand-free RNase A are also available in the protein database at atomic resolution (PDB code: 2E3W) [25], but the superposition of this structure with other RNase A enzymes complexed with any ligands (e.g., 1RUV, 1RPG) revealed that, probably upon ligand binding, one of the important active-site histidine residues (His119) changes its orientation, since in the ligand-free enzyme the imidazol ring of His119 is flipped over. For this reason, the ligand-free enzyme was not suitable for docking.

In the first set of calculations we started from the productive binding complex of RNase A with deoxycytidyl-3′,5′-deoxyadenosine (PDB code: 1RPG, resolution 1.40 Å) [26]. According to the generally accepted mechanism [1, 27], His12 was unprotonated, and Lys41 and His119 were protonated. In addition, to evaluate our docking method, we also used the RNase A–cyclic uridyl phosphate (CUP) complex from previously published QM/MM calculations [28] to perform docking studies between CUP and the protein in order to reproduce this structure for verification.

The starting point for the second set of calculations was the crystal structure of the human RNase H1 catalytic domain mutant D210N in complex with a 14mer, the DNA:RNA duplex (PDB code: 2QKK) [22]. Prior to protein preparation as described above, the residue Asn210 was mutated in silico to Asp210 in order to reproduce the wild-type protein. Moreover, the optimized complex served as comparison to the docking results.

Ligand preparation

To verify the procedure of the RNase A docking, we first used the ligands from the original optimized enzyme–substrate complexes described above (deoxycytidyl-3′,5′-deoxyadenosine and cyclic uridyl-phosphate, CUP, respectively).

For the desired active reactant state complexes, RNA(poly (U)) oligomers (dimer, trimer, tetramer nucleotides) from the PubChem [29] database served as ligands (CID: 439261). The appropriate ligand files were prepared by adding and optimizing the hydrogen atoms using the Protein Preparation Wizard of Maestro 9.0 [17], and then the 3D conformation of each molecule was rebuilt using LigPrep 2.3 [17] at a pH of 7.0. Afterwards, the rebuilt ligands were subjected to a conformational search using Macromodel 9.7 [17] with the OPLS 2005 force field in a water solvent model. For minimization, the "steepest descent" method was applied with a maximum number of iterations of 500. The energy window for retaining structures was set to 5.02 kcal mol^{-1}, and the RMSD cutoff value was set to 0.5 Å to generate the lowest energy conformations.

For the docking experiment with RNase H, a tetramer RNA:DNA hybrid duplex was extracted from the original ligand of the X-ray structure. Since none of the docking programs can simultaneously consider two molecules such as those present in the RNA:DNA double helix, a covalent ether bridge was built between the RNA and the DNA chain without changing the conformation of the double helix. This ligand was used for docking after molecular mechanics optimization in MOE with Amber99 as force field. After docking, the ether bridge was removed and the corresponding nucleotide base was optimized to build up the hydrogen bond between the two ribonucleotide chains. Additional ligands were also generated by using a single-stranded dimer, trimer, and tetramer RNA nucleotide chain of the hybrid duplex. These structures were then docked rigidly in order to retain their original geometries.

All figures containing molecular structures were prepared with MOE2009.

Docking programs and parameters

MOE 2009.10 [16] In the potential energy setup panel, Amber99 was chosen as the force field. The implicit Born model [30] was selected for solvation. Two placement methods, "alpha triangle" and "triangle matcher," were employed to find the optimal docking parameters. The alpha triangle method was found to be faster and to yield more suitable docking poses that show the required enzyme–ligand interactions. In all cases, scoring was done by the London dG [31] method, and force field refinement was applied, allowing for flexibility of the catalytic site within 7.0 Å. Each run was adjusted to retain 30 docked conformations as a cut-off unless less suitable poses were found. The top poses were retained for visual analysis (investigating the H-bond distances) for each nucleotide per enzyme. For RNase A, the residues Gln11, His12, Lys41 and His119 were defined as the binding pocket [1], while the residues Asp145, Glu186, Asp210 and Asp274 defined the pocket for RNase H [22]. "Rotate bond" of the Docking Simulation panel was enabled in both cases.

Glide 5.5 [17] For each protein, a grid box of 30×30×30 Å with a default inner box (10×10×10 Å) was first centered on the catalytic pocket using the same active site residues to describe the pocket as above. Default parameters were used, and for RNase A four constraints were defined. One positional constraint was set for the phosphate group and three donor–acceptor (H-bond) constraints were applied between His12, Lys41 and His119 and the given sugar moiety oxygen atoms, which represent the most important

interactions between the enzyme and the ligand. The positional constraint was used for RNase H only, since the amino acids of the catalytic site are not involved in the substrate binding interaction [21]. For all experiments, the standard precision mode of GlideScore was selected as the scoring function, and the options of "dock flexibly" and "dock rigidly" were selected for RNase A and RNase H, respectively, in order to retain the original helix-like conformation of the RNA strand. The top 10 poses were kept and the H-bond between the ligand and the protein analyzed.

QXP-Flo+0802 [18] In a multistep procedure [7], the full Monte Carlo docking (SDOCK+) was followed by a local Monte Carlo run (MCDOCK), which is also part of the program package. The 25 hits of the full MC run were ranked, rescored and redocked in a local MC simulation as described by Alisaraie et al. [7]. The oxygen atoms of the catalytic site water were colored purple and the hydrogens blue to enable some movement. The ligand was kept rigid for the RNase H docking runs. Docking results were evaluated according to the IBAC scheme [6].

Autodock 4.0 [19] The active site residues were defined as flexible residues, and the rest of the protein was designated as rigid. The grid maps representing the protein in the docking process were calculated with Autogrid 4.0. The dimension of the grid was 80×80×80 points, with a spacing of 0.375 Å employed between the grid points, and the center was close to the ligand.

Default docking parameters were used, and the Lamarckian genetic algorithm (LGA) was applied for the search parameters [32]. For the ligands of RNase A, the number of rotatable torsion angles was set by AutoTors. To disable bond rotation for the RNase H substrates, the number of torsions was set to 0. The resulting structures were automatically ranked according to their mean docking energies by the scoring function of AutoDock. Subsequently, the relevant H-bond lengths were analyzed.

Evaluation of docking results

ProFit V3.1 [33] For the results of the docking studies with single-stranded RNA, RMSD values were calculated using the McLachlan algorithm [34], as implemented in the program. For the simulations with the modified RNA:DNA double helix, no RMSD calculation is possible in this form, since the mobile and reference structures must have the same number of atoms. Therefore, the ether bridge was removed after docking and the 14mer double helix of the reference structure was shortened to a tetramer for the RMSD calculations.

IBAC (interaction-based accuracy classification [6]*)* The predicted poses were analyzed with respect to the essential key interactions with the protein.

LPC (ligand–protein contacts) [35] Extended analysis of all heavy atom contacts was performed using the LPC CSU server of the Weizmann Institute.

Results and discussion

Software evaluation for nucleotide docking

The goal of our simulations was to produce reliable starting structures for further QM/MM reaction mechanism studies.

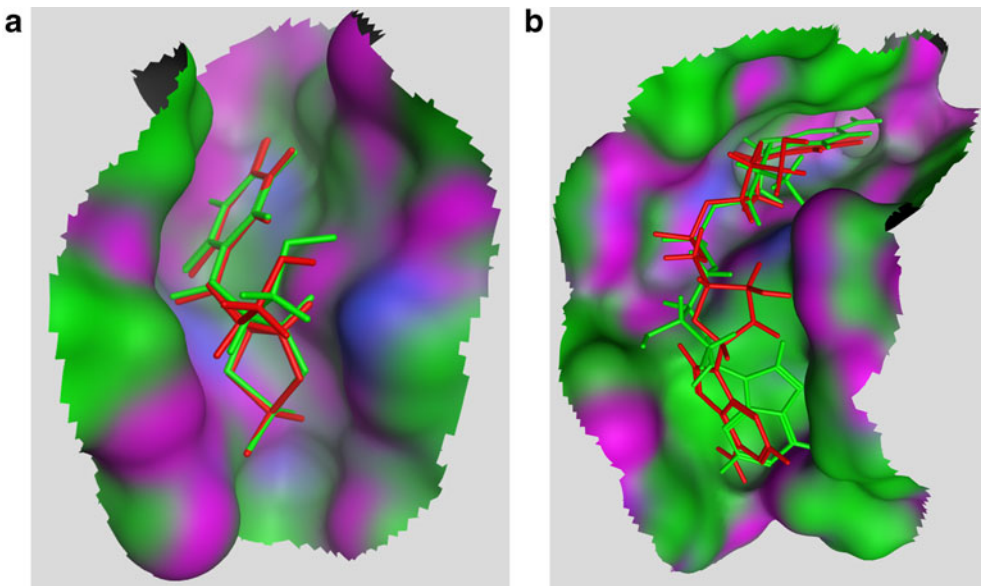

Fig. 2a–b Superposition of the top binding poses from bound docking with MOE. Reproduced complexes of RNase A and cyclic uridyl-phosphorane (RMSD=0.097; **a**), and deoxy-cytosyl-adenyl-dinucleotide (RMSD=0.327; **b**). The original pose is colored *green* and the docking hit is *red*. The displayed surface illustrates the catalytic pocket, in which a *purple* color identifies H-bond interactions, a *green* color characterizes areas with hydrophobic contacts, and a *blue* color denotes dipole–dipole interactions

Although currently available crystal structures from the PDB database provide fundamental information about the active sites of both RNase A and RNase H, it is not possible to generate the coordinates of their reactant state complexes experimentally since the nucleotide is hydrolyzed immediately upon binding. Therefore, the use of docking programs to build a productive initial complex structure is crucial. Since docking programs are not particularly optimized for nucleotide docking, we have initially evaluated four docking programs, MOE2009 [16], Glide 5.5 [17], QXP-Flo+0802 [18], and Autodock 4.0 [19] for their accuracy in reproducing the experimentally known crystal structures of inactive RNase–ligand complexes. In all docking simulations we started from the DFT/B3LYP QM/MM optimized complexes (reference molecule) derived from the available X-ray data, as described in the "Materials and methods" section. The ligand of the existing complex was removed from its binding site and docked back into the catalytic pocket (bound docking). This method affords an unbiased setup to evaluate the software and the scoring function. The resulting structures were analyzed and the docking accuracy was determined by comparing the H-bond distances and the distances between the superposed structures of the given docking pose and the reference molecule. Additionally, ProFitV3.1 [33] was applied to calculate the RMSD values between the docking pose and the reference molecule by superposing the complete enzyme–substrate complex.

MOE and Glide demonstrated good accuracy in the prediction of binding modes, with RMSD values for the best docking hits of below 1 Å. Figure 2 shows the superposition of the top binding pose with the original structure.

QXP and Autodock, however, could not reproduce the initial structure by bound docking at all. The resulting complex did not match the productive binding structure, as the ligand was placed incorrectly into the catalytic pocket and therefore neither a RMSD validation nor the LPC analysis would be applicable to describe these docking results.

As one can see from Tables 1 and 2, although MOE finds fewer correct docking poses than Glide, the resulting hits have significantly better RMSD values. Another interesting point is that in both docking studies, the order of docking scores does not correlate with the order of RMSD values and hence the reproducibility of the reference structure. In addition, LPC analysis [35] revealed that, with respect to the X-ray structure, MOE could recover 76% and Glide 61% of all heavy atom contacts (below 4.5 Å) with a deviation of 0.6 Å or less.

Table 1 RMSD values of the best docking poses of the RNase A–cyclic uridyl-phosphate complex. Only RMSD values below 1 Å were considered

	Hit 1	Hit 2	Hit 3	Hit 4	Hit 5
MOE	0.097	0.122	0.117	0.181	
Glide	0.775	0.557	0.748	0.882	0.546

Table 2 RMSD values of the best docking poses of the RNase A–deoxy-cytosyl-adenyl-dinucleotide complex. Only RMSD values below 1 Å were considered

	Hit 1	Hit 2	Hit 3	Hit 4	Hit 5
MOE	0.392	0.327	0.511	0.494	
Glide	0.860	0.792	0.910	0.869	0.974

For the docking studies of RNase H and the RNA:DNA hybrid duplex (see the "Materials and methods" section), the ligand was modified by connecting the RNA and DNA chain with an ether bridge so that docking could be performed with the double helix. However, the modification did not affect the conformation of the hybrid duplex. Since the original ligand has been modified, comparison of the RMSD values for the whole system was possible only when the ether bridge was removed after the docking simulation (Fig. 3, RMSD=0.711). In addition, the position of the ligand was compared by measuring mechanistically relevant distances between the enzyme and the nucleotide in a similar manner to the IBAC method of Kroemer et al. [6].

Table 3 compares the most relevant distances between the docked ligand and RNase H to the reference structure. Note that although the docked complexes match the low-resolution (3.20 Å) experimental structure quite well, the measured distances are slightly overestimated by both software packages. This may be due to the flexible docking applied, where the active site residues were allowed to move, with the consequence that the binding pocket is a

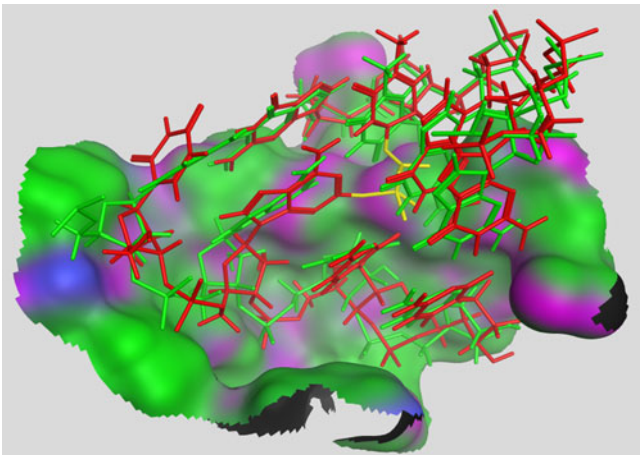

Fig. 3 Top hit of bound docking of RNase H and the modified tetramer RNA:DNA hybrid duplex (RMSD=0.711; Fig. 2c). The original pose is colored *green* and the docking hit is *red*. The displayed surface illustrates the catalytic pocket. Color codes are as specified in Fig. 2

Table 3 Comparison of distances (in Å) between docking results (best hit) and the corresponding experimental structure of the RNase H–tetramer hybrid duplex (PDB code: 1RPG). *pro*O1P is the O1P atom of the following nucleotide. To measure the distance between the centroids of the sugar rings, a dummy atom was placed at the center of the five-membered ring

	Reference structure	MOE		Glide	
		Distance	Difference	Distance	Difference
O3′–(OD1)Asp210	3.35	3.71	0.36	3.83	0.48
O3′–(OE1)Glu186	3.11	3.21	0.10	3.41	0.30
O1P–(OE1)Glu186	3.38	3.86	0.48	3.71	0.33
O1P–(OD1)Asp145	2.87	3.25	0.38	3.66	0.79
O1P–(OD2)Asp145	3.26	3.73	0.47	3.75	0.49
O1P–(OD1)Asp274	4.50	4.65	0.15	4.46	-0.04
*pro*O1P-(OD1)Asp274	3.83	4.28	0.45	4.50	0.67
P(ref)−P(docked)	-		0.37		0.73
Centroid of sugar ring for nucleotide 1	-		0.37		1.85
Centroid of sugar ring for nucleotide 2	-		0.68		2.18

little expanded upon force-field optimization of the side chain positions.

To evaluate MOE and Glide for docking a tetramer RNA:DNA double helix into the active site of RNase H, the experimental and docked structures were manually compared by measuring the distances of atoms that are relevant to the mechanism (Table 3). Although MOE and Glide are both suitable for reproducing the experimental results, again MOE showed significantly better performance in terms of the average accuracy for docking nucleotides. According to LPC analysis [35], both MOE and Glide recovered roughly 50% of all heavy atom contacts (below 4.5 Å), with a deviation of 0.6 Å or less from the reference structure.

The RNase A reaction state complex

Based on these results, we chose to use MOE and Glide to find suitable reactant states for our studies of the reaction mechanisms of RNase A and RNase H. For RNase A, the literature provides a putative pathway [1] of the trans-phosphorylation step (Fig. 4) and defines a number of H-bonds between the enzyme and the ligand. The first step in the proposed mechanism involves intramolecular trans-phosphorylation to form a stable cyclic phosphate intermediate product and displacement of the O5P-nucleotide product, as illustrated in Fig. 4. Since RNase A cleaves single-stranded RNA behind uridyl and cytosyl nucleotides only, dimer, trimer, and tetramer uridyl nucleotide oligomers were used for the docking simulation in the presence of crystallographic water molecules.

According to the putative reaction mechanism [36] as presented in Fig. 4, the first step involves the protonation of His12 by the OH group of O2′, followed by a proton transfer from His119 to O5P. Accordingly, strong H-bonds between (NE2)His12 and O2′ and between (ND1)His119 and O5P are indispensible, and the corresponding distances should therefore be as short as possible in correct docking poses.

At the intermediate or transition state stage, the phosphorane and the partial negative charge on the phosphate oxygens must be stabilized by further interactions between (NZ)Lys41–O2′, (NE2)Glu11–O2P and (NE2)His12–O1P. Therefore, the resulting hits were analyzed with respect to the corresponding H-bond distances. Tables 4 and 5 show that the alpha-triangle placement method, as implemented in MOE, yielded significantly more possible docking poses than GlideScore SP. The H-bond

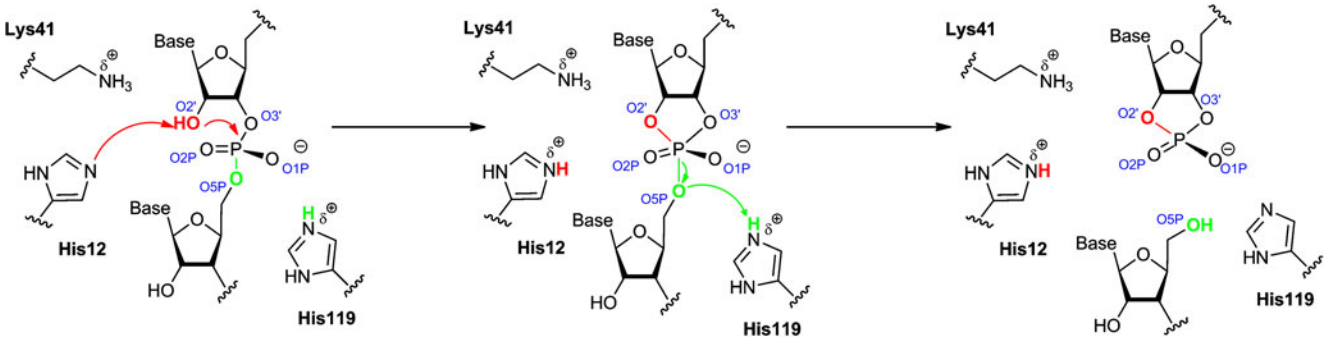

Fig. 4 Putative reaction mechanism of the first step of RNase A hydrolysis

Table 4 Relevant H-bond distances (in Å) between the uridyl nucleotide dimer and active site residues of RNase A after docking with MOE using the alpha-triangle placement method. The top hit, with the shortest average H-bond distances, is shown in bold

MOE	Hit 1	Hit 2	Hit 3	Hit 4	Hit 5	Hit 7
(NZ)Lys41–O2′	2.94	2.98	2.97	**2.98**	3.02	2.94
(NE2)His12–O2′	3.23	3.19	3.12	**3.15**	3.26	2.93
(NE2)Glu11–O2P	2.98	3.01	3.47	**2.93**	3.15	3.56
(NE2)His12–O1P	3.54	3.89	3.54	**3.30**	3.94	3.95
(ND1)His119–O5P	2.77	3.13	2.70	**2.96**	2.98	2.82

analysis reveals that, aside from the number of hits found, the docking efficiency of Glide and MOE are comparable. Figure 5 illustrates the catalytic pocket with the ligand (uridyl nucleotide dimer) of the docking pose that best fits the desired H-bond interactions.

In order to adjust the enzyme ligand complex to the natural structure and investigate the interactions far from the scissile phosphate bond, the size of the ligand was increased by docking a trimer and a tetramer uridyl phosphate, respectively, into the active site of RNase A. This resulted in a drastic decrease in meaningful docking poses; just one acceptable hit for each ligand was obtained for both docking tools.

Surprisingly, the derived structure reproduced the natural structure along with all of the important interactions between protein and nucleotide quite well (Table 6).

Cation titration experiments suggest that the interaction between the enzyme and a single-stranded RNA extends well beyond the phosphate group through coulombic interactions [37]. The binding subsites of RNase A can be divided into P and B subsites, in which the phosphate and the nucleotide base interact with the protein as described by Raines [1]. In our docking complex, the uridyl base of the B1 subsite interacts with Thr45, and the phosphate group of the P2 subsite interacts with Lys7 and Arg10, as illustrated in the ligand interaction map depicted in Fig. 6. This map also shows that both short- and long-range interactions between the protein and the trimer nucleotide were reproduced in the docking simulation.

Table 5 Relevant H-bond distances (in Å) between the uridyl nucleotide dimer and active site residues of RNase A after docking with GlideScore SP. The top hit, with the shortest average H-bond distances, is shown in bold

Glide	Hit 1	Hit 5	Hit 8
(NZ)Lys41–O2′	3.01	**2.99**	2.86
(NE2)His12–O2′	2.84	**2.87**	2.89
(NE2)Glu11–O2P	2.94	**2.84**	2.93
(NE2)His12–O1P	3.76	**3.48**	3.83
(ND1)His119–O5P	2.76	**2.87**	2.88

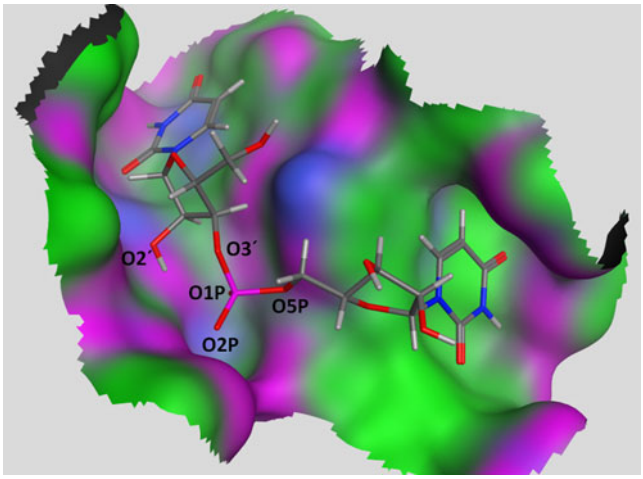

Fig. 5 Best docking pose from MOE docking of dimeric uridyl nucleotide into the active site of RNase A. The displayed surface was generated by MOE and illustrates the catalytic pocket. Color codes are as specified in Fig. 2

Lys41 and His12 are both strongly H-bonded to the O2′ oxygen and Gln11 is coordinated to O2P. Furthermore, His12 interacts with the O1P oxygen and His119 with the O5P atom. All of these interactions are required for the reaction mechanism to proceed. In addition, the active site water molecules as well as other crystallographic water molecules play important roles in stabilizing the complex through H-bridges between the protein and the ligand.

To summarize these results, our docking simulations of the uridyl nucleotides and RNase A yield a ligand pose that reproduces all of the important H-bonds at the catalytic site as described in the literature [1]. In addition, the interactions beyond the cleavage site were also reproduced correctly. Accordingly, we have used the trimer uridyl docking pose of MOE as the desired reaction state complex, which can then be used as a starting structure for further calculations.

Table 6 Relevant H-bond distances (in Å) between the ligand and active site residues of RNase A after docking with MOE and Glide, respectively. The top hit is shown in bold

	MOE trimerU	MOE tetramerU	Glide trimerU	Glide tetramerU
(NZ) Lys41–O2′	**2.93**	3.86	3.00	**2.69**
(NE2) His12–O2′	**3.05**	3.06	2.75	**2.72**
(NE2) Glu11–O2P	**2.84**	3.54	2.96	**2.72**
(NE2) His12–O1P	**2.95**	4.83	3.94	**4.36**
(ND1) His119–O5P	**2.85**	2.97	2.74	**3.83**

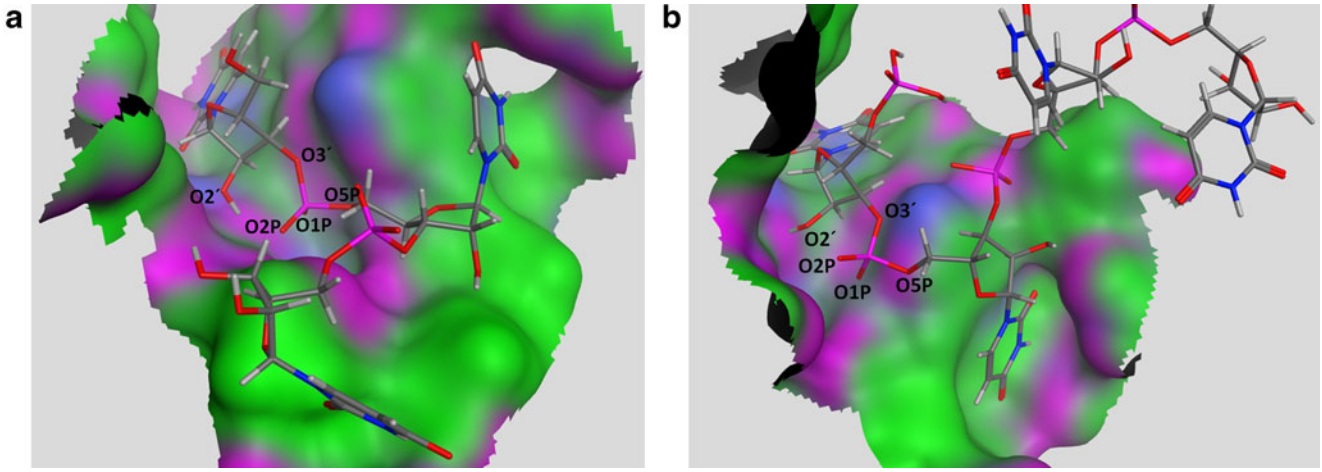

Fig. 6a–c Top hit docking poses of trimer (**a**) and tetramer (**b**) uridyl nucleotide and RNase A. Corresponding ligand interaction map as generated by MOE (**c**). The displayed surface illustrates the catalytic pocket. Color codes are as specified in Fig. 2

The RNase H reaction state complex

Recently, the first crystal structures of the catalytic domain mutant of human RNase H [22] and *Bacillus halodurans* RNase H [38] complexed with an RNA:DNA hybrid substrate were reported. Nowotny et al. have shown that the RNA strand of the hybrid duplex is recognized by the protein through its interaction with the 2′-OH groups, and that the active site of human RNase H consists of the amino acids Asp145, Glu186, Asp210 and Asp274, which are all essential to achieve the required activity [22]. However, the amino acids of the catalytic site are not involved in the substrate binding interaction [21].

In our reaction mechanism studies [39] (Fig. 7), Nowotny's human RNase H1 catalytic domain mutant D210N in complex with the 14-mer RNA/DNA hybrid was used as a starting structure for our calculations (PDB code 2QKK [22]). In order to turn the system back to the natural active enzyme–substrate complex, we mutated Asn210 to Asp210. Separate QM/MM optimization proved that the mutation hardly affects the structure of the system or the active site. To further confirm the validity of this procedure, we performed docking calculations to generate the corresponding reactant state complex.

Since none of the docking tools used can handle a duplex structure like the RNA:DNA double helix, we have employed RNA trimer and tetramer nucleotides (cut from the RNA:DNA double helix) for docking into the active site of RNase H in the presence of water molecules and the Mg^{2+} ions that exist in the binding site. In order to retain the geometry of the original double helix the ligand was kept rigid during the simulations by allowing no bond rotations at all for the nucleotide. Afterwards the resulting structures were compared with the QM/MM optimized active complex from our previous calculations (reference structure) [39] and the relevant protein-ligand distances were analyzed and are summarized in Table 7.

Due to the fact that we used single-stranded oligomers in our studies, it is not possible to calculate RMSD values for the whole system. Therefore, we compared the distances

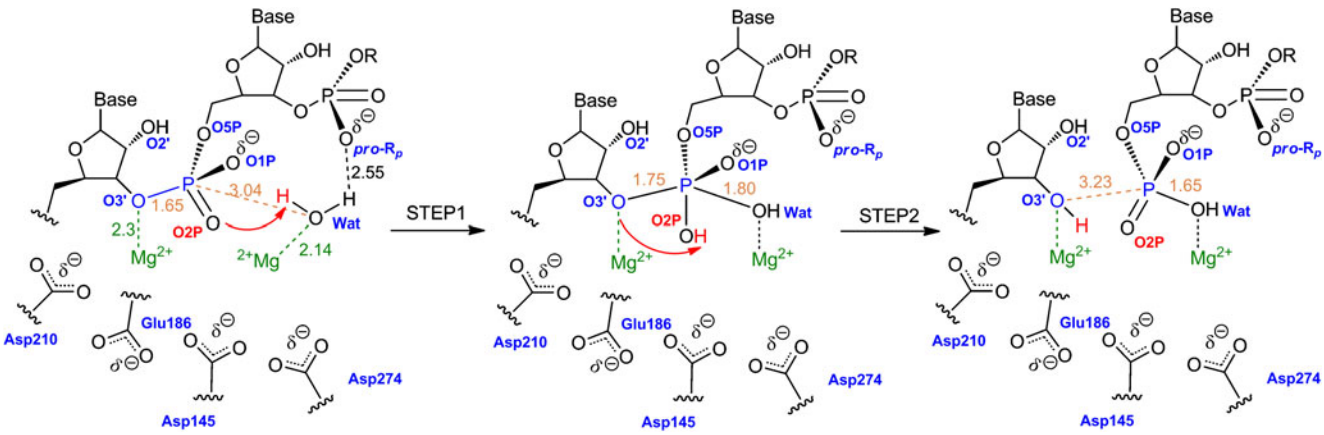

Fig. 7 Proposed reaction pathways [1] for the O3′−P cleavage of human RNase H

Table 7 Comparison of relevant protein–ligand interactions between docking results (best hit) and the reference structure. *pro*O1P is the O1P atom of the following nucleotide. Distance and different values are given in Å

	Ref.	MOE tetramer		MOE trimer		Glide tetramer		Glide trimer	
		Dist.	Diff.	Dist.	Diff.	Dist.	Diff.	Dist.	Diff.
O3′–(OD1)Asp210	3.35	3.57	0.22	2.94	-0.41	3.41	0.06	2.88	-0.47
O3′–(OE1)Glu186	3.11	3.23	0.12	3.16	0.05	3.12	0.01	3.24	0.13
O1P–(OE1)Glu186	3.38	3.52	0.14	3.63	0.25	3.53	0.15	3.60	0.22
O1P–(OD1)Asp145	2.87	3.12	0.25	2.72	-0.15	3.02	0.15	2.67	-0.20
O1P–(OD2)Asp145	3.26	3.44	0.18	3.09	-0.17	3.33	0.07	3.03	-0.23
O1P–(OD1]Asp274	4.50	4.60	0.10	4.30	-0.20	4.44	-0.06	4.30	-0.20
*pro*O1P–(OD1)Asp274	3.83	3.86	0.03	4.03	0.20	4.24	0.41	3.96	0.13

between the nucleotide and significant residues of the active site with our QM/MM optimized structure. The number of possible docking poses obtained by MOE is much larger than that obtained with Glide, but the resulting top hits are comparable, as can be judged from the analyzed interactions.

Both structures are found to be very close to the reference structure (Fig. 8). This strongly supports the idea that mutating Asn210 to Asp210 in order to construct the active enzyme did not affect the geometry of the catalytic site. As can be seen in Fig. 8, the nucleotides in the deeper regions of the pocket show excellent fits, but the nucleotide base at the end of the pocket instead turns towards the exit of the pocket due to the fact that there are fewer interactions with the protein in that region.

Conclusions

With the aim of predicting reactant state structures that can be used to calculate the reaction mechanisms for RNase A and RNase H hydrolysis, we evaluated four protein-based docking tools with respect to their enzyme–nucleotide docking capabilities. To this end, different ribonucleotide oligomers were placed into the active sites of RNase A and RNase H. As molecular docking tools were originally developed to place small ligands into protein-binding sites, or for protein–protein interactions, we tested the docking tools Glide, QXP-Flo+, and Autodock as well as the docking algorithm implemented in MOE for their abilities to reproduce inactive protein–RNA complexes known from the literature. While QXP and Autodock could not reproduce the crystal structures in a bound docking experiment with a known structure, MOE and Glide demonstrated good accuracy in the prediction of binding modes, with RMSD values for the best docking hits below 1 Å, even for large ligands like uridyl tetramer nucleotide. According to our findings, these two docking programs could successfully reproduce the reference experimental structures and interactions beyond those at the active site. Since our initial goal was to find suitable active reactant state complexes for our studies of the reaction mechanisms of RNase A and RNase H, we chose MOE and Glide to accomplish our project.

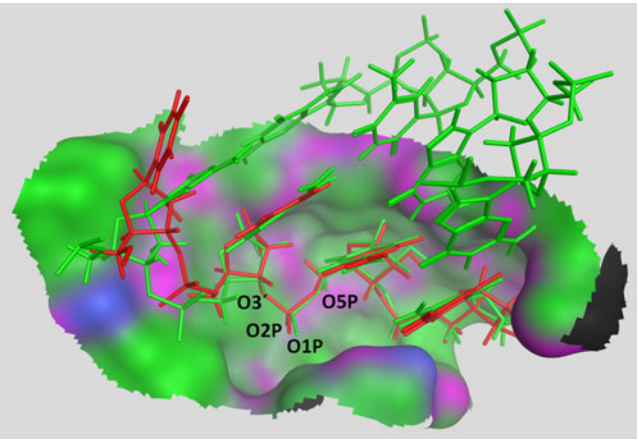

Fig. 8 Superposition of the reference structure (*green*) and the top tetramer nucleotide docking hit (*red*). Color codes are as specified in Fig. 2

In the docking studies of the mutated RNase H and a modified tetramer RNA:DNA hybrid duplex, we were able to correctly position the ligand into the active site of the enzyme. Therefore, we can assume that docking experiments between the nucleotide and the active enzyme provided a reasonable complex structure, which further supports the notion that the complex in our reaction mechanism studies [39] was the correct starting structure.

The results of the present studies prove that current protein-based docking programs can also be effectively used for nucleotide docking, and—at least in the special cases we have investigated—they can generate reactant state complexes for further theoretical studies. Our simulations with RNase A identified a few appropriate docking poses that were subjected to further analysis in terms of the important interactions between the enzyme and the tetramer nucleotide. In addition, MOE provided the desired productive initial structure for our intended reaction mechanism studies. Previously, we have studied [28] the second step (hydrolysis step) of the RNase A catalytic cycle, and

utilizing the docked reactant state structure derived from our docking simulation, we can now continue our investigation towards the first step of the reaction mechanism. These calculations should eventually result in the first description of the complete pathway on a high-level DFT-based QM/MM basis.

Acknowledgments The QM/MM optimization of the generated complexes was performed using EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

References

1. Raines RT (1998) Chem Rev 98:1045–1065
2. Wu HJ, Lima WF, Crooke ST (2001) J Biol Chem 276:23547–23553
3. Hughes SH, Arnold E, Hostomska Z (1998) In: Crouch RJ, Toulme JJ (eds) Ribonuclease H. Inserm, Paris, pp 195–224
4. Schatz O, Cromme FV, Gruningerleitch F, Legrice SFJ (1989) Febs Lett 257:311–314
5. Tanese N, Goff SP (1988) Proc Natl Acad Sci USA 85:1777–1781
6. Kroemer RT, Vulpetti A, McDonald JJ, Rohrer DC, Trosset JY, Giordanetto F, Cotesta S, McMartin C, Kihlen M, Stouten PFW (2004) J Chem Inf Comput Sci 44:871–881
7. Alisaraie L, Haller LA, Fels G (2006) J Chem Inf Model 46:1174–1187
8. Li YZ, Shen J, Sun XG, Li WH, Liu GX, Tang Y (2010) J Chem Inf Model 50:1134–1146
9. Sandor M, Kiss R, Keseru GM (2010) J Chem Inf Model 50:1165–1172
10. van Dijk ADJ, de Vries SJ, Dominguez C, Chen H, Zhou HX, Bonvin A (2005) Proteins Struct Funct Bioinf 60:232–238
11. van Dijk M, Bonvin A (2008) Nucleic Acids Res 36:e88
12. van Dijk M, van Dijk ADJ, Hsu V, Boelens R, Bonvin A (2006) Nucleic Acids Res 34:3317–3325
13. Issur M, Despins S, Bougie I, Bisaillon M (2009) Nucleic Acids Res 37:3714–3722
14. Kamphuis MB, Bonvin A, Monti MC, Lemonnier M, Munoz-Gomez A, van den Heuvel RHH, Diaz-Orejas R, Boelens R (2006) J Mol Biol 357:115–126
15. Suydam IT, Levandoski SD, Strobel SA (2010) Biochemistry 49:3723–3732
16. Hostomska Z, Matthews D, Hostomsky Z (1993) J Acq Im Def Syn Hum Ret 6:673–673
17. Schrödinger LLC (2009) Macromodel v.9.7, Glide v.5.5, Maestro v.9.0, LigPrep v.2.3. Schrödinger LLC, New York
18. McMartin C, Bohacek RS (1997) J Comput Aided Mol Des 11:333–344
19. Morris GM, Goodsell DS, Huey R, Lindstrom W, Hart WE, Kurowski S, Halliday S, Belew R, Olson AJ (2007) Autodock 4.0. http://autodock.scripps.edu/
20. Fedoroff OY, Salazar M, Reid BR (1993) J Mol Biol 233:509–523
21. Katayanagi K, Okumura M, Morikawa K (1993) Proteins Struct Funct Genet 17:337–346
22. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W (2007) Mol Cell 28:513–513
23. Research Collaboratory for Structural Bioinformatics (2010) RCSB protein data bank. Rutgers/SDSC, Piscataway/La Jolla
24. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, de Jong WA (2010) Comput Phys Commun 181:1477–1489
25. Boerema DJ, Tereshko VA, Kent SBH (2008) Biopolymers 90:278–286
26. Zegers I, Maes D, Daothi MH, Poortmans F, Palmer R, Wyns L (1994) Protein Sci 3:2322–2339
27. Wladkowski BD, Krauss M, Stevens WJ (1995) J Am Chem Soc 117:10537–10545
28. Elsässer B, Valiev M, Weare JH (2009) J Am Chem Soc 131:3869–3871
29. Mueller GA, Pari K, DeRose EF, Kirby TW, London RE (2004) Biochemistry 43:9332–9342
30. Onufriev A, Case DA, Bashford D (2002) J Comput Chem 23:1297–1304
31. Eisenschitz R, London F (1930) Z Phys A Hadrons Nuclei 60:491–527
32. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) J Comput Chem 19:1639–1662
33. Andrew CR, Martin Craig TP (2009) ProFit v.3.1. http://www.bioinf.org.uk/software/profit/
34. McLachlan AD (1982) Acta Crystallogr A 38:871–873
35. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M (1999) Bioinformatics 15:327–332
36. Pauling L (1960) The nature of the chemical bond. Cornell University Press, Ithaca
37. Record MT, Lohman TM, Dehaseth P (1976) J Mol Biol 107:145–158
38. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Cell 121:1005–1016
39. Elsässer B, Fels G (2010) Phys Chem Chem Phys 12:11081–11088