

# Atomistic details of the molecular recognition of DNA-RNA hybrid duplex by ribonuclease H enzyme

GORLE SURESH and U DEVA PRIYAKUMAR\*

Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500 032, India  
e-mail: deva@iiit.ac.in

MS received 27 April 2015; revised 22 June 2015; accepted 2 July 2015

**Abstract.** *Bacillus halodurans* (Bh) ribonuclease H (RNase H) belongs to the nucleotidyl-transferase (NT) superfamily and is a prototypical member of a large family of enzymes that use two-metal ion ( $Mg^{2+}$  or  $Mn^{2+}$ ) catalysis to cleave nucleic acids. Long timescale molecular dynamics simulations have been performed on the BhRNase H-DNA-RNA hybrid complex and the respective monomers to understand the recognition mechanism, conformational preorganization, active site dynamics and energetics involved in the complex formation. Several structural and energetic analyses were performed and significant structural changes are observed in enzyme and hybrid duplex during complex formation. Hybrid molecule binding to RNase H enzyme leads to conformational changes in the DNA strand. The ability of the DNA strand in the hybrid duplex to sample conformations corresponding to typical A- and B-type nucleic acids and the characteristic minor groove width-seem to be crucial for efficient binding. Sugar moieties in certain positions interacting with the protein structure undergo notable conformational transitions. The water coordination and arrangement around the metal ions in active site region are quite stable, suggesting their important role in enzymatic catalysis. Details of key interactions found at the interface of enzyme-nucleic acid complex that are responsible for its stability are discussed.

**Keywords.** HIV-1 reverse transcriptase; MD simulations; ribonuclease H activity; retroviral therapy; conformational transition; protein-nucleic acid interactions.

## 1. Introduction

Ribonucleases are enzymes that cleave phosphodiester bonds of nucleic acid backbones using divalent cations (preferably  $Mg^{2+}$  or  $Mn^{2+}$ ) as cofactors and yield 5'-phosphate and 3'-hydroxyl terminals.<sup>1,2</sup> It was believed that these enzymes have played important roles in the transition from RNA world to DNA world.<sup>3</sup> These enzymes were also implicated in removing 5'-ends in lagging strand synthesis in DNA replication,<sup>3</sup> DNA repair, and RNA transcription.<sup>4</sup> Ribonucleases are present in almost all kingdoms<sup>5</sup> and exhibit essential role in the life cycle of retroviruses and LTR retrotransposons by performing three different functions. Several ribonucleases were identified and grouped into different categories (mainly type I & II) depending on the substrate recognition and their structure.<sup>6,7</sup> *E.coli* RNase HI has been widely studied among all the known ribonucleases by using X-ray,<sup>8–10</sup>

site-directed mutagenesis<sup>11,12</sup> and NMR studies.<sup>13,14</sup> The crystal structure revealed a  $\alpha/\beta$  fold containing four carboxylate residues and two  $Mg^{2+}$  ions, located at specific positions in the catalytic site. Similar fold has been found in RNase HII,<sup>15</sup> Holliday junction resolvase RuvC,<sup>16,17</sup> DNA transposases and retroviral integrases<sup>18</sup> and Argonautes.<sup>19</sup>

The current study focuses on a specific enzyme ribonuclease H (also known as RNase H), which belongs to the nucleotidyl-transferase (NT) superfamily. These enzymes are capable of discriminating DNA-RNA hybrid duplexes from other duplexes such as single and double stranded DNA and RNA, and cleave the RNA strand of the hybrid duplex by hydrolysing the phosphodiester bonds without affecting the complementary DNA strand in an endonucleotic manner.<sup>2</sup> DNA-RNA hybrid duplex is a heterogeneous nucleic acid in which one DNA strand compliments with another RNA strand and exhibit distinct structural and dynamic features compared to pure duplexes.<sup>20,21</sup> RNase H is a metalloenzyme, comprised of a  $\alpha/\beta$  fold with conserved active site carboxylate motif (DDE

\*For correspondence

motif) which is essential for metal binding. The enzymatic activity depends on metal ion concentration and nature of the ions.<sup>2</sup> Crystal structures of the complexes of ribonuclease H and DNA-RNA hybrid have been reported by Nowtony *et al.* for human RNase H<sup>22</sup> and *Bacillus halodurans* (*Bh*) RNase H<sup>2</sup> families. Examination of these crystal structures has indicated that these two active sites have similar catalytic domains and these highly conserved residues are required for the enzymatic activity. The catalytic action occurs via a two metal ion mechanism, proposed by Steitz *et al.*<sup>23</sup> Recent QM/MM studies of De Vivo *et al.* on RNase H – DNA-RNA hybrid complex<sup>24</sup> suggested a two-step mechanism for this enzymatic activity supporting the previous studies.<sup>25,26</sup> The crystal structure of RNase H-DNA/RNA hybrid complex from *Bh* indicated that the enzyme recognizes the DNA-RNA hybrid duplex by interacting with 2'-OH groups of the RNA strand. The primary difference between the crystal structures of ribonuclease H bound to hybrid duplex from human and *Bh* is the “basic protrusion”. This might make a difference in the substrate recognition mechanism of these ribonuclease enzymes. Deletion of this basic protrusion resulted in decreased enzymatic affinity towards DNA-RNA hybrid duplex in *E.coli*.<sup>27</sup> These active site residues are conserved in most of the ribonucleases. *In vivo* studies have shown that the inactivation of RNase H results in several diseases indicating the essential role played by these enzymes on survival of some of organisms. The studies performed on mice embryo have showed that the lack of RNase H resulted in malfunction in their DNA synthesis which led to death.<sup>28</sup> Furthermore, these enzymes play an essential role in HIV-1 reverse transcription process.<sup>11,29</sup> Understanding the structural and energetic relationships among protein and DNA-RNA hybrid duplex, and the activity between them is crucial for designing new organic molecules which can inhibit the enzymatic activity, and subsequently help in controlling the HIV-1 reverse transcription process.

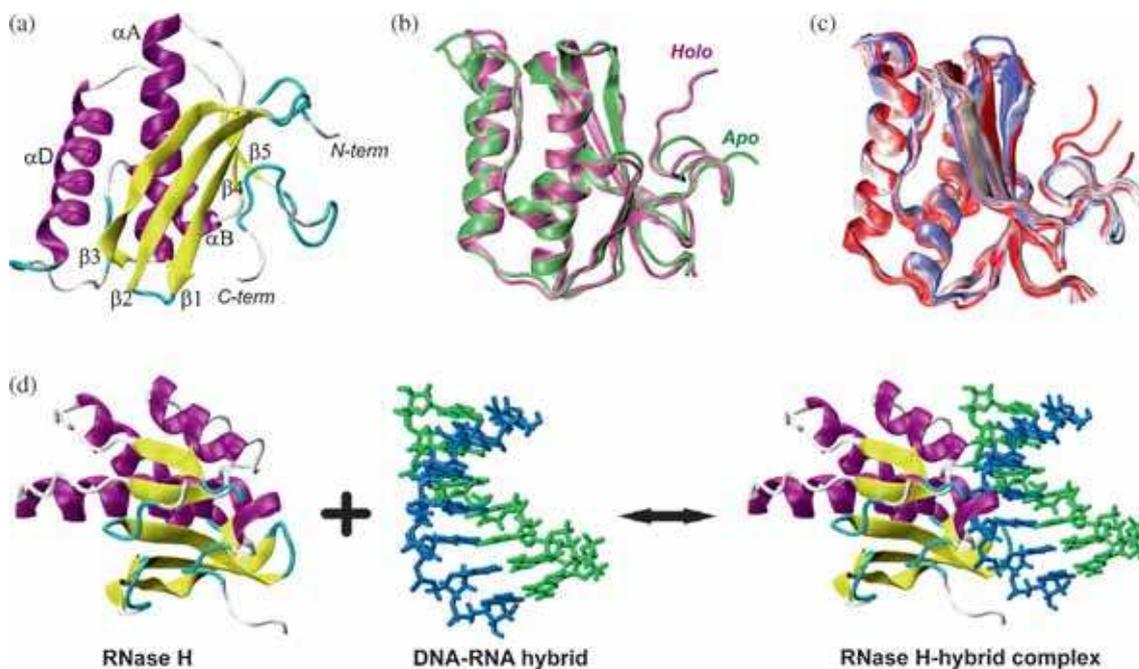
Molecular interactions between protein and nucleic acids play crucial role in regulating biological processes.<sup>30</sup> One of the important goals of modern molecular biology and biophysics is to understand these crucial interactions between protein and nucleic acids, and the associated conformational changes. Of particular interest is investigating how a small change in sequence affects protein-nucleic acid binding. Understanding these aspects is helpful in designing small molecules with desired properties which are capable of controlling cellular processes. Several experimental techniques like X-ray, NMR and site-directed mutagenesis are instrumental in studying protein-nucleic acid

interactions.<sup>31–35</sup> Molecular modelling and molecular dynamics (MD) simulations, along with the available experimental data, are extensively used to understand these interactions in atomistic detail.<sup>36–41</sup> In this study, MD simulations have been performed on the RNase H enzyme, DNA-RNA hybrid duplex, and their complex from *Bh* in explicit solvent environment to understand the associated conformational changes in both the monomers during the formation of the complex and the stability of the key interactions involved in complex formation.

## 2. Materials and Methods

### 2.1 Model systems

The model systems used in the present study were conceived based on the X-ray crystallographic data of *Bh* ribonuclease H complexed with a DNA-RNA hybrid (PDB ID:1ZBI).<sup>2</sup> This complex structure is preferred over other available ribonuclease H-DNA-RNA hybrid complexes because of its better resolution<sup>42</sup> and this has been extensively studied using experiments.<sup>2,22</sup> Additionally, the RNase H from *Bh* and HIV-1 are structurally similar, and share the conserved DDE motif. The original experimental structure contains two RNase H proteins complexed with a dodecamer of DNA-RNA hybrid duplex in which each protein molecule is bound to the substrates without any protein-protein contact. Each protein is 135 amino acid long, arranged as three  $\alpha$ -helices (named as  $\alpha$ A,  $\alpha$ B, and  $\alpha$ D) and a mixed  $\beta$ -sheet containing five  $\beta$ -strands, in which three  $\beta$ -strands are arranged in antiparallel and next two are in parallel arrangement (figure 1A). The  $\alpha$ A runs in between  $\beta$ 3- $\beta$ 4 strands and  $\alpha$ B and  $\alpha$ D runs in between  $\beta$ 4- $\beta$ 5 strands which attain a  $\alpha\beta\alpha$  Rossmann-like fold. Apart from the protein and DNA-RNA hybrid duplex, two Mg<sup>2+</sup> ions and 240 water molecules are present in the experimental crystal structure. The active site region contains three aspartic acids (D10, D71 and D131) and two glutamic acids (E48 and E127) along with two Mg<sup>2+</sup> ions and water molecules. One of the protein chains and five base pairs in DNA-RNA hybrid duplex were removed from the original experimental structures.<sup>2,43</sup> Since the binding of protein to hybrid is non-sequence specific and the two halves of the complex are exactly identical, the current model system (one protein + 7 bp DNA-RNA hybrid) is deemed adequate. The monomer systems were isolated from the ribonuclease H-DNA-RNA hybrid complex by removing the respective monomer. These three model systems represent two different forms, **apo** (free) and



**Figure 1.** (a) Three dimensional representation of secondary structural elements in RNase H enzyme; (b) Alignment of holo (pink) and apo (green) RNase H enzymes; (c) Structural evolution of apo RNase H enzyme during first 20 ns simulation time from initial conformation (red) to conformation at 20 ns (gradient blue); and (d) Schematic representation of formation of complex between RNase H and DNA-RNA hybrid.

**holo** (bound), of the protein and the DNA-RNA hybrid duplex.

## 2.2 MD simulation protocol

Initial preparation of the systems, equilibration and trajectory analysis were done using the CHARMM biomolecular simulation program<sup>44</sup> following the protocols that are described elsewhere.<sup>21,41,45,46</sup> Briefly, the systems were subjected to a 500-step steepest decent (SD) minimization with harmonic restraints on heavy atoms of the protein and DNA-RNA hybrid together with the crystal  $Mg^{2+}$  and water molecules. These minimized systems were then immersed in a pre-equilibrated water box and were made electrically neutral by adding sufficient number of ions. Modified TIP3P water model<sup>47</sup> for water and SHAKE algorithm for constraining the covalent bonds involving hydrogens were used.<sup>48</sup> Then these systems were subjected to each of 500 step SD and adopted basis Newton-Rapson (ABNR) minimizations with mass weighted harmonic restraints with force constant  $4 \text{ kcalmol}^{-1} \text{ \AA}^{-2}$  on non-hydrogen atoms and crystal ions. These systems with the same constraints were subjected to 100 ps MD simulations in NVT ensemble for equilibration. Periodic boundary conditions were used<sup>49</sup> and particle mesh Ewald (PME) summation method<sup>50,51</sup> was employed to treat the long range electrostatic interactions. The usage of a force smooth switch function from

$10 \text{ \AA}$  to  $12 \text{ \AA}$  allowed to truncate Lennard-Jones (LJ) interactions at  $12 \text{ \AA}$ .<sup>52</sup> The structures obtained from the equilibration step were used as starting configurations for running production simulations. Production simulations were performed using the NAMD program<sup>53</sup> by employing the CHARMM27 all atom nucleic acid force field for nucleic acids<sup>54,55</sup> and CHARMM22 all atom protein force field with CMAP corrections for proteins.<sup>56,57</sup> A weak harmonic constraint was applied on the central hydrogen bond of the terminal base pairs of the DNA-RNA hybrid duplex using a force constant of  $4 \text{ kcalmol}^{-1} \text{ \AA}^{-2}$  to prevent their opening.<sup>46,58</sup> All the simulations were performed in NPT ensemble and Leapfrog integrator was employed to integrate Newton's equations of motion. Langevin piston for pressure<sup>59</sup> and Hoover thermostat for temperature were used.<sup>60</sup> MD simulations were performed up to 100 ns by saving the coordinates for every 5 ps and used for further analysis.

## 2.3 Analysis of the trajectories

Extensive analyses on the resultant trajectories were performed to understand the structural differences and energetic factors responsible for the conformational rearrangements in both protein and DNA-RNA hybrid duplex during binding process, and the factors responsible for the stability of complex of ribonuclease H

with DNA-RNA hybrid. Final 80 ns simulation trajectories were used for analysis by treating first 20 ns as the equilibration period. The root mean square deviations (RMSD) were calculated with respect to the initial conformation and the atomic fluctuations were calculated for heavy atoms of the solute by aligning all the conformations on their average conformation obtained from simulations. Root mean square fluctuations (RMSF) corresponding to the experimental temperature factor (B-factors) were calculated using the following relation:<sup>61</sup>

$$B = \frac{8\pi^2}{3} RMSF^2 \quad (1)$$

The interaction energies were calculated as the sum of electrostatic and LJ terms which were calculated using the "INTER" command in CHARMM, with an infinite non-bonded cut-off. Visual molecular dynamics (VMD)<sup>62</sup> was used for viewing trajectories and rendering images depicted in the manuscript. The helical parameters and groove widths of the DNA-RNA hybrid duplex were computed by using the Curves+ program.<sup>63</sup>

### 3. Results and Discussion

#### 3.1 Structures and dynamics of apo state of the RNase H enzyme and DNA-RNA hybrid

To assess the structural changes during the simulations and to quantify the convergence of the simulations, RMSD were calculated with respect to their respective starting structures. Average values corresponding to the protein, hybrid and individual strands in both apo and holo forms were calculated and are given in table 1. The time series of RMSD indicate that the structures have attained equilibrium conformation within short time interval (figure S1A-C, figure 1C). The alignment of average structures of proteins showed that the protein maintains its overall fold upto a large extent upon binding (figure 1B). The details of the significant structural differences of RNase H in the apo and holo forms were examined by calculating RMSD values for various regions of the protein. The values suggest considerable changes in the orientation in  $\beta$ 1 strand, and amino acids in regions 83–90, and 126–135. It is also observed that the large conformational changes in active site region of protein are highly restricted after binding (figure S1C). Most of the non-active site residues of the holo protein deviate much more than active site residues, indicating the highly rigid nature of the active site, which is in agreement with previous studies.<sup>64</sup> This

kind of conformational preorganization of the active site was also observed in recent MD simulation studies on ribonuclease HI.<sup>64,65</sup> The large deviations observed in protein secondary structure were also reflected in the change in number of residues in secondary structural elements. The number of residues present in  $\alpha$ -helices and  $\beta$ -strands of protein varies from free form to bound form. Most variations were observed in the number of residues in apo protein especially in helices i.e., the  $\alpha$ -helices largely unfold and fluctuate from helix to coil (table 1). When the substrate binds to the protein, the number of residues present in helices vary from 42 to 43 which is similar to that in the experimental structure (44) but there is a large variation from 20 in free form to 23 in holo form in case of  $\beta$ -strands where the experimental value is 27.

RMSF values have been calculated with respect to the average structure obtained from the final 80 ns of the simulations. The protein fluctuates similarly irrespective of the presence and absence of the hybrid (figure 2 and table 1). The atomic fluctuations corresponding to the experimental structure using B-factors were also computed and presented along with the fluctuations obtained from MD simulations, and the two are in good qualitative agreement with each other. Notable differences are in reduction of fluctuations of the  $\beta$ 1 strand, turn1, amino acids 123–124 and increment in fluctuations of the amino acid region 28–30 upon binding of the hybrid duplex. The reduction in atomic fluctuations is caused by the interaction of DNA-RNA hybrid duplex with the amino acids present in these regions. This indicates that the  $\beta$ 1 strand of the protein interacts with the DNA-RNA hybrid in a similar manner as observed in the crystal structure of the complex of RNase H2 with RNA-DNA chimera.<sup>66</sup>

The solvent accessible surface areas (SASA) of protein secondary structural elements were calculated when the protein is in free and bound forms and the differences between the apo and holo forms are shown in figure 3. While positive values represent buried surface area, negative values represent exposed surface area upon binding. As shown in figure 3, the residues which are present in the regions loop1, loop 4, and  $\alpha$ D are exposed to the solvent during binding whereas the residues present in  $\beta$ 1 strand, loop 2 and  $\alpha$ B are buried. It follows that the residues that are buried lie in the protein-hybrid interface as expected. The residues that are identified to be present in the interface are in good agreement with the experimental structure of RNase H2 -DNA-RNA chimera complex, which suggests that the protein binds the nucleic acids through  $\beta$ 1 strand, and the helices  $\alpha$ A, and  $\alpha$ B.<sup>66</sup> The buried secondary structural elements that are identified here exhibit increased

**Table 1.** Average values of root mean square deviations<sup>a</sup>, fluctuations and radius of gyration calculated for RNase H enzyme and DNA-RNA hybrid duplex in apo and holo states; average numbers of residues present in secondary structural elements of RNase H enzyme were also calculated.

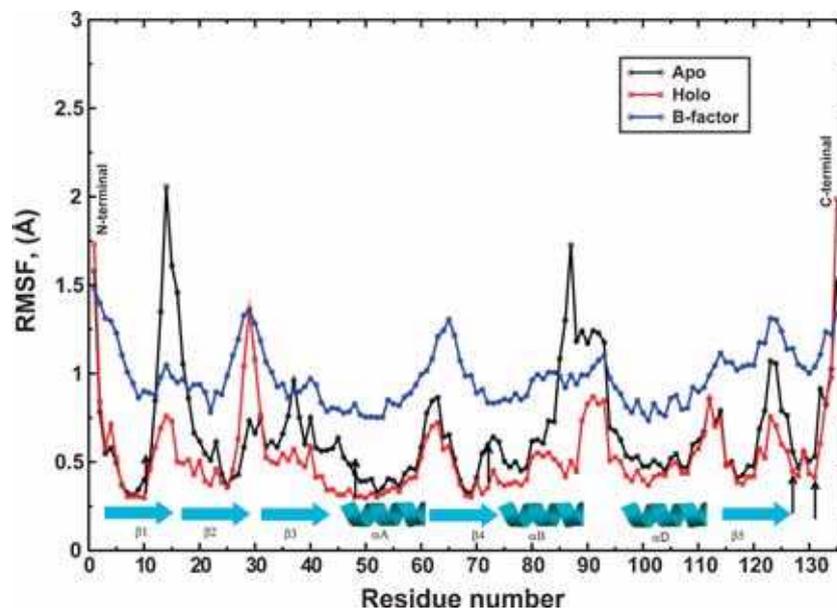
Moiety	Apo	Holo
	RMSD (Å)	
Complex		2.09 ± 0.01
Protein	3.35 ± 0.02	2.14 ± 0.01
Protein-I <sup>b</sup>		1.89 ± 0.01
Hybrid	1.44 ± 0.01	1.64 ± 0.02
Hybrid-bases	1.25 ± 0.01	1.46 ± 0.03
Hybrid-backbone	1.78 ± 0.01	2.00 ± 0.01
DNA strand	1.43 ± 0.01	1.32 ± 0.02
RNA strand	1.44 ± 0.01	1.90 ± 0.02
Active site region	1.81 ± 0.01	0.72 ± 0.00
Non-active site region	3.33 ± 0.04	2.14 ± 0.01
	RMSF (Å)	
Protein	0.67	0.54
Hybrids	1.05	0.68
Hybrid-bases	0.86	0.64
Hybrid-backbone	1.19	0.71
DNA strand	0.78	0.68
DNA strand-bases	0.69	0.62
DNA strand-backbone	0.87	0.73
RNA strand	1.16	0.65
RNA strand-bases	0.93	0.66
RNA strand-backbone	1.33	0.64
	RGYR (Å)	
Complex		15.55 ± 0.00
Protein	14.90 ± 0.01	14.19 ± 0.00
Hybrid	10.04 ± 0.00	10.07 ± 0.00
	Secondary structure elements	
Protein	$\alpha$ -Helix	$\beta$ -strands
Crystal	44	27
Apo	41.8 ± 1.5	20.7 ± 0.8
Holo	43.7 ± 1.6	23.7 ± 0.9

<sup>a</sup> The definitions for different regions are shown in table S1. <sup>b</sup>The RMSD corresponding to protein-I was calculated by aligning only protein in the complex with its initial structure.

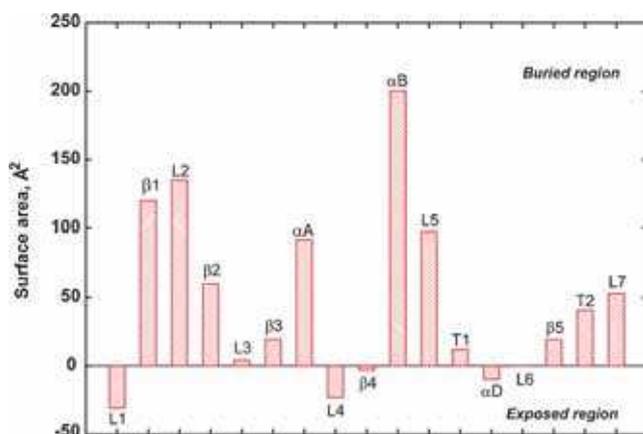
atomic fluctuations of these regions in the free form (figure 2).

In case of DNA-RNA hybrid duplex, the small deviations observed for hybrid in free and bound forms indicate their closeness to initial structure (table 1). The close average RMSD values for hybrid in free (1.44 Å) and bound (1.64 Å) forms indicate minor structural changes in the hybrid upon binding to the protein. RMSD contributions from individual strands, bases, and backbone atoms were also calculated to further understand the conformational changes in detail. These indicate large conformational change in RNA strand of the hybrid during the binding suggesting the fact that the enzyme binds the hybrid by contacting the backbone of the RNA strand. Further, the calculated atomic

fluctuations indicate that the enzyme binding reduces the fluctuations of the hybrid duplex suggesting that the enzyme binds strongly to the hybrid (table 1). This reduction in duplex fluctuations arises from the reduction in fluctuations of the bases and backbone. The contributions from individual strands, backbone, and bases to fluctuations show that the reduction in hybrid fluctuations arises due to the decreased fluctuation of the RNA strand's backbone. This reduction in fluctuations of RNA indicates that the protein binds to the DNA-RNA hybrid by attacking its RNA strand backbone. Combination of these results suggests that the hybrid fluctuations those of especially the RNA strand which are important for hydrolysis are reduced upon binding to the protein.



**Figure 2.** The root mean square fluctuations (RMSF) (in Å) of the apo and holo RNase H enzymes. The RMSF values obtained from experimental B-factors were also included.



**Figure 3.** Buried and exposed surface regions of ribonuclease H enzyme during binding to nucleic acid duplex.

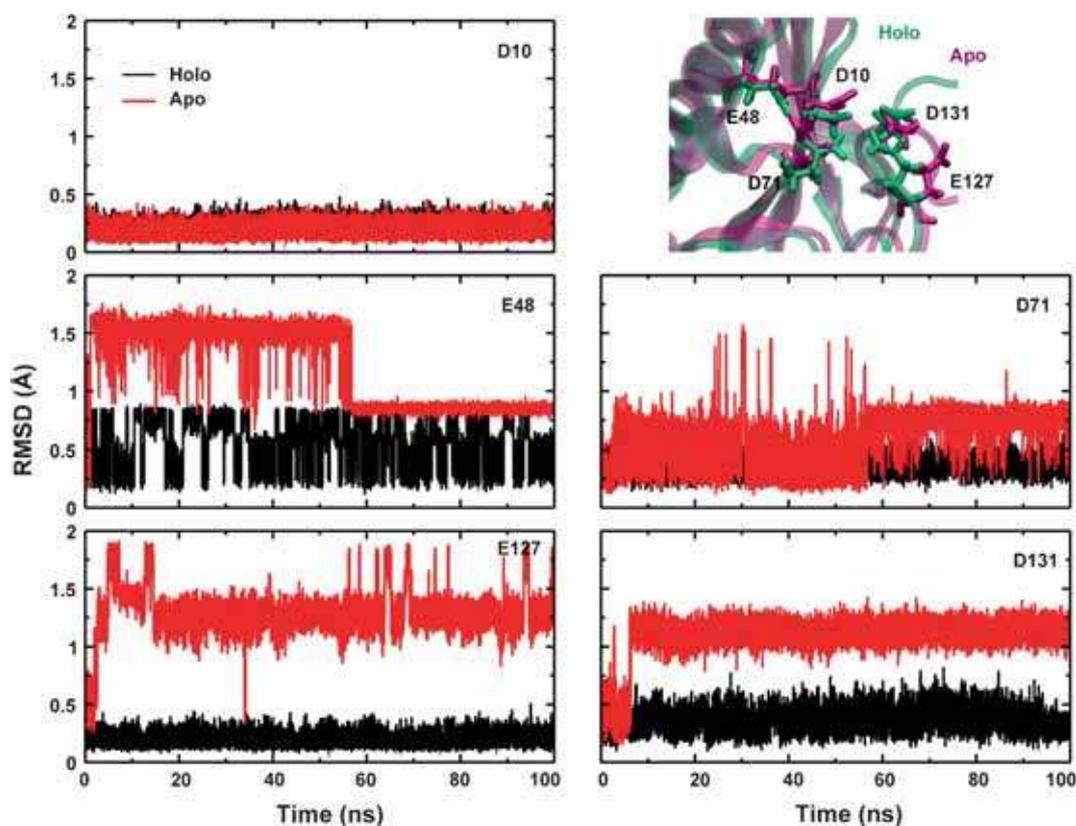
The average value of radius of gyration (RGYR) for protein in free form is higher than its bound form, indicating that the protein is more compact in the holo form than the apo form further confirming significant structural differences between the two (table 1). However, the differences between the RGYR values of the hybrid duplex in the apo and holo forms are similar, further confirming retention of the integrity of the overall structure.

The time series of RMSD show that the complex reaches a particular conformational state in a small time interval and stays there for the remainder of the simulation time of 100 ns (figure S1A). The average RMSD value of the complex is 2.08 Å which is comparable to the resolution of original crystal structure 1.85 Å

used to model the starting structure, and significantly represents small deviations from the initial conformation during the MD simulations (table 1). The major deviations in complex arise from the rearrangement of enzyme in complex. To further understand the compactness of the complex, radius of gyration analysis has been done on the complex. The time series of radius of gyration of enzyme-hybrid complex indicate that the complex is pretty stable during the long simulations by allowing small structural deviations and maintenance of its shape and structure, supporting RMSD values (figure S1D).

### 3.2 Activesite dynamics

Previous studies on ribonuclease HI have shown that the active site preorganizes before binding to the substrate.<sup>64</sup> Understanding the factors that are responsible for this preorganization has been useful in studying the catalytic mechanism. The time series of the RMSD corresponding to the individual residues in active site showed that the substrate binding does not affect residue D10 dynamics (figure 4). Major changes are observed for E127 and D131 whose RMSD converges after few nanoseconds of the simulations. However, the behaviour exhibited by E48 and D71 residues is different. Further, the residues are structurally more flexible in free form and become rigid after substrate binding, indicating interaction of the protein with the substrate (figure S2). Briefly, the results suggest that the active site is preorganized for binding to a larger extent



**Figure 4.** Root mean square deviations of residues (in Å) corresponding to the active site region of RNase H enzyme in holo (black) and apo (red) states.

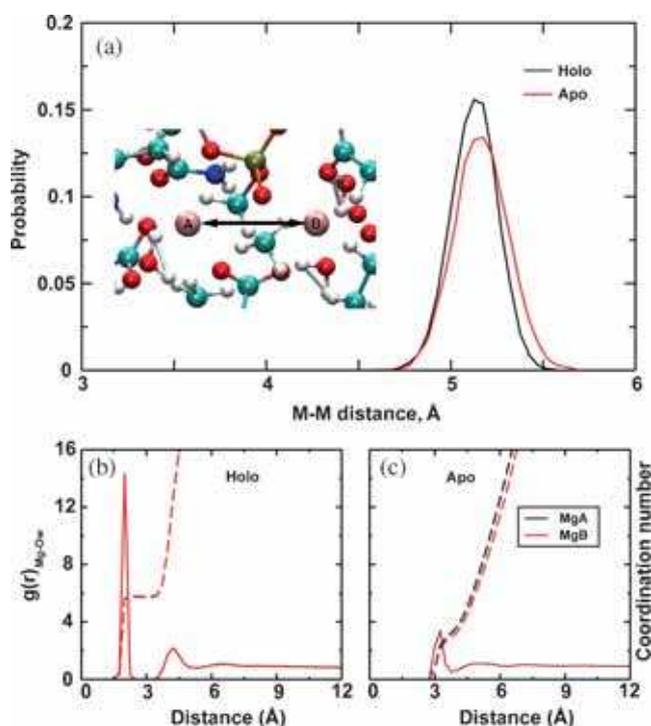
in the apo state albeit with higher flexibility. Upon binding such higher flexibility decreases to form a more rigid structure.

Experiments have shown that the water coordination around  $Mg^{2+}$  ions and their separation distance play a key role during the catalytic mechanism.<sup>2,22</sup> The distance between two metal ions was calculated (figure 5A) and it was almost constant throughout the simulations. The observed average metal–metal distance during the simulations is longer than 5 Å which is slightly higher than experimental value of 4.5 Å.<sup>2,22</sup> The longer metal–metal distances have also been observed in previous MD simulation study on mutant enzyme–hybrid complexes.<sup>24</sup> Further, the water arrangement and its geometric structure around the metal ions were characterized by calculating the radial distribution functions,  $g(r)$ , and coordination numbers, and compared with their corresponding free forms. The highest peak at 2.7 Å indicates the strong binding of water molecules around the metal ions (figure 5B–C). The lowest peak for ions in free form shows their weak interactions with the water molecules i.e., rapid exchange of water. These plots also infer that the substrate binding restricts the rapid exchange of water around metals which makes the system stable. Furthermore, it provides a perfect water

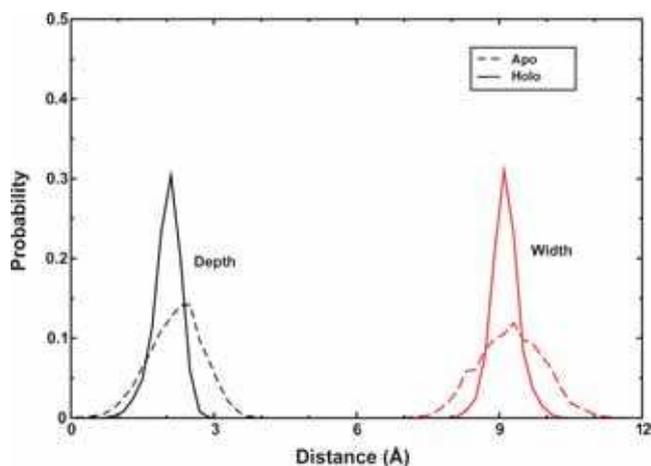
arrangement around the two metal ions, supporting the experimental studies. The average coordination number observed for both metal ions is  $\sim 5.8$  which is close to the experimental values.<sup>2,22</sup> The observed water geometry around the metal ions was very stable during the simulations, indicating their importance in the stability of enzyme–hybrid complex and catalysis.

### 3.3 Helical integrity

To examine the impact of protein binding on the structural integrity of the hybrid duplex, analyses such as the helical parameters, groove widths, hydrogen bond distances, base pair interaction energies, and base stacking interaction energies were performed. Probability distributions of hydrogen bond distances corresponding to Watson–Crick base pairs in DNA–RNA hybrid duplex indicate that the enzyme binding does not affect base pairing (figure S3). The enzyme binding reduces the base stacking within the strand but increases stacking with the bases in the complementary strand showing overall marginal decrease in interactions among base pairs (table S3). However, the protein binding narrows the width and depths of the minor groove



**Figure 5.** (a) Probability distribution of distance between two metal ions present in active site of apo and holo RNase H enzymes. Radial distribution functions and corresponding coordination numbers of two metal ions attached with the enzyme active site in holo; (b) and apo; (c) forms.



**Figure 6.** Probability distributions of depth and width regions of minor groove of DNA-RNA hybrid duplex in apo and holo forms.

(figure 6). This may be due to the interaction of enzyme with the minor groove region of the hybrid similar to the complex of RNase H2 and DNA-RNA chimera where the protein interacts with the chimera through its minor groove<sup>66</sup> such a study is expected to impart high rigidity.

Helical parameters corresponding to base pairs and base pair steps were calculated to obtain more details

of the changes induced in hybrid duplex upon binding. As shown in table 2, the protein binding reduces the average deformations of hybrid from a propeller of  $-9.29^\circ$  to  $-8.33^\circ$ . The free hybrid shows large positive inclination value of  $16.62^\circ$  to the helical axis and is shifted towards the major groove ( $X\text{-disp} = -2.97$ ). The protein binding increases in the inclination value to  $21.52^\circ$  and is shifted even towards the major groove ( $X\text{-disp} = -3.51$ ). During protein binding, the rise does not change much but decreases the twist to  $28.83^\circ$ . The total bend of hybrid increases from  $9.34^\circ$  to  $10.22^\circ$ , and the roll increases from  $8.92^\circ$  to  $10.51^\circ$ . Enzyme binding reduces the interactions among base pairs (stacking interactions) observed by the variations seen in the translational and rotational parameters of individual base pair steps (figure S4). The enzyme binding affects the helical parameters of base pairs and base pair steps involving dTrA base pair. Several base pair and axis parameters such as X-displacement, inclination, tip, base opening, buckle, stretch and propeller twist are different for dTrA base pair at which the enzyme binds directly to the nucleic acid (figure S5). These observed variations suggest that the enzyme binding imparts changes to the hybrid structure at a local level rather than global level.

### 3.4 Desolvation around the hybrid duplex

Solvent plays a significant role in the binding of protein and nucleic acid. Hydration numbers and SASA values have been calculated to examine the desolvation effects around the helix, especially individual base pairs. The calculated hydration numbers around the major and minor groove regions of each base pair indicate that the enzyme binding increases the extent of hydration around the major groove, and significantly decreases the hydration number around the minor groove (table S4). Decrease in hydration numbers were observed around the major groove regions adjacent to the base pairs that are interacting with enzyme. It is also observed that the protein binding dehydrates the RNA backbone especially at A4, U5, U6, and C7 nucleotides. These values indicate that enzyme interacts with the hybrid through minor groove of base pairs. The surface area accessible to the water is calculated around the minor, major and phosphate oxygen atoms of all the base pairs using a probe with radius  $1.4 \text{ \AA}$ . The obtained values indicate that the surface area accessible to the solvent is reduced around the minor groove regions of dTrA and dCrG base pairs after interacting with the enzyme (table S4). These suggest that the enzyme strongly interact with these two base pairs.

**Table 2.** Average values of helical parameters for DNA-RNA hybrid duplex in apo and holo states.

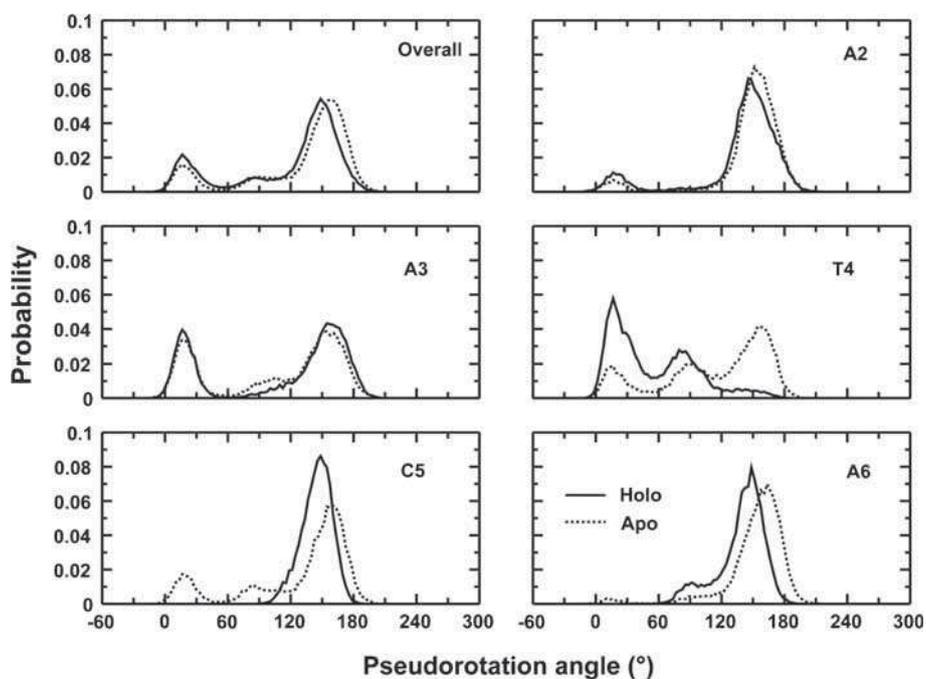
Parameter	Apo	Holo
X-displacement	$-2.97 \pm 0.00$	$-3.51 \pm 0.00$
Y-displacement	$-0.06 \pm 0.00$	$-0.34 \pm 0.00$
Tip	$7.17 \pm 0.02$	$6.82 \pm 0.02$
Inclination	$16.62 \pm 0.03$	$21.52 \pm 0.02$
Propeller twist	$-9.29 \pm 0.03$	$-8.33 \pm 0.03$
Shear	$-0.04 \pm 0.00$	$0.08 \pm 0.00$
Stretch	$-0.03 \pm 0.00$	$-0.07 \pm 0.00$
Stagger	$-0.26 \pm 0.00$	$-0.39 \pm 0.00$
Buckle	$5.67 \pm 0.04$	$5.47 \pm 0.04$
Opening	$2.99 \pm 0.02$	$3.05 \pm 0.02$
Shift	$-0.44 \pm 0.00$	$-0.36 \pm 0.00$
Slide	$-0.76 \pm 0.00$	$-0.87 \pm 0.00$
Rise	$3.40 \pm 0.00$	$3.46 \pm 0.00$
Tilt	$-4.17 \pm 0.01$	$-3.98 \pm 0.01$
Twist	$30.29 \pm 0.01$	$28.83 \pm 0.01$
Roll	$8.92 \pm 0.01$	$10.51 \pm 0.01$
Total bend	$9.34 \pm 0.04$	$10.22 \pm 0.04$

### 3.5 Sugar and backbone conformation of hybrid duplex

The backbone conformations of the hybrid duplexes play an important role in substrate recognition and enzyme catalysis.<sup>66</sup> Previous studies have shown that the DNA-RNA hybrid exists in mixed A/B-type conformation. Furanose sugars in RNA strand sample C3'-endo conformations similar to pure RNA whereas the furanose sugars in DNA strand show dynamic transition between C3'-endo and C3'-exo.<sup>21,46,58</sup> The evaluation of pseudorotation angles of individual nucleotides along the simulation time indicate that the furanose sugars of DNA and RNA strands in apo hybrid sample conformations similar to pure DNA and RNA respectively. Probability distributions of the pseudorotation angles of furanose sugar puckering of the bound hybrid show sampling of mixed conformations for RNA and DNA strands (figure 7). The ribose sugars in RNA strand mostly samples in the C3'-endo and C2'-exo regions similar to the apo hybrid supporting the previous experimental results.<sup>21</sup> The enzyme binding induces transition in conformation of the deoxyribose sugars from C3'-exo to O4'-endo and C3'-endo regions. These conformational transitions are high in case of thymine (T4) and cytosine (C5) where the enzyme makes direct contacts with the oxygen atoms of the nucleobases (figure 8). In general, the complementary strand to the strand binding to enzyme has to reorient in order to make the binding process more feasible which results in stabilization of the entire system.<sup>66</sup> The experimental data indicated that the number of amino acids

interacting with the DNA and RNA strands is same but the major conformational changes are observed in deoxyribose sugars only, and not in the ribose sugars. Comparison of individual phase angles illustrates that the deoxythymine changes its conformation from B-form to A-form (table S2) which indicates the enzyme binding induces conformational transition from B-form to A-form, similar to other DNA binding proteins.<sup>38,41</sup> Analysis of the backbone phosphodiester dihedral angles indicates that backbone rearrangement in DNA-RNA hybrid is not required for ribonuclease H enzyme binding (figure S6). Thus, the ability of the DNA strand of the hybrid to assume A-type and B-type conformations is key for binding to take place.

Previous studies have proposed that the characteristic minor groove widths of the DNA-RNA hybrid duplexes are recognized by the RNase H enzyme. In addition, the ability of the DNA strand to sample conformational space corresponding to a typical A-form nucleic acid has also been suggested to be a recognizing factor. It is well-known that there exists a significant overlap between the distributions of the minor groove widths of RNA and hybrid duplexes, and RNA exists in the A-form. Hence, this does not explain how the RNase H enzyme is able to differentiate the hybrid duplex from a RNA duplex. Figure 7 shows two extreme conformational transitions those of especially T4 and C5. Upon binding to the enzyme, sampling of the C3'-endo in case of T4, but it is completely diminished in case of C5, which samples the C2'-endo region. While RNA strand can sample the C3'-endo region, ribose cannot exist in

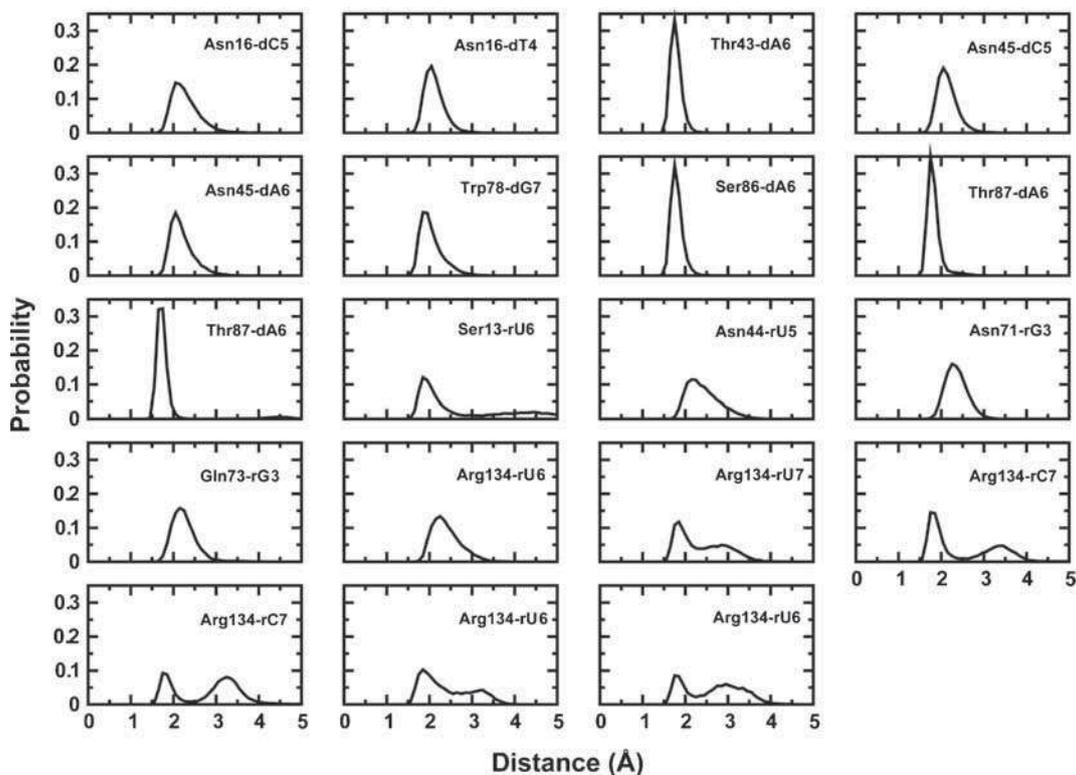


**Figure 7.** Probability distributions of pseudorotation angles of deoxyribose sugars present in DNA-RNA hybrid duplexes in holo and apo forms.

the C2'-endo conformation, which seem to be essential for binding. Overall, we conclude that the flexibility of the backbone of the DNA strand of the hybrid duplex enables optimal binding to the RNase H enzyme.

### 3.6 Ribonuclease H enzyme-hybrid interactions

The calculated solvent accessible surface areas of entire protein-hybrid complex and individual monomers



**Figure 8.** Probability distributions of distances corresponding to hydrogen bonds formed when RNase H enzyme interacting with DNA-RNA hybrid duplex.

**Table 3.** Statistics for the hydrogen bonds observed at the interface of RNase H enzyme-DNA-RNA hybrid duplex complex during MD simulations on binary complex (resolution 5.0 ps, occupancy  $\geq 30\%$ , life time 5.0 ps).

Atom pair	Occupancy (%)	Average life time (ps)	Number of events	Distance (Å)	
				MD	Exp
Asn16:HD21-dC5:O4'	68.1	16.8	3240	2.2 (0.1)	2.7
Asn16:HD22-dT4:O2	89.1	48	1485	2.0 (0.1)	2.1
Thr43:HG1-dA6:O1P	99.9	4995.3	16	1.7 (0.1)	1.8
Asn45:HD21-dC5:O2	87.3	40.4	1729	2.1 (0.1)	2.1
Asn45:HD22-dA6:O5'	79.8	26.1	2449	2.1 (0.1)	2
Trp78:HE1-dG7:O1P	88.5	49.1	1442	2.0 (0.1)	1.8
Ser86:HG1-dA6:O1P	99.7	1449.6	55	1.7 (0.1)	2.9
Thr87:HN-dA6:O2P	97.9	410.2	191	1.8 (0.1)	1.8
Thr87:HG1-dA6:O2P	94.9	1025.9	74	1.8 (0.1)	3.3
Ser13:HN-rU6:O1P	56.6	43.5	1042	2.7 (0.2)	1.9
Asn44:HD22-rU5:O4'	53.4	11.5	3729	2.3 (0.1)	2.5
Asn71:HD21-rU5:O1P	59.9	13.1	3669	2.3 (0.1)	2.5
Asn71:O-rG3:H2'	74	22.9	2581	2.1 (0.1)	3.1
Gln73:HN-rG3:O2'	56.8	15.2	2989	2.3 (0.1)	2.2
Arg134:HH11-rU6:O1P	55.7	28.7	1549	2.3 (0.1)	13.6
Arg134:HH12-rC7:O1P	56.5	45.6	991	2.5 (0.1)	12.1
Arg134:HH12-rC7:O2P	35	31.4	890	2.7 (0.1)	9.8
Arg134:HH21-rU6:O1P	57.5	29.1	1579	2.4 (0.1)	14.1
Arg134:HH21-rU6:O2P	37.5	24.5	1225	2.5 (0.1)	13.3

indicate that an average of  $1770 \text{ \AA}^2$  surface area has buried during the binding process (figure S7). The subtraction of surface area of complex from the sum of individual monomer surface areas result the buried surface area which is the surface area buried during binding. The resulted interface area is around  $885 \text{ \AA}^2$  which is equal to half in buried surface area and is in the order of surface area required for antigen-antibodies recognition. To examine the interactions present at the protein-hybrid interface, several analysis were performed. The enzyme binds to the nucleic acid by forming several hydrogen bonds, hydrophobic and electrostatic interactions, and water mediated hydrogen bonds. The statistical analysis performed to understand the formation and stability of hydrogen bonds between RNase H and hybrid suggests that the residues Thr43, Trp78, and Thr87 forms strong hydrogen bonds with phosphoryl oxygen atoms of DNA strand (table 3). The Asn16, Asn44 and Asn45 residues have hydrogen bond interactions with O2 atom of T4 and C5 bases of DNA strand respectively. The residues Asn16 and Asn 45 have hydrogen bond interactions with O4' of sugar of C5 residue and O5' of sugar of A6 residue of DNA strand respectively. The residues Ser13, Lys119, and Thr122 have interactions with phosphoryl oxygen atoms and residues Gly15, Gln73 and Lys119 with sugar oxygen atoms of RNA strand. The residue Gln73 have extra hydrogen bonds with H2' of sugar and O2 atom of U2 base of RNA strand. Formation of many new hydrogen bonds is observed during the simulations along with

the hydrogen bonds observed in experimental structure with higher occupancy. The hydrogen bonds observed in experimental structure were fairly stable during simulations. The probability distribution of hydrogen bonds show that the hydrogen bonds are fairly stable during simulation (figure 8). The hydrogen bond between thr43 and dA6 is very stable during the simulations and indicate that this bond may have important role in stabilizing the complex (figure S8). It is also noticed that the donor-acceptor pairs of new hydrogen bonds formed during simulation are at longer distance in the experimental structure. A wide range of hydrophobic contacts (table S5), electrostatic interactions (table S6) and water mediated hydrogen bonds (table S7) were observed during the simulations. Even though the crystal structure has very few water mediated bonds, many new water mediated hydrogen bonds were observed during simulations. These bonds may assist in non-specific interactions between protein and hybrid.

#### 4. Conclusions

MD simulations in explicit solvent environment have been performed on the binary complex of ribonuclease H and DNA-RNA hybrid complex, and the respective monomers to understand the structural differences in monomers, factors responsible for their conformational preorganization, active site dynamics and the stability of the key interactions between protein and hybrid duplex. MD simulations for each system are extended

to 100ns and the results obtained from the analysis of long MD trajectories are summarized as below:

Many structural and energetic changes are observed between protein in apo and holo forms. These changes are especially observed in the regions loop6 and  $\beta$ 1 strand which substantially interact with the substrate, DNA-RNA hybrid duplex. Most conformational rearrangements are observed at the C-terminal region. The residues present in the active site are preorganized before binding to hybrid duplex and become rigid. The protein binding induces conformational changes in DNA-RNA hybrid duplex, in particular the DNA strand. The protein changes the internal arrangement of base pairs, especially at dArU/dTrA base pair step and changes their deoxyribose conformation to C3'-endo or C2'-endo. The analyses also indicate that the protein binds the hybrid at minor groove similar to many protein-nucleic acid complexes and induce conformational transition in deoxyribose sugars. Water coordination around the metal ions present at the active site region is stable throughout the simulations, indicating its importance in catalytic mechanism. The protein-hybrid complex is very stable due to various interactions of protein with the hybrid through hydrogen bonds, hydrophobic, electrostatic and water mediated hydrogen bonds. These regions also have less fluctuation values in holo form suggesting their strong interactions with the nucleic acid. Several new interactions have been observed between enzyme and hybrid duplex during the simulations. Because of the structural similarity of the model used here with ribonuclease H of HIV-RT, this study offers insights that has potential applications in designing small molecules to inhibit its function.

### Supplementary Information

This includes methods to calculate hydration numbers and hydrogen bonds (Supplementary methods), RMSD and radius of gyration (figure S1; table S1), Ramachandran plot (figure S2), hydrogen bond distances (figure S3), helical parameters (figures S4–S5), backbone angles (figure S6), buried surface area (figure S7), hydrogen bond distances (figure S8), pseudorotation angles (table S2), stacking interactions (table S3), hydration numbers and sasa (table S4), and statistics of hydrogen bonds (tables S5, S6, S7). Supplementary Information is available at [www.ias.ac.in/chemsci](http://www.ias.ac.in/chemsci).

### Acknowledgements

We thank AICTE, and Department of Atomic Energy-BRNS (37(2)/14/05/2015/BRNS/20046) for financial

assistance. GS thanks Council of Scientific and Industrial Research (CSIR), India for senior research fellowship.

### References

1. Nakamura H, Oda Y, Iwai S, Inoue H, Ohtsuka E, Kanaya S, Kimura S, Katsuda C, Katayanagi K and Morikawa K 1991 *Proc. Natl. Acad. Sci. USA* **88** 11535
2. Nowotny M, Gaidamakov S A, Crouch R J and Yang W 2005 *Cell* **121** 1005
3. Qiu J, Qian Y, Frank P, Wintersberger U and Shen B 1999 *Mol. Cell. Biol.* **19** 8361
4. Crouch R J and Toulme J J 1998 In *Ribonucleases H* (INSERM Editions: Paris)
5. Tadokoro T and Kanaya S 2009 *FEBS J.* **276** 1482
6. Worrall J A R and Luisi B F 2007 *Curr. Opin. Struct. Biol.* **17** 128
7. Cerritelli S M and Crouch R J 2009 *FEBS J.* **276** 1494
8. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Ikehara M, Matsuzaki T and Morikawa K 1990 *Nature* **347** 306
9. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Nakamura H, Ikehara M, Matsuzaki T and Morikawa K 1992 *J. Mol. Biol.* **223** 1029
10. Yang W, Hendrickson W A, Crouch R J and Satow Y 1990 *Science* **249** 1398
11. Kanaya S, Kohara A, Miura Y, Sekiguchi A, Iwai S, Inoue H, Ohtsuka E and Ikehara M 1990 *J. Biol. Chem.* **265** 4615
12. Haruki M, Noguchi E, Nakai C, Liu Y Y, Oobatake M, Itaya M and Kanaya S 1994 *Eur. J. Bio. Chem.* **220** 623
13. Kanaya S and Crouch R J 1983 *J. Biol. Chem.* **258** 1276
14. Oda Y, Iwai S, Ohtsuka E, Ishikawa M, Ikehara M and Nakamura H 1993 *Nucleic Acid Res.* **21** 4690
15. Lai L, Yokota H, Hung L W, Kim R and Kim S H 2000 *Struct. Fold. Des.* **8** 897
16. Ariyoshi M, Vassylyev D G, Iwasaki H, Nakamura H, Shinawaregawa H and Morikawa K 1994 *Cell* **78** 1063
17. Ceschini S, Keeley A, McAlister M S, Oram M, Phelan J, Pearl L H, Tsaneva I R and Barrett T E 2001 *EMBO J.* **20** 6601
18. Rice P A and Baker T A 2001 *Nat. Struct. Biol.* **8** 302
19. Yang W and Steitz T A 1995 *Structure* **3** 131
20. Gyi J I, Lane A N, Conn G L and Brown T 1998 *Biochemistry* **37** 73
21. Suresh G and Priyakumar U D 2014 *Phys. Chem. Chem. Phys.* **16** 18148
22. Nowotny M, Gaidamakov S A, Ghirlando R, Cerritelli S M, Crouch R J and Yang W 2007 *Mol. Cell* **28** 264
23. Steitz T A and Steitz J A 1993 *Proc. Natl. Acad. Sci. USA* **90** 6498
24. De Vivo M, Dal Peraro M and Klein M L 2008 *J. Am. Chem. Soc.* **130** 10955
25. Krakowiak A, Owczarek A, Koziolkiewicz M and Stec W J 2002 *ChemBiochem* **3** 1242
26. Cassano A G, Anderson V E and Harris M E 2004 *Biopolymers* **73** 110
27. Haruki M, Noguchi E, Kanaya S and Crouch R J 1997 *J. Biol. Chem.* **272** 22015
28. Cerritelli S M, Frolova E G, Feng C, Grinberg A, Love P E and Crouch R J 2003 *Mol. Cell* **11** 807

29. Luisi B F, Xu W X, Otwinowski Z, Freedman L P, Yamamoto K R and Sigler P B 1991 *Nature* **352** 497
30. Nelson D L and Cox M M 2005 In *Lehninger's Principles of Biochemistry* 4th ed. (New York: W H Freeman)
31. Draper D E 1993 *Proc. Natl. Acad. Sci. USA* **90** 7429
32. Duan Y, Wilkosz P and Rosenberg J M 1996 *J. Mol. Biol.* **264** 546
33. Arndt J W, Gong W, Zhong X, Showalter A K, Liu J, Dunlap C A, Lin Z, Paxson C, Tsai M D and Chan M K 2001 *Biochemistry* **40** 5368
34. Tishchenko S, Nikonova E, Nikulin A, Nevskaya N, Volchkov S, Piendl W, Garber M and Nikonov S 2006 *Acta Cryst. D* **62** 1545
35. Yang X, Gérczei T, Glover L and Correll C C 2001 *Nature Struct. Biol.* **8** 968
36. Sen S and Nilsson L 1999 *Biophys. J.* **77** 1782
37. Reyes C M, Nifosi R, Frankel A D and Kollman P A 2001 *Biophys. J.* **80** 2833
38. Chen L, Zhang J, Yu L, Zheng Q C, Chu W T, Xue Q, Zhang H X and Sun C C 2012 *J. Phys. Chem. B* **116** 12415
39. Driessen R P, Meng H, Suresh G, Shahapure R, Lanzani G, Priyakumar U D, White M F, Schiessel H, van Noort J and Dame R T 2013 *Nucleic Acids Res.* **41** 196
40. Furini S, Barbini P and Domene C 2013 *Nucleic Acids Res.* **41** 3963
41. Priyakumar U D, Harika G and Suresh G 2010 *J. Phys. Chem. B* **114** 16548
42. Rosta E, Nowotny M, Yang W and Hummer G 2011 *J. Am. Chem. Soc.* **133** 8934
43. Maláč K and Barvík I 2013 *J. Mol. Graph. Model.* **44** 81
44. Brooks B R, Brooks C L, MacKerell A D, Nilsson L, Petrella R J, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner A R, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R W, Post C B, Pu J Z, Schaefer M, Tidor B, Venable R M, Woodcock H L, Wu X, Yang W, York D M and Karplus M 2009 *J. Comput. Chem.* **30** 1545
45. Priyakumar U D, Ramakrishna S, Nagarjuna K R and Reddy S K 2010 *J. Phys. Chem. B* **114** 1707
46. Suresh G and Priyakumar U D 2014 *J. Phys. Chem. B* **118** 5853
47. Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 *J. Chem. Phys.* **79** 926
48. Ryckaert J P, Ciccotti G and Berendsen H J C 1977 *J. Comput. Phys.* **23** 327
49. Field M J and Karplus M 1992 In *CRYSTAL: Program for Crystal Calculations in CHARMM* (Harvard University: Cambridge, MA)
50. Darden T, Perera L, Li L P and Pedersen L 1999 *Structure* **7** R55
51. Essmann U, Perera L, Berkowitz M L, Darden T A, Lee H and Pedersen L G 1995 *J. Chem. Phys.* **103** 8577
52. Steinbach P J and Brooks B R 1994 *J. Comput. Chem.* **15** 667
53. Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L and Schulten K 2005 *J. Comput. Chem.* **26** 1781
54. Foloppe N and MacKerell A D 2000 *J. Comput. Chem.* **21** 86
55. MacKerell A D and Banavali N K 2000 *J. Comput. Chem.* **21** 105
56. MacKerell A D, Bashford D, Bellott D R L, Evanseck J D, Field M J, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F T K, Mattos C, Michnick S, Ngo T, Nguyen D T, Prodhom B, Reiher W E, Roux B, Schlenkrich M, Smith J C, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D and Karplus M 1998 *J. Phys. Chem. B* **102** 3586
57. MacKerell A D, Feig M and Brooks C L 2004 *J. Comput. Chem.* **25** 1400
58. Suresh G and Priyakumar U D 2013 *J. Phys. Chem. B* **117** 5556
59. Feller S E, Zhang Y, Pastor R W and Brooks R W 1995 *J. Chem. Phys.* **103** 4613
60. Hoover W G 1985 *Phys. Rev. A* **31** 1695
61. Priyakumar U D and MacKerell A D 2010 *J. Mol. Biol.* **396** 1422
62. Humphrey W, Dalke A and Schulten K 1996 *J. Molec. Graphics* **14** 33
63. Lavery R, Moakher M, Maddocks J H, Petkeviciute D and Zakrzewska K 2009 *Nucleic Acids Res.* **37** 5917
64. Stafford K A and Palmer A G 2014 *F1000 Research* **3** 67
65. Rychlik M P, Chon H, Cerritelli S M, Klimek P, Crouch R J and Nowotny M 2010 *Molecular Cell* **40** 658
66. Babu C S, Dudev T and Lim C 2013 *J. Am. Chem. Soc.* **135** 6541