



HAL
open science

Nucleotide docking: prediction of reactant state complexes for ribonuclease enzymes

Brigitta Elsässer, Gregor Fels

► **To cite this version:**

Brigitta Elsässer, Gregor Fels. Nucleotide docking: prediction of reactant state complexes for ribonuclease enzymes. *Journal of Molecular Modeling*, 2010, 17 (8), pp.1953-1962. 10.1007/s00894-010-0900-8 . hal-00646937

HAL Id: hal-00646937

<https://hal.science/hal-00646937>

Submitted on 1 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nucleotide docking: prediction of reactant state complexes for Ribonuclease enzymes

Received: 18.08.2010 / **Accepted:** 04.11.2010

Brigitta Elsässer[✉] and Gregor Fels

University of Paderborn, Department of Chemistry, Warburgerstr. 100, D-33098 Paderborn

[✉]Phone: +49-5251-605752; Fax: +49-5251-603245; Email: elsaesse@mail.upb.de

Abstract

Ribonuclease enzymes (RNases) play key roles in maturation and metabolism of all RNA molecules. Computational simulation of the processes involved help understanding the underlying enzymatic mechanism and is often employed in a synergistic approach together with biochemical experiments. Theoretical calculation require atomistic details on starting geometry of the molecules involved which in the absence of crystallographic data can only be achieved from computational docking studies. Fortunately, docking algorithms during the past years have tremendously improved so that today reliable structures of enzyme-ligand complexes can successfully be obtained from computation. However, most docking programs are not particularly developed 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 have 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 have analyzed the docking results with respect to the known key interactions between the protein and the nucleotide. In addition, we have compared the predicted complexes with X-ray structures of the mutated enzyme as well as with structures of previous calculations. In this manner we could prepare the desired reaction state complex as starting structure for further DFT/B3LYP QM/MM reaction mechanism studies.

Keywords Ribonuclease · Docking

Introduction

The genetic information of living cells is encoded in the DNA and is translated into proteins via messenger RNA-molecules (mRNA) that are copied from the DNA during transcription. Both, messenger RNAs, that carry genetic material for making proteins, as well as non-coding RNAs, that function in varied cellular processes, are degraded by Ribonuclease enzymes (RNase) as part of their life cycle. In addition, active RNA degradation systems are the first defense mechanism against RNA viruses, and provide the underlying machinery for more advanced cellular immune strategies. Therefore, inhibitors of these enzymes could provide new drugs against many diseases. As important biological molecules, both enzymes, *human* RNase H and RNase A, represent key pharmaceutical targets and their understanding on a molecular basis is a keystone in the development of corresponding inhibitory drugs.

RNase A is part of the human immune responses and 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. Contrary to RNase A, RNase H specifically cleaves the RNA strand of a DNA/RNA duplex after transcription at position 7-12 nucleotides from the 3'-DNA/5'-RNA terminus (Fig. 1) [2] and leaves 5'-phosphorylated and 3'-hydroxy products. A 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 inactivation of RNase H results in non-infectious virus particles [4, 5].

<Figure 1>

In order to investigate the hydrolytic cleavage mechanism of the two key Ribonuclease enzymes human RNase H and RNase A, by computational approaches reliable starting geometries of the two enzyme-ligand complexes are indispensable. Here docking approaches can help by predicting 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 with the conceptual difference that the ligand is either treated in its entirety or it is built 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 comparing and evaluating the different docking tools for a given enzyme-ligand combination [6-9]. Almost all papers, however, consider protein-based docking of small ligand molecules or protein-protein docking, with the exception of van Dijk et al. who have recently developed a new method for protein-DNA docking [10-12]. Predicting the correct binding between proteins and RNA oligonucleotide chains is of growing interest [13-15], although rarely described in the literature.

Since our main goal is to investigate the hydrolytic mechanism of the above described 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 an optimal docking procedure we have compared four docking programs (MOE2009 [16], Glide 5.5 [17], QXP-Flo+0802 [18], and Autodock 4.0 [19]) to find the most suitable one for our purposes. Here we describe our investigation of possible Ribonuclease-ribonucleotide oligomer complexes by molecular docking. Although calculation of RMSD (root mean square deviation) values between docking results and crystal structures is a well-established method for evaluation of docking poses we rather applied the interaction-based accuracy classification (IBAC) method as described by Kroemer et al. [6] since for each complex the necessary protein-ligand interactions are well described in the literature [6, 20-22]. Therefore, we have determined the correct docking poses by comparing the hydrogen bond (H-bond) distances between enzyme and nucleotide. In this way we have on the one hand identified the docking program which is most suitable for our case of nucleotide docking and on the other hand we could 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*

Crystal structures of both Ribonuclease-ribonucleotide complexes were retrieved from the PDB Database [23]. The protein preparation for both Ribonucleases was carried out by 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 was optimized 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 is also available in the protein database at atomic resolution (PDB code: 2E3W) [25], however, 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 to reproduce this structure for verification.

The starting point for the second set of calculations was the crystal structure of *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 *in silico* mutated to Asp210 in order to reproduce the wild-type protein. Moreover, the optimized complex served as comparison for 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] followed by rebuilding the 3D conformation of each molecule using LigPrep 2.3 [17] at pH of 7.0. Afterwards, the rebuilt ligands were subjected to a conformational search using MacroModel 9.7 [17] with OPLS 2005 force field in a water solvent model. For minimization the “steepest descent” method was applied with a maximum iteration of 500. The energy window for keeping structures was set to 5.02 kcal mol⁻¹ and the RMSD cutoff value 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 as 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. Furthermore, additional ligands were 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 geometry.

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 force field. For solvation the implicit model of “Born” [30] was selected. Two placement methods, “alpha triangle” and “triangle matcher”, were employed to find the optimal docking parameters with the result that the alpha triangle method is faster and yields more suitable docking poses, that

show the required enzyme-ligand interactions. In all cases the scoring was 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 binding pocket [1] while for RNase H the residues Asp145, Glu186, Asp210 and Asp274 defined the pocket [22]. “Rotate bond” of “the Docking Simulation” panel was enabled in both cases.

Glide 5.5. [17] For each protein first a grid box of $30 \times 30 \times 30$ Å with a default inner box ($10 \times 10 \times 10$ Å) was 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, that represent the most important interactions between the enzyme and the ligand. For RNase H only the positional constraint was used, 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 scoring function, and the option of “Dock flexibly” was selected for RNase A and “Dock rigidly” for RNase H in order to keep the original helix-like conformation of the RNA strand. The top 10 poses were kept and 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 catalytic site water oxygen atoms 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 $80 \times 80 \times 80$ points with a spacing of 0.375 Å between the grid points and the center close to the ligand.

Default docking parameters were used and for the search parameters the “Lamarckian Genetic Algorithm (LGA)” was applied [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 energy 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 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 the reference structure 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 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 docking of nucleotides

The goal of our simulations was to produce reliable starting structures for further QM/MM reaction mechanism studies. Although currently available crystal structures of the PDB data base provide fundamental information about the active site 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 developed 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 non-active RNase-ligand complexes. In all docking simulations we started from the DFT/B3LYP QM/MM optimized complexes (reference molecule) of 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 for the evaluation of the software and the scoring function. The resulting structures were analyzed and the docking accuracy was determined by comparison of 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.

<Figure 2>

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

<Table 1>

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.

<Table 2>

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.

For the docking studies of RNase H and the RNA:DNA hybrid duplex (see Materials and Methods section) the ligand was modified by connecting the RNA and DNA chain with an ether bridge to be able to perform the docking with the double helix. However, the modification did not affect the conformation of the hybrid duplex. Since the original ligand has been modified the comparison of the RMSD values for the whole system was possible only when subsequent to docking simulation the ether bridge was removed (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 similarly to the IBAC method of Kroemer et al. [6]

<Figure 3>

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 over-estimated by both softwares. This might be due to the applied flexible docking, where the active site residues were allowed to move, with the consequence that the binding pocket is a little expanded upon force field optimization of the side chain positions.

<Table 3>

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 distances of atoms relevant for the mechanism (Table 3). Although, MOE and Glide are both suitable to reproduce 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 atoms contacts (below 4.5 Å) with a deviation of 0.6 Å or less in comparison with the reference structure.

The RNase A reaction state complex

On the basis of these results we have chosen MOE and Glide to find suitable reactant states for our studies of RNase A and RNase H reaction mechanism. For RNase A the literature

provides a putative pathway [1] of the transphosphorylation 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 transphosphorylation to form a cyclic phosphate stable 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, for the docking simulation dimer, trimer, and tetramer uridyl nucleotide oligomers were used in the presence of crystallographic water molecules.

<Figure 4>

According to the putative reaction mechanism [36] as presented in Fig. 4, the first step involves 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' as well as between (ND1)His119 and O5P are indispensable and the corresponding distances should therefore be as short as possible in correct docking poses.

<Table 4>

<Table 5>

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

<Figure 5>

In order to adjust the enzyme ligand complex to the natural structure and to investigate the interactions far-off 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 of meaningful docking poses, but rather yielded – with both of the two docking tools employed – just one acceptable hit each for each ligand.

<Table 6>

Surprisingly, the derived structure reproduced the natural structure with all 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 shows that as well the short as the wide range interactions between the protein and the timer nucleotide could be reproduced in the docking simulation.

<Figure 6>

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 these interactions are essential for the reaction mechanism to proceed. In addition not only the active site water molecules but further crystallographic water molecules also 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 yields a ligand pose that reproduces all of the important H-bond in the catalytic site as described in the literature [1]. In addition, also the interactions beyond the cleavage site were reproduced correctly. Accordingly, we have used the trimer uridyl docking pose of MOE as the desired reaction state complex as starting structure for further calculation.

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 was 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 for 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 14-mer RNA/DNA hybrid was used as 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 have performed docking calculations to generate the corresponding reactant state complex.

<Figure 7>

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.

<Table 7>

Due to the fact that we used single stranded oligomers in our studies the calculation of RMSD values for the whole system is not feasible. Therefore, we compared the distances 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 for Glide but the resulting top hits are comparable as can be judged from the analyzed interactions.

<Figure 8>

Both structures turn out to be very close to the reference structure (Fig. 8). This strongly supports that replacement of the mutated Asn210 to Asp210 in order to construct the active enzyme did not affect the geometry of the catalytic site. As can be seen on Fig. 8 the nucleotides which are in the deeper regions of the pocket fit excellent, however the nucleotide base at the end of the pocket turns rather towards the exit of the pocket due to the fact that in that region there are less interactions with the protein.

Conclusions

With the aim of predicting reactant state structures for reaction mechanism calculation of RNase A and RNase H hydrolysis, we have evaluated four protein-based docking tools with respect to their capability for enzyme-nucleotide docking. To this end different ribonucleotide oligomers were placed into the active sites of RNase A and RNase H. As molecular docking tools were developed originally for placing 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 ability to reproduce non-active 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 the interactions beyond the active site. Since our initial goal was to find suitable active reactant state complexes for our studies of RNase A and RNase H reaction mechanism we have, therefore, chosen MOE and Glide to accomplish our project.

In the docking studies of the mutated RNase H and a modified tetramer RNA:DNA hybrid duplex we could 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 supports further 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 well be used for nucleotide docking and – at least in the special cases we have investigated - can generate reactant state complexes for further theoretical studies. Our simulations with RNase A identify a few appropriate docking poses, which 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 on the basis of the docked reactant state structure as 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 bases.

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 SF (1989) *J Febs Letters* 257:311-314
5. Tanese N, Goff SP (1988) *PNAS* 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, Bisailon 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. Macromodel version 9.7, Glide version 5.5, Maestro version 9.0, LigPrep version 2.3 (2009) Schrödinger LLC, New York, NY
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. Autodock 40
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 Genetics* 17:337-346
22. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W (2007) *Mol Cell* 28:513-513

23. RCSB Protein Data Bank
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 and 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 V3.1
<http://www.bioinf.org.uk/software/profit/>
34. McLachlan AD (1982) *Acta Crystallogr Sect 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, New York
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

Tables

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

	hit1	hit2	hit3	hit4	hit5
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

	hit1	hit2	hit3	hit4	hit5
MOE	0.392	0.327	0.511	0.494	
Glide	0.860	0.792	0.910	0.869	0.974

Table 3 Comparison of distances (in Å) between docking results (best hit) and the corresponding experimental structure of 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 into the center of the five-membered ring

	reference structure	MOE		Glide	
		dist	diff.	dist	diff.
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
<i>pro</i> 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 nucleotide 1	-		0.37		1.85
centroid of sugar ring nucleotide 2	-		0.68		2.18

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

MOE	hit1	hit2	hit3	hit4	hit5	hit7
(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

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

Glide	hit1	hit5	hit8
(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

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. Top hit is marked in blue

	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

Table 7 Comparison of relevant protein-ligand interactions between docking results (best hit) and the reference structure. *proO1P* 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
<i>proO1P</i> -(OD1)Asp274	3.83	3.86	0.03	4.03	0.20	4.24	0.41	3.96	0.13

Figure captions

- Fig. 1** Cleavage mechanism of the two Ribonuclease enzymes, RNase A and RNase H
- Fig. 2** Superposition of the top binding poses from bound docking with MOE. Reproduced complexes of RNase A and cyclic uridyl-phosphorane RMSD=0.097 (Fig. 2a), and deoxy cytosyl-adenyl-dinucleotide RMSD=0.327 (Fig. 2b). The original pose is colored green and the docking hit is red. The displayed surface illustrates the catalytic pocket, in which purple color identifies H-bond interactions, green color characterizes areas of hydrophobic contacts, and blue color denotes dipole-dipole interactions
- 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
- Fig. 4** Putative reaction mechanism of the first step of the RNase A hydrolysis
- Fig. 5** Best docking pose from MOE-docking of dimeric uridyl nucleotide into RNase A active site. The displayed surface was generated by MOE and illustrates the catalytic pocket. Color codes are as specified in Fig. 2
- Fig. 6** Top hit docking pose 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
- Fig. 7** Proposed reaction pathways [1] for the O3'-P cleavage of *human* RNase H
- Fig. 8** Superposition of the reference structure (green) and the top tetramer nucleotide docking hit (red). Color codes are as specified in Fig. 2

















