The dynamics of protein-RNA interfaces using all-atom molecular dynamics simulations

Afra Sabei,[†,¶] Cécilia Hognon,[†,¶] Juliette Martin,[‡,§] and Elisa Frezza[*,†]

†*Université Paris Cité, CiTCoM, CNRS, F-75006 Paris, France*

‡*Univ Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5086 MMSB, Lyon, France*

¶*The authors contributed equally to the work*

§*Laboratory of Biology and Modeling of the Cell, Université de Lyon, ENS de Lyon, Université Claude Bernard, CNRS UMR 5239, Inserm, U1293, Lyon, France.*

E-mail: elisa.frezza@u-paris.fr

Phone: +33 (0)1 53 73 15 69

Abstract

Facing the current challenges raised by human health diseases requires the understanding of cell machinery at a molecular level. The interplay between proteins and RNA is key for any physiological phenomenon, as well protein-RNA interactions. To understand these interactions many experimental techniques have been developed, spanning a very wide range of spatial and temporal resolutions. In particular, the knowledge of tridimensional structures of protein-RNA complexes provides structural, mechanical and dynamical pieces of information essential to understand their functions. To get insights into the dynamics of protein-RNA complexes, we carried out all-atom molecular dynamics simulations in explicit solvent on nine different protein-RNA complexes with different functions and interface size by taking into account the bound and unbound forms. First, we characterized structural changes upon binding and for the

RNA part the change in the puckering. Second, we extensively analyzed the interfaces and the structural waters involved in the binding. Based on our analysis, the interfaces rearranged during the simulation time showing alternative and stable residue-residue contacts with respect to the experimental structure.

Introduction

Protein-RNA interactions are crucial for any cellular function.[1] RNA-binding proteins are key to RNA export, processing, splicing, localization, translation, degradation, and the regulation of all these steps.[2,3] The dysfunction of such complexes is implicated in many humans and animal pathologies.[4,5] For example, the formation of ribonucleoprotein particles (RNPs) including mRNAs is key for the post-transcriptional regulation of gene expression.[6,7] As another example, many viruses manipulate the translation initiation complexes to ensure their replication, by recruiting the host ribosomes to translate their mRNA.[8] Hence, understanding and predicting protein-RNA interactions is a crucial step to develop new therapeutic approaches and improve our knowledge of cell machineries.

Despite their significance, the specific mechanisms of protein-RNA interactions are not completely understood. Our ability to predict such mechanisms is hindered by several factors, among which the conformational changes undergone by the structures during the binding process and the lack of available tridimensional complex structures are to be mentioned.[9,10] In fact, obtaining high-resolution tridimensional structures through high-resolution techniques is still a challenging task as shown by the relatively small number of structures deposited in the Nucleic Acid Data Bank (NDB).[11,12]The limited number of known complex structures is nevertheless constrained by the inherent flexibility of proteins and/or RNA and the size of the complexes. The ability of binding proteins to change their conformation is often made possible by RNA flexibility, which is a crucial component in the binding process. Among other techniques, NMR spectroscopy is largely used and one of the most powerful methods for studying RNA's and protein-RNA structure and conformational dynamics in solution.

Despite that, the current application of NMR methodology imposes a constraint on the size of the molecules, typically limiting them to approximately 150 nucleotides (nt), achieved through the utilization of site-selective labeling strategies.[13–15]

Another key aspect is the proper characterization of the role of energetic, structural, and thermodynamic factors in the specificity and mechanism of complexation. Such factors are still beyond our understanding for a number of reasons. Thermodynamic data are very sparse and often obtained in different conditions, and the determination of the actual affinity strongly depends on the method used and on the strength of the interactions that are being measured.[16,17] Therefore, most analyses have focused on structural explanations, and only a limited number of complexes have been analyzed in an attempt to characterize the common features of the interactions.[18,19] In the dataset developed by Krüger and co-workers, 322 complexes are analyzed and their structures present a large diversity of interface size and properties.[20] It has been shown that there are important differences in RNA-protein complexes as compared to DNA-protein complexes, most notably related to the nature of contacting surfaces of the two molecules.[18–20] For example, for specific protein-RNA binding the number and types of intermolecular interactions and preferred amino acids have been observed.[21,22] It is crucial to understand the specific characteristics of protein-RNA complexes with respect to protein-DNA interactions, because RNA and DNA can bind the same protein, but on orthogonal binding sites.[23] DNA and RNA double helices also differ in conformation: the former assumes a B-form and the latter an A-form. On the RNA side, the phosphate contributes less, and the sugar more, to the interaction than in protein-DNA complexes.[18,19] Regarding the amino acid propensity to be at the protein-RNA interface, it has been observed that it is close to the one obtained for protein-DNA interfaces that are enriched in positively charged and polar residues.[18,24] A crucial player in protein-RNA recognition is the 2′OH of the ribose.[19]

Although the information above is undoubtedly valuable to understand many aspects of protein-RNA interfaces, it gives a rather static view of such complexes and at the same time

does not take into account the relationship between experimental structural features and energetic costs (binding affinity). In fact, the anatomic description of protein-RNA interfaces and their physico-chemical properties generally relies on a static point of view. However, the dynamics of interfaces plays a major role in the mechanisms of complexation (involving both specific and non-specific complexes) and it can influence the resulting binding strength. Moreover, to date, in the large datasets of protein-RNA complexes available[20,25] most complexes don't dispose of both the structure of the cognate unbound RNA or proteins and the related binding affinity. This probably explains why so little is known on the role of dynamics of protein-RNA interfaces and on the energetic aspects of such interactions. Only few studies have been conducted on specific systems using all-atom Molecular Dynamics (MD) simulations,[26,27] for example short single-stranded RNAs bound to RNA recognition motifs (RRM) proteins (HuR-RRM3 and SRSF1-RRM2),[28] the adaptive binding of spliceosomal U1A-RNA binding to a hairpin RNA and an internal loop RNA,[29] a GGUG motif bound to Fused in sarcoma RRM protein,[30] reduced and oxidized TFIIIA Zinc Fingers interacting with 5S RNA.[31] Shaw and co-workers performed microsecond MD simulations to analyze the SARS-CoV-2 helicase replication–transcription complex, in particular the non-specific helicase 13, to highlight their mechanism and to characterize the interfaces.[32,33] However, no systematic study has been conducted to assess if protein-RNA interfaces can undergo significant changes in simulations and if changes in the interfaces of native complexes can be observed as pointed out for protein-protein[34,35] and protein-DNA interfaces also from some of us.[36–39]

In this context, water molecules also play a crucial role in the process of binding, in specific recognition, in interactions between protein-RNA as well as influence the structure and function of isolated biomolecules.[40] In fact, they seem to be involved in mediating interactions,[41] as well in the formation of H-bonding network to stabilize the interactions.[42] At the interface, water molecules play a crucial role in determining the specificity and stability of the complex. However, with respect to protein(s) and DNA-protein interfaces,[43] the water dy-

namics of RNAs and for protein-RNA complexation has been poorly investigated. Although several PDB structures have been analyzed, a static view has often been presented[41,44,45] and only few studies have been conducted that take into account the interplay between water and protein/RNA dynamics and their interfaces.[46]

In this study, we aim to characterize the dynamics of protein-RNA complexes and their interfaces at molecular level by performing a more systematic analysis. To get insights on the dynamics of protein-RNA complexes, we investigated nine protein-RNA complexes for which at least the 3D structures of the complex and the unbound RNA molecule are available. We carried out standard all-atom MD simulations of complexes and unbound proteins/RNAs in explicit water with K^+Cl ions at physiological conditions. MD simulations were extensively characterized and several analyses were performed. Among them, we focused on the characterization of the stability of their interfaces and interfacial water molecules. In each system, we analyzed in detail the time evolution of the contacts between interface residues and between interface residues and water molecules.

Materials and Methods

Data set

In order to analyze the protein-RNA interfaces, we took into account the datasets available in the literature for protein-RNA complexes.[16,47–50] We performed a structural characterization of each candidate complex in order to avoid redundant structures and with the aim to assure a diversity in the size and type of the interfaces. Based on that, we selected 9 complexes as reported in Table 1. The nine cases span various interface sizes (1341 - 3937 $Å^2$). The 3D structures of the 9 complexes under study are shown in Figure 1.

To model our starting structures, several steps were necessary with the aim to use the same number of residues for the complex and the free protein and RNA molecule. Terminal residues present in the free protein/RNA but absent from the complex were chopped, and

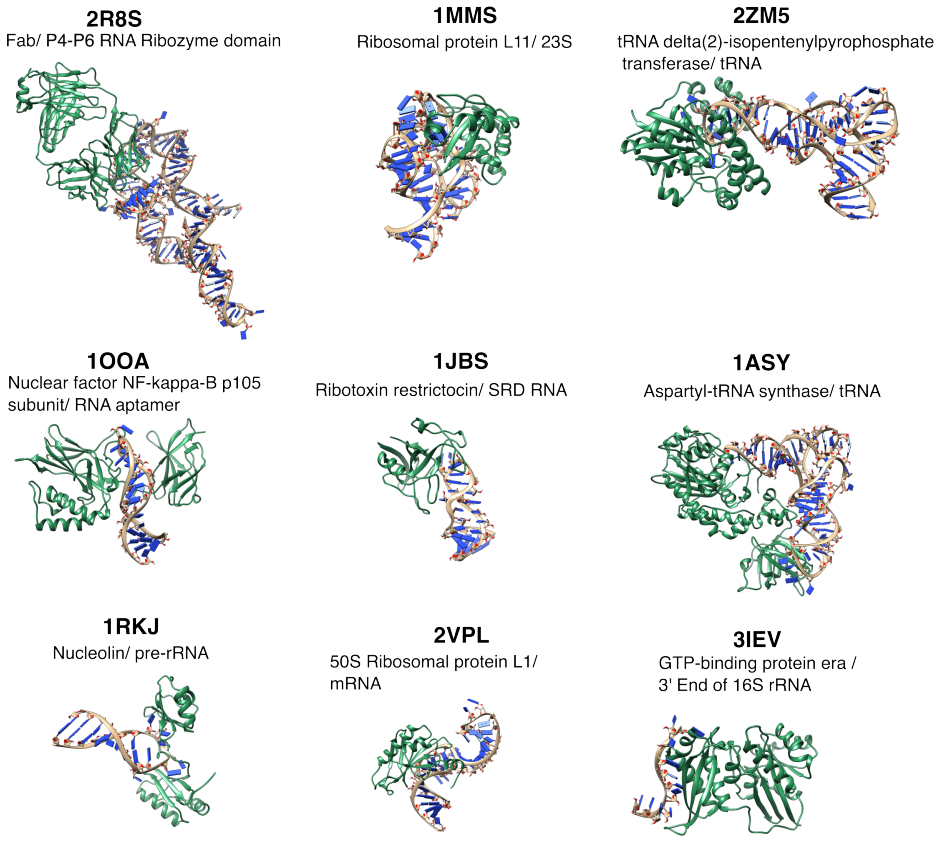

Figure 1: 3D structures of the nine protein-RNA complexes studied in this study, with one partner colored in green (protein) and the other one in beige (RNA).

residues present in the complex but absent in the free protein/RNA were modeled based on the complex. In the absence of the PDB structure for the unbound protein, the bound form was taken into account. When one or more residues were absent, homology modeling was used to add the missing residue using either Modeller 10.2[51] or ModeRNA 1.7.0[52] for the protein or the RNA molecule, respectively. An exception was made for the unbound form of the 1RKJ complex protein, due to the absence of a correct homology model. For this reason, we used the structure extracted from the PDB of the complex. For the PDB structures obtained via NMR, we first carefully analyzed the overall quality of the produced models and then we used the software DSSR (Dissecting the Spatial Structure of RNA)[53] to examine and annotate the RNA tertiary structures. As a result, we were able to determine both canonical and non-canonical base pairs and the number of hydrogen bonds in each structure. Then, in order to select the most representative model, we clustered the available models and constructed arc representations of the structures. To prepare the structures, we also used the crystallographic water molecules, the monovalent ions and Mg^{2+} ions. Finally, for the protein, we computed a pKa calculation on the complex structure using the software PDB2PQR[54] and the algorithm PROPKA[55] to define the protonation state of each residue. To deal with the same chemical species, we applied the same protonation state of the protein in the complex on the relative free protein(s).

Table 1: Dataset of unbound-to-bound transition. b: bound; u: unbound; r: RNA; p: protein; $RMSD_{exp}$: RMSD between the unbound and bound RNA structure.*: different NMR structures.

PDB ID_b	PDB $ID_{u,r}$	PDB $ID_{u,p}$	$RMSD_{exp}$ (Å)	Interface size (Å^2)
1OOA(A:C)	2JWV*	3DO7	5.4	1910
1ASY(A:R)	2TRA	1EOV	4.9	3585
2ZM5(A:C)	3L0U(A)		4.2	3937
1MMS	3I9C(A)	2K3F*	1.6	2456
3IEV(A:D)	1SDR(B)		7.6	2311
2R8S(HL:R A)	1HR2	3IVK(AB)	4.4	2416
2VPL(A:B)	1U63	2OV7(A)	3.7	2388
1JBS(A:C)	1AQZ	1AQZ	3.4	1341
1RKJ(A:B)*	1QWA		4.9	2218

All-atom Molecular dynamics simulation

All-atom MD simulations were performed with the GROMACS 2018 package[56–59] using the Amber ff14SB force field for proteins[60] and the the Amber ff99bsc0 force field with OL3 modification (ff99bsc0χ_{OL3}) for RNA molecules.[61] Each RNA molecule was placed in a cubic box and solvated with TIP4P-EW to a depth of at least 14 Å.[62] Each system was neutralized by adding potassium cations and then K^+Cl^- ion pairs to reach a physiological salt concentration of 0.15 M and by adding $MgCl_2$ to reach a concentration of 0.02 M.[63] Long-range electrostatic interactions were treated using the particle mesh Ewald method[64,65] with a real-space cutoff of 10 Å. The hydrogen bond lengths were restrained using P-LINCS,[66] allowing a time step of 2 fs. Translational movement of the solute was removed every 1000 steps to avoid any kinetic energy build-up.[67]

Before the MD production, we carried out the energy minimization and equilibration. During equilibration (at least 10 ns) a Berendsen thermostat ($\tau_T = 0.1$ ps) and Berendsen pressure coupling ($\tau_P = 0.5$ ps)[68] were used. The production part was carried out in an NTP ensemble at a temperature held at 300 K and a pressure held constant at 1 bar using the Bussi velocity-rescaling thermostat ($\tau_T = 0.1$ ps)[69] and the Parrinello-Rahman barostat ($\tau_P = 0.5$ ps).[70] During minimization and heating, RNA heavy atoms and protein backbone remained fixed by using positional restraints. During the equilibration, the restraints were gradually relaxed. Bond lengths were restrained using P-LINCS, allowing a time step of 2 fs. Initially, each simulation was 500 ns long. Due to the stability obtained during the simulations, most of them were extended, to a maximum of 1500 ns. All the simulation times are available in Table S1. A total simulation time of 21.75 μs was performed.

RMSD/RMSF analysis of complexes and single protein/RNA

We calculated RMSD time series on the unbound RNA, bound RNA, and complex using their starting structures as reference in order to determine system stability and binding effect. We computed RMSF for all atoms, averaged fluctuations for each residue, and superimposed the

trajectory on the initial structure for unbound RNA and separated chains for complexes. Using this method, we were able to examine how each chain's flexibility changed as it was bound. Finally, we performed a hierarchical cluster analysis using the software TTClust[71] to determine the clusters for the unbound and bound systems.

RNA analysis

Puckering analysis

To take the ribose flexibility and its conformation, we computed the sugar pucker. It is important to highlight that the sugar ring is well described using pseudorotation parameters. Although there are four possible pseudorotation parameters for a five-membered ring (Frezza et al.[72]), two in particular are useful to characterize the sugar conformation: the so-called phase (Pha) and amplitude (Amp). While the amplitude describes the degree of ring puckering, the phase describes which atoms are most out of the mean ring plane. We calculated these parameters using the expressions given below:

$$\mathrm{Amp} = \sqrt{(a^2 + b^2)}; \mathrm{Pha} = \cos^{-1}\left(\frac{a}{\mathrm{Amp}}\right) \tag{1}$$

where $a = 0.4\sum_{i=1}^{5} \nu_i \cos[0.8\pi(i-1)]$ and $b = -0.4\sum_{i=1}^{5} \nu_i \sin[0.8\pi(i-1)]$ with ν_i the ring dihedral angle i. This approach has the advantage of processing the ring dihedrals ν_1 (C1' -C2' -C3' -C4') to ν_5 (O4' - C1' -C2' -C3') in an equivalent manner. Conventionally, sugar ring puckers are divided into 10 families described by the atom which is most displaced from the mean ring plane (C1', C2', C3', C4' or O4') and the direction of such displacement (*endo* for displacements on the side of the C5 atom and *exo* for displacements on the other side). In different sugar conformations, the distance between neighboring C2 atoms in the nucleobases and the orientation of P relative to the sugar/bases are different from each other. Based on previous works, these quantities are supposed to be related to the chemical reactivity of RNA with SHAPE reagents. In order to understand the interplay between

the sugar conformation and protein-RNA binding, we grouped the sugar puckers into two large families. The sugar puckers C1'-*exo*, C2'-*endo*, C3' -*exo*, C4'-*endo* belong to the B-like family, while C1'-*endo*,C2'-*exo*, C3'-*endo*, C4'-*exo* belong to the A-like family. For the single angular quantities, because of their periodicity, standard statistics cannot be used to compute the average and the fluctuations of angular values. For example, the arithmetic mean of 0° and 360° is 180°, which is misleading because for most purposes 360° is the same thing as 0°. Hence, we used the directional statistics instead, as explained in.[72]

Interface Analysis

The interface analysis is conducted on MD snapshots taken every 40 ps (i.e. 12500 to 25000 snapshots depending on the system). Part of this analysis was computed with in-house python scripts that are available at `https://github.com/juliettemartin/DYN_INTERFACES`.

Interface definition

Interface contacts are defined at the residue level using a 5 Å distance cutoff between heavy atoms. The solvent accessible surface area (ASA) was computed with the NACCESS software using the default radius of 1.4 Å to take into account the water molecule.[73]

Following the criteria introduced by Levi for the protein-protein interfaces, distinct regions were defined at the interface.[74] Hence, residues with a change of ASA between the isolated chain and the complex are classified as: core if relative ASA in the isolated chain is greater than 25% and lower than 25% in the complex, rim if relative ASA in the complex is greater than 25%, support if the relative ASA in the isolated chain extracted from the complex is lower than 25%.

Interface properties

The accessible surface area buried by the interface is defined by :

$$\Delta ASA = ASA_A + ASA_B - ASA_{AB} \tag{2}$$

where ASA_{A} and ASA_{B} are the accessible surface areas of separate chains A and B respectively, and ASA_{AB} represent the accessible surface area of the complex. The gap index evaluates the complementarity of the interacting surfaces in a complex and is defined as

$$\Delta \mathrm{gap\,index} = \frac{2\,\mathrm{gap\,volume}}{\Delta ASA} \tag{3}$$

where the gap volume estimates the volume enclosed by any two molecules and is computed using the procedure developed by Laskowski and implemented in the SURFNET software.[75]

Interface clustering

To conduct a comparison between interfaces at various time points, we analyze the interface contacts of the respective snapshots, S1 and S2. The interface similarity is then measured by the Jaccard index J defined as:

$$J(C1, C2) = \frac{|C1 \cap C2|}{|C1 \cup C2|} \tag{4}$$

where C1 and C2 represents the two sets of contacts. The Jaccard index is equal to 1 if S1 and S2 have identical contacts. On the contrary it is equal to 0 if S1 and S2 have no common contacts.

First, the Jaccard indexes are converted into dissimilarity matrices by taking 1-J, followed by the computation of hierarchical clustering to identify clusters at the interfaces. To do so, the Ward.D2 method implemented in R was used.[76] No generic method can be applied to choose the optimal number of clusters since the choice depends on the application. In this

study, we aimed to have clusters that are different enough from each other, of reasonable size, and relatively stable in time. The choice of the optimal number of clusters was thus guided by the topology of the clustering dendrograms and the assessment of cluster size and stability in time when varying the number of clusters. We thus computed the size of the largest cluster, the size of the smallest cluster (number of snapshots in this cluster), and the number of cluster changes during the simulation time.

Choice of representative structures in each cluster

In each cluster, we defined the centroid as the snapshot with the highest average Jaccard similarity with respect to the other snapshots of the same cluster. To visualize the location of these centroids in each cluster, we applied principal component analysis on the Jaccard dissimilarity matrices and plotted the first two dimensions. In the SI, projections are presented.

Statistics of interface contacts in the different clusters

We computed the relative frequency of contacts in each cluster, defined by the proportion of snapshots exhibiting a contact in the cluster. Each contact is thus associated with FN_c values of relative frequencies with N_c the number of clusters. For each contact, we calculated the variance of these values among clusters. This quantity points out which contacts have similar frequencies (low variance) and which have variable frequencies across clusters (high variance). Variance values refer to contacts between residue pairs. To map these variance values at the level of residues, for each residue we selected the maximum variance, $\mathrm{var}(X)$, observed among contacts:

$$\mathrm{var}(X) = \max_{Y \text{ in contacting residues}} \{\mathrm{var}(XY)\} \tag{5}$$

where var(XY) is the variance of the relative frequency of the XY contact among clusters. For example, a residue X seen in contact with residue Y with a variance of 0.2 and in with residue Z with a variance of 0.4 will receive a value of 0.4.

Water analysis: interfacial water molecules

To analyze the water at the interface, we used the same approach proposed by some of us in.[39] Hence, first, we determined the water molecules whose distance, d_w, from both protein and RNA is below 4 Å. To compute this distance, only the heavy atoms of the biomolecules and the oxygen of the water molecules were considered for determining the number of water molecules at the interface, also called interfacial water molecules. Interfacial water molecules were determined for the representative structures and along MD simulations. For the latter, to investigate the contacts mediated by water we also analyzed the contacts formed for each interfacial water molecule with e protein and RNA partners using the same distance cutoff d_w. The data obtained from the MD simulations were analyzed using Python scripts, the packages `mdtraj`,[77] `NumPy`[78] and `SciPy`[79] libraries.

Results and Discussions

Global conformational changes observed in molecular dynamics: RMSD profiles, RMSD clustering and RMSF profiles of complexes and single protein/RNA

As presented above in the Section Materials and Methods, first we performed hierarchical clustering analysis using TTClust[71] based on RMSD on the complexes and the single protein and RNA and we computed the RMSD time series. RMSD time series, based on starting and clustering structures results, are summarized in Supplementary Figures S1-9. The simulations carried out on the unbound proteins highlight low (e.g., 1JBS), moderate and high

conformational changes, including confined to tails, specific loops and relative movements between protein domains (e.g. 1ASY). When unbound RNA is simulated, low (e.g. 1MMS), moderate and high conformational changes are observed, including for example the tails (e.g. 2ZM5), specific loops in the hairpin (e.g. 1JBS) or backbone conformational changes which impact the whole structure and arise to relative movements between different parts of the RNA structure (e.g. 2R8S). RMSF profiles are shown in Figure S10-12, with 3D structures colored according to the change of flexibility between unbound states and complexes, defined as the difference between the RMSF computed in the bound state and the unbound ones. Based on these profiles, the flexibility of proteins and RNAs in complexes can be comparable to what is seen in the unbound forms, but also higher or lower than the one obtained in the free states. In fact we can observed the flexibilization either a the level of the interface or at sites distant from the interface probably due to allosteric effects. For some complexes (1ASY, 1OOA, 1RKJ and 2Z5M), large pivotal movement of one biomolecule with respect to the other are observed which induce a high global change in the RMSD.

Geometric interface parameters along MD simulations

We computed the geometric interface parameters, the ΔASA, the gap volume and the gap index, along the MD trajectories and on the experimental starting structures, to investigate the impact of the dynamics on these parameters. Figure 2 and Supplementary Figure S13-18 show the distribution and Supplementary Table S2 reports the relative values. All geometric interface parameters show variability during the simulation. While the initial values may deviate from the crystallographic ones, in the majority of cases, the simulations explore values in proximity to the crystallographic ones. By comparing these properties between the experimental structure, the starting structure ($t = 0$) and the average along the MD simulation, we highlighted different profiles. The 3IEV complex shows the greatest stability for the 3 parameters studied, while 2ZM5 and 1MMS, which are very stable for the gap index, show greater deviation for the other two. The greatest deviation of the gap index values,

both experimental and baseline, is shown for the 1OOA and 1JBS complexes. However, this pattern is not found for ΔASA and gap volume, properties for which these complexes have an average deviation approximately equal to the other cases. These results show the considerable variability of the different complexes studied.

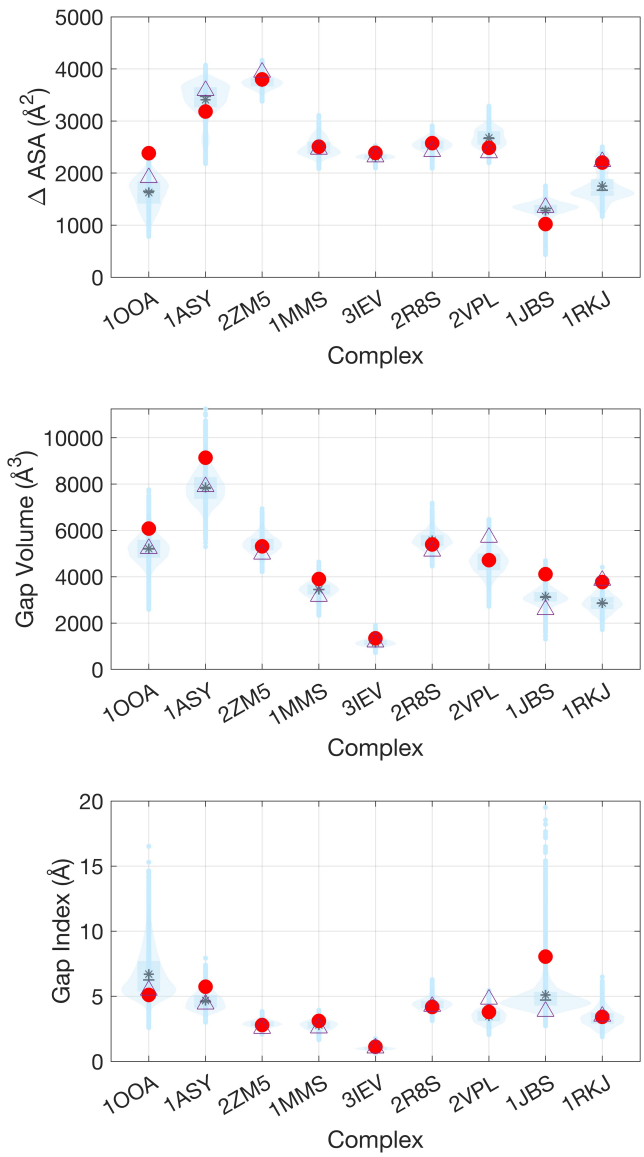

Figure 2: Distribution of geometric interface parameters for the nine complex represented in light blue as violin plots, with the average value indicated by an asterisk: Interface size measured by ΔASA (top), gap volume (middel) and gap index (bottom). Red filled circle: starting structure for MD production ($t = 0$). Purple triangle: experimental structure.

Dynamics of interfaces

Conservation of initial interface contacts along MD simulations

Along the MD trajectories, we determined the conservation of initial interface contacts for each complex, as explained in the Material and Method. As shown in Figure S19, the preservation of initial contacts varies according to the protein-RNA complex under investigation. We highlighted two different patterns, a high very stable interface, with half of all complex maintaining at least 80% of initial contacts established (1MMS, 2R8S, 2VPL, 2ZM5 and 3IEV), and a major switch in contacts for the others complexes (1ASY, 1JBS, 1OOA and 1RKJ). The 3IEV complex, composed of *Escherichia coli* Ras-like protein bound to 3'end of 16S rRNA, has a medium size interface (ΔASA = 2,311 Å^2) and presents the most conserved interface: about 90% of conserved contacts during the 500 ns of simulation. The MiaA protein complex with tRNA, 2ZM5, with the largest surface area (ΔASA = 3,937 Å^2) also has a very stable interface (80%). Concerning 1JBS complex, the restrictocin and a 29-mer SRD RNA analog, with the smallest surface area (ΔASA = 1,341 Å^2), the complex looses the majority of the initial contacts at the beginning of the MD simulation. The fraction of conserved contacts is decreasing to 25% at 15ns, then goes back to 75% during the rest of the simulation. On the contrary, 1OOA and 1RKJ retain less than 50% of the initial contacts. The highest change is noticed for 1RKJ complex, formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target, (ΔASA = 2,218 Å^2) with only 37% of preserved contacts. Regarding 1OOA complex, distinct conservation levels are observed suggesting different substates of protein-RNA interfaces, one at 50% and another one at 25%.

Analysis of the substates of protein-RNA interfaces

As explained in the Material and Methods section, the interface similarity was measured by the Jaccard index J, and then we performed hierarchical clustering based on interface

contacts. Clustering interface results are reported in Figure 3 for 1RKJ complex and in the Supplementary Figure S20-28 for the other complexes. In regards to the results, four of the nine complexes analyzed in this study exhibit different substates of their protein-RNA interaction interface. The nucleolin/rRNA complexes, 1RKJ, displays five distinct substates along the MD trajectory. Except for cluster 1, the other clusters are evenly distributed along the simulation, as shown in Figure 3C. We obtained 3 clusters for 1ASY, 2 clusters for 1JBS and 4 clusters for 1OOA complex. Remarkably, interface clusters were found for the complexes with major switch in initial contacts, as described in the previous paragraph. That is confirmed by the conservation of initial interface contacts for the five other complexes (1MMS, 2R8S, 2VPL, 2ZM5 and 3IEV). Moreover, the lack of relationship between ΔASA and number of clusters is again confirmed by these results, as previously mentioned for protein/protein interaction.[39]

To explore the structural differences between substates, we computed the variance of each contact across the interface clusters. The matrix representation of the contact variance for 1RKJ is shown in Figure 4A. As highlighted in Figure 4B, at RNA level, all the residues initially in contact with the protein have high contact variance between clusters. Same observation is made for the protein residues. Thus, a majority of contacts are not maintained between the five clusters, with significant changes in all regions for protein and RNA. Different characteristics are found for the other complexes. For 1ASY, 1JBS and 1OOA complexes, the matrix are available in the Supplementary Figure S29-31. In contrast to 1RKJ, a majority of contacts, in the case of 1ASY, have stable frequencies across clusters. The results highlight one region at protein/RNA interface which varies considerably for the three interface clusters, between C 649 to C 674 for RNA and Asp 229 to Pro 557 for protein. 1JBS complex follows the same pattern than 1ASY complex (localized variations), whereas 1OOA has the same profile as 1RKJ (extensive variations).

Therefore, for the interface clusters found for the complexes, the geometric interface parameters were investigated. Mean values with standard deviations of all the properties

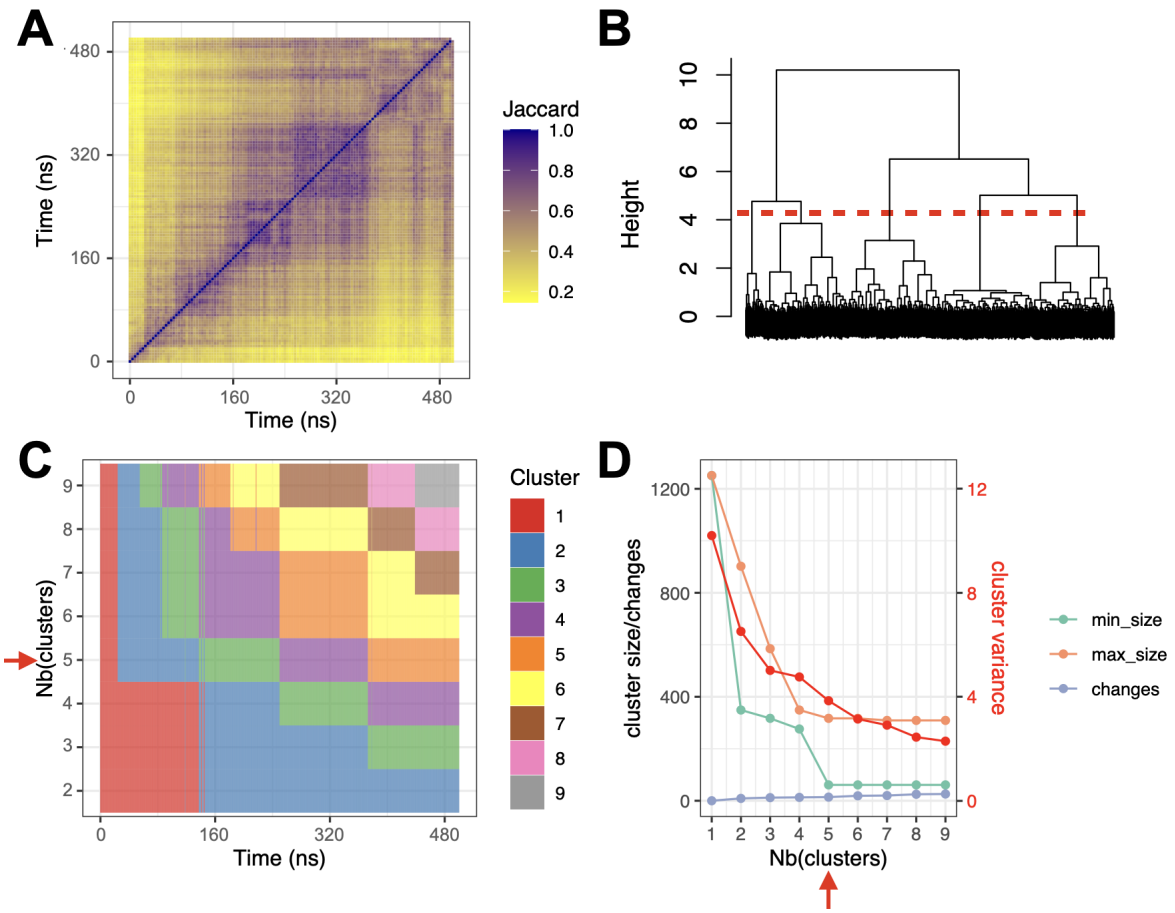

Figure 3: Clustering results for the 1RKJ complex. A. Jaccard similarity matrix. For each pair of snapshots, a yellow pixel indicates low interface similarity, and a purple pixel indicates high interface similarity. B. Clustering dendrogram. C. Cluster membership along the MD simulation, for different numbers of clusters. D. Cluster size, number of changes, and intra-cluster variance in red, for different numbers of clusters; min_size: size of the smallest cluster, max_size: size of the largest cluster, changes: number of cluster changes during the simulation. The red dashed line in panel B. and the red arrows in panels C.-D. indicate the optimal number of clusters.

(ΔASA, gap volume and gap index) are available in the Supplementary Table S2 and the distributions in the Supplementary Figure S14, S16 and S18. For all the complexes studied, the results indicate a clear difference between the clusters for the following two parameters: ΔASA and the gap index. The last parameter, gap volume, revealed a clear difference for the 1RKJ complex.

By considering the distribution of these parameters for each cluster, the gap index and

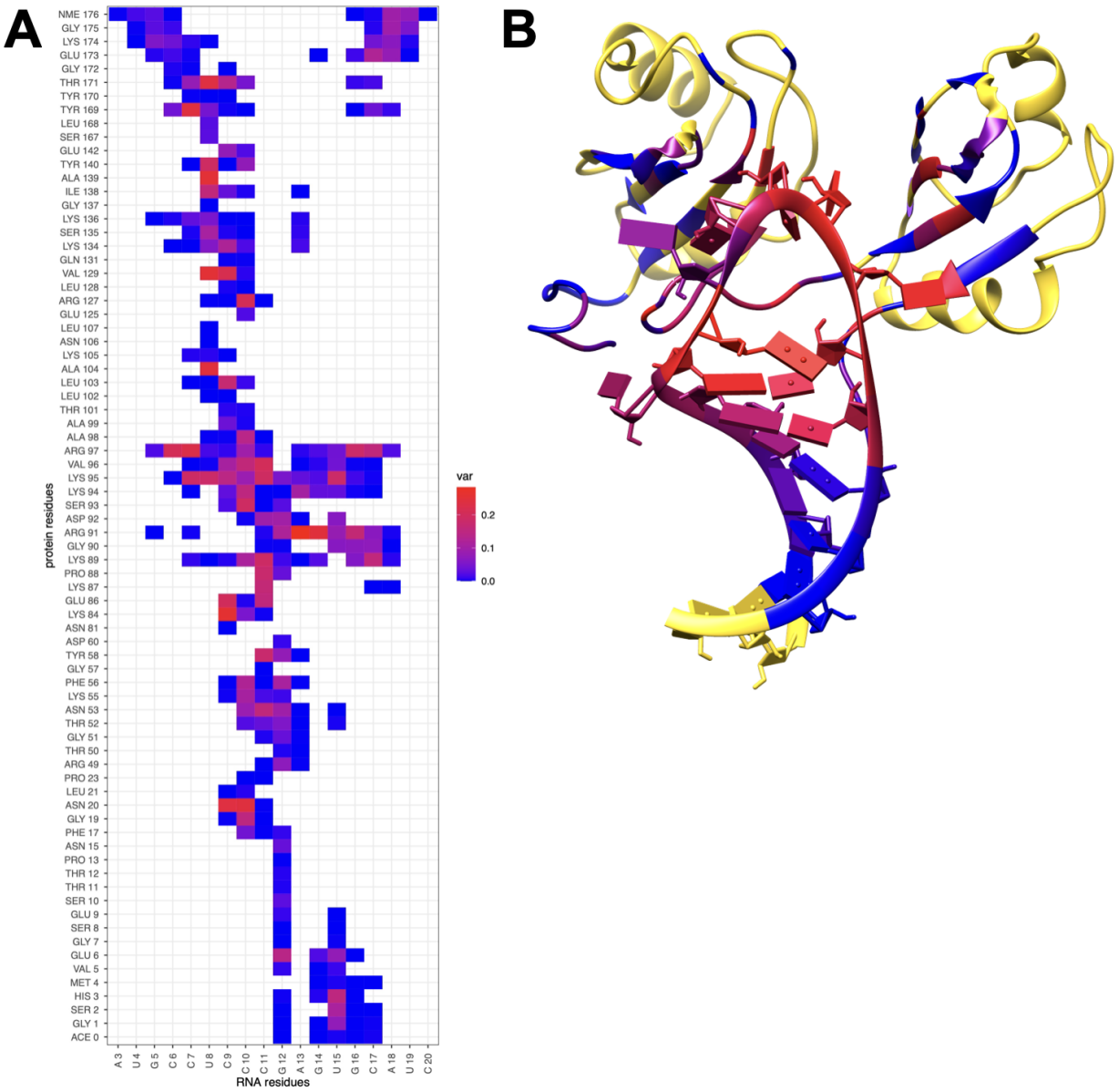

Figure 4: Contact variance in complex 1RKJ. A. Matrix representation of each contact variance across the five interface clusters. Blue: low variance; Red: high variance. B. 3D structure representation of 1RKJ complex. Interface residues are colored by the maximum of contact variance (blue to red) and the other residues are colored in yellow.

gap volume are distributed over the same region for all the clusters of the different complexes. However, the ΔASA property has different pattern depending on the cluster under consideration. The two clusters sample different ΔASA ranges, indicating interfaces of different size.

The last observation also applies to complexes 1ASY, 1OOA and 1RKJ, where the different clusters correspond to interfaces of different size. All these results highlighted that protein-RNA interface dynamics vary significantly depending on the complex studied.

Puckering analysis

To better characterize the impact of the interaction between protein and RNA on the conformation and the flexibility of the ribose, specially for residues at the interface, we computed the sugar pucker, as mentioned in the Materiel and Methods section.

Comparison between unbound and bound forms

We compared the ribose puckering between unbound and bound forms. The results obtained for 1RKJ complex, shown in Figure 5 A, highlight the influence of protein-RNA interaction on the sugar pucker. For the unbound RNA, C3'-*endo* form, and more widely A-like family, are the predominant ones except for residues 10 and 11. On the contrary, the bound form has a different layout, with an increase of C2'-*endo* form, and B-like family. This change in the conformation of the sugar occurs for residues that are highly involved in interactions with the protein, as illustrated by green bar and stars and in Figure 5 B.

The results on the other protein-RNA complexes studied are summarized in Supplementary Figure S32-36. The same trend concerning interface residues as 1RKJ is found for the other complexes. As with 1ASY, we can also point out a change in the forms of the B-like family towards the A-like family between the unbound and bound forms for residues not present at the interface, such as for residues 15 to 20 and 50 to 60 of this complex. An exception was 2VPL complex, which showed no major changes in ribose pucker between the absence or presence of the protein. All these results highlight the impact of protein-RNA interaction on RNA sugar conformation. Moreover, as previously demonstrated, C2'-*endo* and associated forms are involved in RNA reactivity, which could therefore promote interaction between proteins and RNA.[72,80]

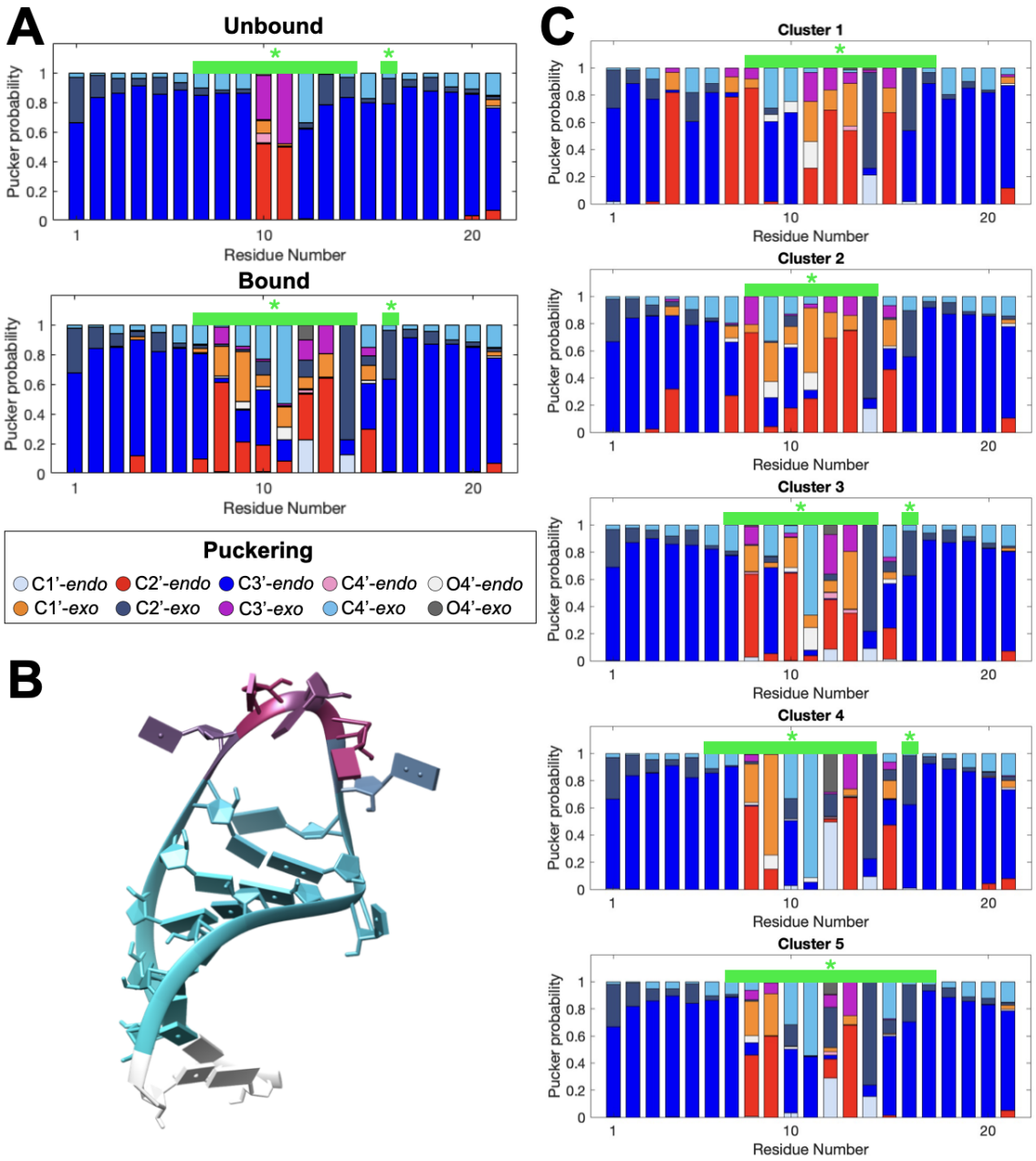

Figure 5: A. Ribose pucker probability of RNA unbound (top) and bound (bottom) for 1RKJ represented with stacked bar plot. B. RNA structure colored by interface contact for 1RKJ. White: not interface residues, cyan: residues at interface with a low frequency and pink: residues at interface with high frequency. C. Ribose pucker probability of RNA for each interface cluster of 1RKJ. For A and C, green bar and stars highlight the RNA residues involved in contacts with protein, with a frequency higher than or equal to 0.75. B-like family: Red: C2'-*endo*, Pink: C4'-*endo*, Orange red: C1'-*exo*, Magenta: C3'-*exo*. A-like family: Blue: C3'- endo, light blue: C1'-*endo*, dark blue: C2'-*exo*, cyan: C4'-*exo*. Light grey: O4'-*endo*, dark grey: O4'-*exo*.

The effect of the interface on the puckering distribution

Due to the results obtained between unbound and bound forms, we performed the same analysis on the complexes with several interface clusters. Figure 5 C shows the ribose pucker probability of RNA for the different interface clusters structures obtained on 1RKJ complex. As indicated by the green bar and stars, we observed a change in the residues implicated at the protein-RNA interface. We notably observe different distributions between clusters of sugar pucker conformations, with a majority in the B-like family for residues at interface. The same observation applies to the 1ASY, 1OOA and 1JBS complexes, as presented in Supplementary Figure S37.

Water analysis

To get more insights into the protein-RNA interfaces, we also evaluated the evolution of interfacial water molecules along the MD simulations, as well as the contacts between interface residues and interface water molecules.

Evolution of interfacial waters

We monitored the number of interfacial water molecule along time. Figure 6 displays the result for complex 1RKJ. The number of interfacial waters rapidly drops from around 100 to 75 during the first tens of nano-seconds of simulations and stays around this value until 250 ns, during which two sub-states are visited. Then we see two augmentations, corresponding each to a new interface substate. It appears that the interface substates correlate with the number of interfacial waters, contrary to what was seen with proteins.[39] The results for the eight others complexes are available in the Supplementary Figure S38. Similar to 1RKJ complex, a switch in the number of water molecules in contact with the residues at the interface was also observed for clusters of complexes 1ASY and 1JBS. However, the 1OOA complex, which has 4 clusters at the interface, has a quite variable number of water molecules (sd=20), but the link with the cluster is less obvious. Finally, complexes with a single cluster

(1MMS, 2R8S, 2VPL, 2ZM5 and 3IEV) generally show a low variability in their number of interfacial waters.

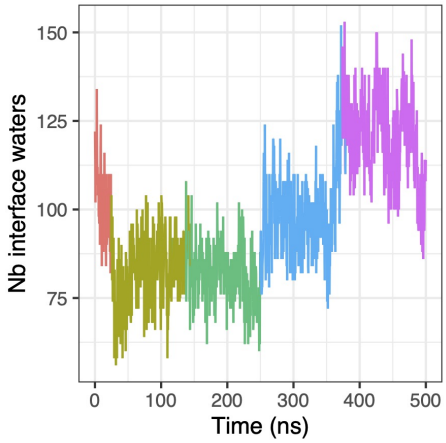

Figure 6: Number of interfacial water molecules in contact with interface residues, colored according to MD clusters for 1RKJ protein-RNA complex.

Contacts between interface residues and interface water molecules

We computed the number of interfacial water molecules in contact with interface residues at each time step of the MD trajectory. Figure 7 presents the distribution of this number, in each cluster of 1RKJ complex, for all the residues with a variable number of contacts interfacial waters (here, with a standard deviation over the trajectory greater than 2). Some residues are variable, irrespective of the interface substates. For example, residues Asn 106 and Asn 53, and RNA bases G 12 and G 16 have variable number of contacts with interface water molecules, but these variations are not connected to the interface substates. On the contrary, some residues have specific profiles depending of the clusters. For examples, residues Gly 175, Gly 7, Lys 174 and Met 4, and RNA bases C 20, C 21, C 7 have very contacts with interfaces water molecules in clusters 1-4, but high number of contacts in cluster 5. Thus, some interface residues have a variable number of contacts with interfacial waters throughout the simulations, while other residues exhibit specific signatures according to the substates. Same observation was made for the complexes with differents substates (1ASY, 1JBS and

100A), as shown in the Supplementary Figures S39-46. For the sake of completeness, single cluster complexes are also included in the Supplementary Figures.

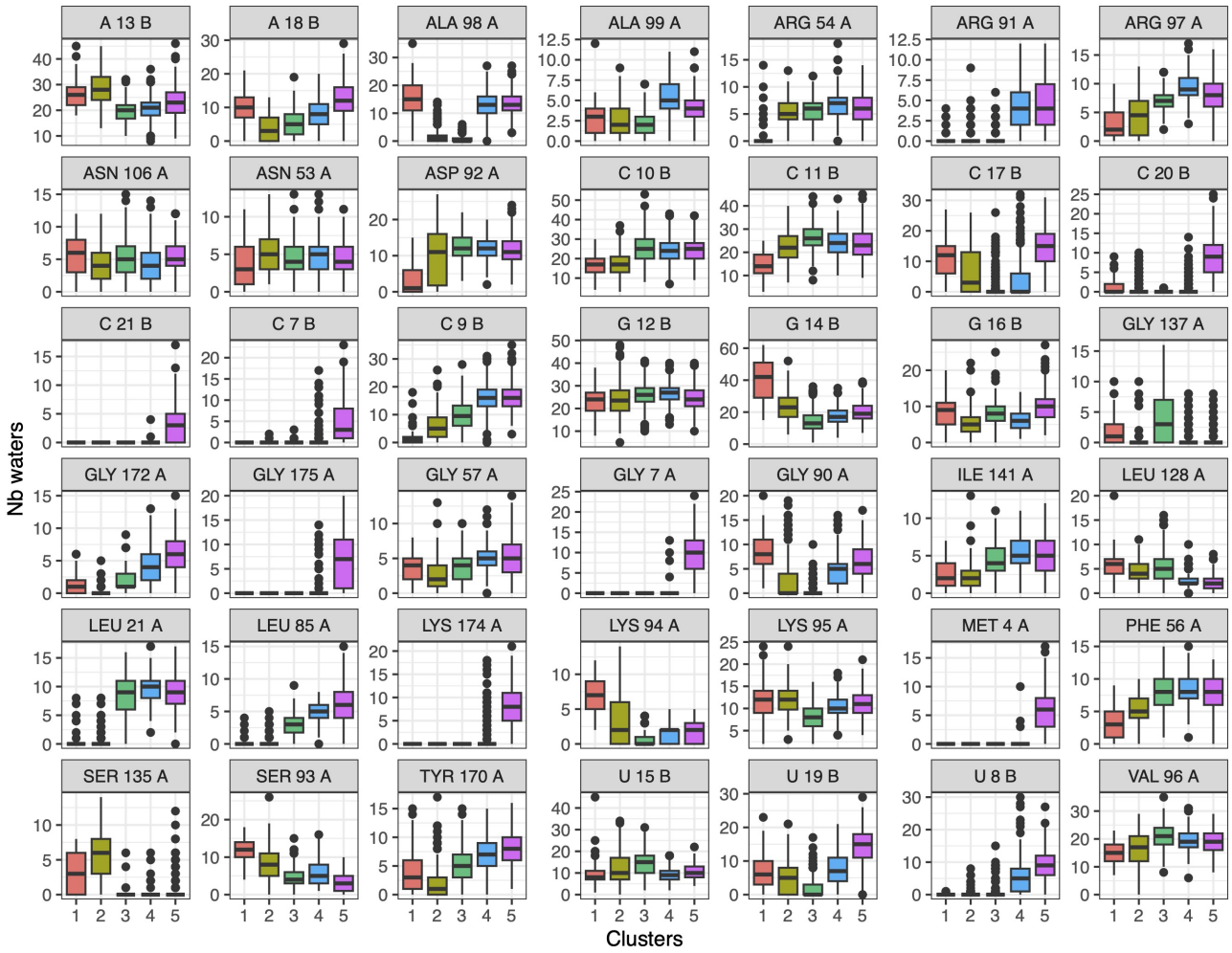

Figure 7: Number of interfacial water molecules in contact with interface residues, in each MD clusters for 1RKJ complex.

Conclusions

In this work, we aimed at investigating the dynamics of protein-RNA interfaces using molecular dynamics simulations. We performed all-atom MD simulations on nine complexes, in particular on their bound and unbound forms. In this context, we analyzed the change of the flexibility upon binding, the conservation of the initial contacts, the evolution of geometric

parameters and sugar pucker along the MD simulations, as well we determined the presence or the absence of interface substates using a clustering approach. Moreover, we evaluated the role of water molecules at the interface.

First, a change of flexibility upon binding is observed involving either only one biomolecule or both biomolecules. A rigidification or a flexibilization of the biomolecule(s) can be induced for the formation of the complex. We also detected a variability of the geometrical parameters and in their distributions, that can be narrow or large based on the complex. The sugar pucker analysis also points to the impact of the protein-RNA interaction on the sugar conformation of the latter, with the bound form moving to a more reactive ribose conformation for the residue at interface.

We also observed that at the scale of our simulations (up to a μs), almost half of RNA-protein complexes under investigation visit distinct interface substates. Interface substates exhibit altered interface contacts, either through direct interaction or facilitated by water molecules. These adjustments are accompanied by shifts in the arrangement of interface waters, which in turn influence the hydration levels of residues. It's crucial to emphasize that these alterations aren't limited to the edges of the interfaces; they also impact the residues at the core of the interface. The geometric properties between substates are notably influenced in a manner dependent on the specific case. This variation could lead to distinct complex stability and contribute to the reversibility of the interaction. The ribose pucker is also distributed differently between clusters, always with a majority for C2'-*endo* and associated B-like family forms.

To conclude, this work highlights that we cannot any longer consider protein-RNA interfaces as static objects, but we need to be take into account their dynamics to give a more complete picture of protein-RNA complexes. Hence, several descriptors are necessary to proper characterize the dynamical interfaces, as proposed here. Moreover, this work demonstrate that it can be interesting to investigate the sugar pucker for reactivity prediction. Finally, the number of water mediated-contacts observed in the simulations of the

complexes suggests that in protein-RNA docking, the presence of water molecules at the interface may open a new route to predict them.

Author Contributions

The study was designed by EF. AS, CH and EF performed all-atom MD simulations. AS, CH, JM and EF analyzed all-atom MD simulations. EF and JM supervised AS. EF supervised CH. All authors contributed to write the manuscript.

Funding

The Grand Équipement National De Calcul Intensif (GENCI) is acknowledged for funding this research with grants A0120711431 and A0100711431.

Acknowledgement

The authors thank all members of the "Structure et traduction des ARN viraux" team at CiTCoM for useful comments and suggestions all along the process.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Table of simulation time and interface parameters values; RMSD/RMSF profiles; figures of geometric interface parameters distributions; initial contact time series; principal component analysis projection; interface clustering results; contact variance matrix; puckering probability; interface water molecules.

Data and Software Availability

The MD simulations obtained for the unbound and bound structures without water are available in the repository Zenodo. Part of the interface analysis is conducted with in-house python scripts that are available at `https://github.com/juliettemartin/DYN_INTERFACES`.

References

(1) Mackereth, C. D.; Sattler, M. Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology* **2012**, *22*, 287–296, Nucleic acids/Sequences and topology.

(2) Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **2008**, *582*, 1977–1986.

(3) Hentze, M. W.; Castello, A.; Schwarzl, T.; Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **2018**, *19*, 327–341.

(4) Gerstberger, S.; Hafner, M.; Tuschl, T. A census of human RNA-binding proteins. *Nature Reviews Genetics* **2014**, *15*, 829–845.

(5) Gebauer, F.; Schwarzl, T.; Valcárcel, J.; Hentze, M. W. RNA-binding proteins in human genetic disease. *Nature Reviews Genetics* **2021**, *22*, 185–198.

(6) Varani, G.; Nagai, K. RNA RECOGNITION BY RNP PROTEINS DURING RNA PROCESSING. *Annual Review of Biophysics and Biomolecular Structure* **1998**, *27*, 407–445, PMID: 9646873.

(7) Lovci, M. T.; Bengtson, M. H.; Massirer, K. B. In *RNA Processing: Disease and Genome-wide Probing*; Yeo, G. W., Ed.; Springer International Publishing: Cham, 2016; pp 297–317.

(8) Deforges, J.; Locker, N.; Sargueil, B. mRNAs that specifically interact with eukaryotic ribosomal subunits. *Biochimie* **2015**, *114*, 48 – 57, Quality Control in Protein Synthesis.

(9) Leulliot, N.; Varani, G. Current topics in RNA- protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001**, *40*, 7947–7956.

(10) Mukherjee, S.; Bahadur, R. P. An account of solvent accessibility in protein-RNA recognition. *Scientific Reports* **2018**, *8*, 10546.

(11) Berman, H.; Olson, W.; Beveridge, D.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.; Srinivasan, A.; Schneider, B. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal* **1992**, *63*, 751–759.

(12) Coimbatore Narayanan, B.; Westbrook, J.; Ghosh, S.; Petrov, A. I.; Sweeney, B.; Zirbel, C. L.; Leontis, N. B.; Berman, H. M. The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Research* **2013**, *42*, D114–D122.

(13) Alvarado, L. J.; LeBlanc, R. M.; Longhini, A. P.; Keane, S. C.; Jain, N.; Yildiz, Z. F.; Tolbert, B. S.; D'Souza, V. M.; Summers, M. F.; Kreutz, C.; Dayie, T. K. Regio-Selective Chemical-Enzymatic Synthesis of Pyrimidine Nucleotides Facilitates RNA Structure and Dynamics Studies. *ChemBioChem* **2014**, *15*, 1573–1577.

(14) Barnwal, R. P.; Yang, F.; Varani, G. Applications of NMR to structure determination of RNAs large and small. *Archives of Biochemistry and Biophysics* **2017**, *628*, 42–56, Nuclear Magnetic Resonance.

(15) Marušič, M.; Toplishek, M.; Plavec, J. NMR of RNA - Structure and interactions. *Current Opinion in Structural Biology* **2023**, *79*, 102532.

(16) Yang, X.; Li, H.; Huang, Y.; Liu, S. The dataset for protein-RNA binding affinity. *Protein Science* **2013**, *22*, 1808–1811.

(17) Piehler, J. New methodologies for measuring protein interactions in vivo and in vitro. *Current opinion in structural biology* **2005**, *15*, 4–14.

(18) Jones, S. Protein-RNA interactions: a structural analysis. *Nucleic Acids Research* **2001**, *29*, 943–954.

(19) Bahadur, R. P.; Zacharias, M.; Janin, J. Dissecting protein–RNA recognition sites. *Nucleic Acids Research* **2008**, *36*, 2705–2716.

(20) Krüger, D. M.; Neubacher, S.; Grossmann, T. N. Protein–RNA interactions: structural characteristics and hotspot amino acids. *RNA* **2018**, *24*, 1457–1465.

(21) Han, K.; Nepal, C. PRI-modeler: extracting RNA structural elements from PDB files of protein–RNA complexes. *FEBS letters* **2007**, *581*, 1881–1890.

(22) Pérez-Cano, L.; Fernández-Recio, J. Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 25–35.

(23) Hudson, W. H.; Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology* **2014**, *15*, 749–760.

(24) Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. Protein-DNA interactions: a structural analysis11Edited by K. Nagai. *Journal of Molecular Biology* **1999**, *287*, 877–896.

(25) Harini, K.; Srivastava, A.; Kulandaisamy, A.; Gromiha, M. M. ProNAB: database for binding affinities of protein–nucleic acid complexes and their mutants. *Nucleic Acids Research* **2021**, *50*, D1528–D1534.

(26) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *WIREs RNA* **2017**, *8*, e1405.

(27) Sponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P., et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chemical reviews* **2018**, *118*, 4177–4338.

(28) Krepl, M.; Pokorná, P.; Mlýnský, V.; Stadlbauer, P.; Šponer, J. Spontaneous binding of single-stranded RNAs to RRM proteins visualized by unbiased atomistic simulations with a rescaled RNA force field. *Nucleic Acids Research* **2022**, *50*, 12480–12496.

(29) Reyes, C. M.; Kollman, P. A. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *Journal of Molecular Biology* **2000**, *297*, 1145–1158.

(30) Basu, S.; Alagar, S.; Bahadur, R. P. Unusual RNA binding of FUS RRM studied by molecular dynamics simulation and enhanced sampling method. *Biophysical Journal* **2021**, *120*, 1765–1776.

(31) Dreab, A.; Bayse, C. A. Molecular Dynamics Simulations of Reduced and Oxidized TFIIIA Zinc Fingers Free and Interacting with 5S RNA. *Journal of Chemical Information and Modeling* **2022**, *62*, 903–913.

(32) Malone, B.; Chen, J.; Wang, Q.; Llewellyn, E.; Choi, Y. J.; Olinares, P. D. B.; Cao, X.; Hernandez, C.; Eng, E. T.; Chait, B. T.; Shaw, D. E.; Landick, R.; Darst, S. A.; Campbell, E. A. Structural basis for backtracking by the SARS-CoV-2 replication–transcription complex. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2102516118.

(33) Chen, J.; Wang, Q.; Malone, B.; Llewellyn, E.; Pechersky, Y.; Maruthi, K.; Eng, E. T.; Perry, J. K.; Campbell, E. A.; Shaw, D. E.; Darst, S. A. Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication–transcription complex. *Nature Structural & Molecular Biology* **2022**, *29*, 250–260.

(34) Jandova, Z.; Vargiu, A. V.; Bonvin, A. M. J. J. Native or Non-Native Protein–Protein Docking Models? Molecular Dynamics to the Rescue. *Journal of Chemical Theory and Computation* **2021**, *17*, 5944–5954.

(35) Prévost, C.; Sacquin-Mora, S. Moving pictures: Reassessing docking experiments with a dynamic view of protein interfaces. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89*, 1315–1323.

(36) Etheve, L.; Martin, J.; Lavery, R. Dynamics and recognition within a protein–DNA complex: a molecular dynamics study of the SKN-1/DNA interaction. *Nucleic Acids Research* **2015**, *44*, 1440–1448.

(37) Etheve, L.; Martin, J.; Lavery, R. Protein–DNA interfaces: a molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic Acids Research* **2016**, *44*, 9990–10002 .

(38) Carzaniga, T.; Zanchetta, G.; Frezza, E.; Casiraghi, L.; Vanjur, L.; Nava, G.; Tagliabue, G.; Dieci, G.; Buscaglia, M.; Bellini, T. A bit stickier, a bit slower, a lot stiffer: Specific vs. nonspecific binding of gal4 to dna. *International Journal of Molecular Sciences* **2021**, *22*, 3813.

(39) Martin, J.; Frezza, E. A dynamical view of protein-protein complexes: Studies by molecular dynamics simulations. *Frontiers in Molecular Biosciences* **2022**, *9*, 970109.

(40) Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chemical Reviews* **2017**, *117*, 10694–10725.

(41) Li, Y.; Sutch, B. T.; Bui, H.-H.; Gallaher, T. K.; Haworth, I. S. Modeling of the Water Network at Protein–RNA Interfaces. *Journal of Chemical Information and Modeling* **2011**, *51*, 1347–1352.

(42) Barik, A.; Bahadur, R. P. Hydration of protein–RNA recognition sites. *Nucleic Acids Research* **2014**, *42*, 10148–10160.

(43) Chong, S.-H.; Ham, S. Anomalous Dynamics of Water Confined in Protein–Protein and Protein–DNA Interfaces. *The Journal of Physical Chemistry Letters* **2016**, *7*, 3967–3972.

(44) Treger, M.; Westhof, E. Statistical analysis of atomic contacts at RNA–protein interfaces. *Journal of Molecular Recognition* **2001**, *14*, 199–214.

(45) Bahadur, R. P.; Zacharias, M.; Janin, J. Dissecting protein–RNA recognition sites. *Nucleic Acids Research* **2008**, *36*, 2705–2716.

(46) Krepl, M.; Blatter, M.; Cléry, A.; Damberger, F. F.; Allain, F. H.; Sponer, J. Structural study of the Fox-1 RRM protein hydration reveals a role for key water molecules in RRM-RNA recognition. *Nucleic Acids Research* **2017**, *45*, 8046–8063.

(47) Huang, S.-Y.; Zou, X. A nonredundant structure dataset for benchmarking protein-RNA computational docking. *Journal of Computational Chemistry* **2013**, *34*, 311–318.

(48) Barik, A.; C., N.; P., M.; Bahadur, R. P. A protein–RNA docking benchmark (I): Nonredundant cases. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 1866–1871.

(49) Pérez-Cano, L.; Jiménez-García, B.; Fernández-Recio, J. A protein-RNA docking benchmark (II): Extended set from experimental and homology modeling data. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 1872–1882.

(50) Nithin, C.; Mukherjee, S.; Bahadur, R. P. A non-redundant protein–RNA docking

benchmark version 2.0. *Proteins: Structure, Function, and Bioinformatics* **2017**, *85*, 256–267.

(51) Fiser, A.; Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **2003**, *374*, 461–91.

(52) Rother, M.; Rother, K.; Puton, T.; Bujnicki, J. M. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Research* **2011**, *39*, 4007–4022.

(53) Lu, X.-J.; Bussemaker, H. J.; Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Research* **2015**, *43*, e142–e142.

(54) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic acids research* **2004**, *32*, W665–W667.

(55) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* **2011**, *7*, 525–537.

(56) Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **2001**, *7*, 306–317.

(57) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854, Place: England Publisher: Science for Life Laboratory, Stockholm and Uppsala, 171 21 Stockholm, Sweden.

(58) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008**, *4*, 435–447, Publisher: ACS Publications.

(59) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005**, *26*, 1701–1718, Publisher: Wiley Online Library.

(60) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

(61) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E. I.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation* **2011**, *7*, 2886–2902.

(62) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics* **2004**, *120*, 9665–9678.

(63) Allnér, O.; Nilsson, L.; Villa, A. Magnesium ion–water coordination and exchange in biomolecular simulations. *Journal of chemical theory and computation* **2012**, *8*, 1493–1502.

(64) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(66) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

(67) Harvey, S. C.; Tan, R. K.-Z.; Cheatham III, T. E. The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *Journal of computational chemistry* **1998**, *19*, 726–740.

(68) Berendsen, H. J.; Postma, J. P. M. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984**, *81*, 3684–3690.

(69) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 14101, Publisher: AIP Publishing.

(70) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.

(71) Tubiana, T.; Carvaillo, J.-C.; Boulard, Y.; Bressanelli, S. TTClust: a versatile molecular simulation trajectory clustering program with graphical summaries. *Journal of chemical information and modeling* **2018**, *58*, 2178–2182.

(72) Frezza, E.; Courban, A.; Allouche, D.; Sargueil, B.; Pasquali, S. The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study. *Methods* **2019**, *162*, 108–127.

(73) Hubbard, S. J. *NACCESS-Computer program*; London: Department of Biochemistry and Molecular Biology, University College London., 1992.

(74) Levy, E. D. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *Journal of molecular biology* **2010**, *403*, 660–670.

(75) Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics* **1995**, *13*, 323–330.

(76) Herwanda, R.; Murdiono, W. E.; Koesriharti, K. Aplikasi Nitrogen dan Pupuk Daun

terhadap Pertumbuhan dan Hasil Tanaman Bawang Merah (Allium cepa L. var. ascalonicum). Ph.D. thesis, Brawijaya University, 2017.

(77) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(78) Harris, C. R. et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(79) Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(80) McGinnis, J. L.; Dunkle, J. A.; Cate, J. H. D.; Weeks, K. M. The Mechanisms of RNA SHAPE Chemistry. *Journal of the American Chemical Society* **2012**, *134*, 6617–6624, PMID: 22475022.