Similarities and Differences in Ligand Binding to Protein and RNA Targets: The Case of Riboflavin

TITLE:

Similarities and Differences in Ligand Binding to Protein and RNA Targets: the Case of Riboflavin

AUTHOR NAMES

Stefano Bosio[1,2], Mattia Bernetti[1,2,], Walter Rocchia[3], Matteo Masetti[1]*

AUTHOR ADDRESS

[1] Department of Pharmacy and Biotechnology, Alma Mater Studiorum – University of Bologna, Via Belmeloro 6, 40126, Bologna, Italy

[2] Computational and Chemical Biology, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, I-16163 Genova, Italy

[3] Computational mOdelling of NanosCalE and bioPhysical sysTems (CONCEPT) Lab, Istituto Italiano di Tecnologia, Via Melen - 83, B Block, 16152 Genova, Italy

ABSTRACT

It is nowadays clear that RNA molecules can play active roles in several biological processes. As a result, an increasing number of RNAs are gradually being identified as potentially druggable targets. In particular, non-coding RNAs can adopt highly organized conformations that are suitable for drug binding. However, RNAs are still considered challenging targets due to their complex structural dynamics and high charge density. Thus, elucidating relevant features of drug-RNA binding is fundamental for advancing drug discovery. Here, by using Molecular Dynamics simulations, we compare key features of ligand binding to proteins with those observed in RNA. Specifically, we explore similarities and differences in terms of: i) conformational flexibility of the target, ii) electrostatic contribution to binding free energy, and iii) water and ligand dynamics. As a test case, we examine binding of the same ligand, namely riboflavin, to protein and RNA targets, specifically the riboflavin (RF) kinase and flavin mononucleotide (FMN) riboswitch. The FMN riboswitch exhibited enhanced fluctuations and explored a wider conformational space, compared to the protein target, underscoring the importance of RNA flexibility in ligand binding. Conversely, a similar electrostatic contribution to the binding free energy of riboflavin was found. Finally, greater stability of water molecules was observed in the FMN riboswitch compared to the RF kinase, possibly due to the different shape and polarity of the pockets.

TEXT:

INTRODUCTION

The growing awareness of the multifaceted role played by RNA in several functional and regulatory aspects of cellular life[1–4] is opening up novel possibilities for therapeutic intervention

on various diseases related to its dysregulation[5–8]. Aside from mRNA vaccines, whose impact on public health has been demonstrated by the success of the worldwide campaign against the SARS-CoV-2 virus, we can envisage two main routes of RNA-based therapeutics, namely oligonucleotide drugs[9–11] and small-molecule drugs[12]. Among the former, we recall antisense oligonucleotides (ASOs), short interfering RNA (siRNA), and aptamers. These strategies hold great promise and some FDA-approved examples already exist, such as the recent Spinraza ASO[13] and the Patisiran siRNA[14]. Nevertheless, shortcomings associated with intrinsic oligonucleotide instability due to rapid degradation by environmental nucleases, potential immunogenicity, and poor pharmacokinetic (PK) properties make their development particularly challenging. In this respect, the high tunability of small molecules[15], which are traditionally optimized via medicinal chemistry strategies to achieve favorable pharmacodynamic and PK profiles, represents an attractive alternative. In fact, recent studies are gradually pointing out that there is potential to achieve concrete results also through this therapeutic route[12,16–24], and the very recent FDA approval of the RNA-targeting small molecule drug Risdiplam[25], about 20 years after the Linezolid's[26], has renewed enthusiasm in the field. In spite of this, as of today, very few marketed small-molecule drugs targeting RNAs exist and, most importantly, it appears that none were identified through rational design approaches, but rather through phenotypic screenings[12,25]. This is related to the limited amount of known RNA binders and the paucity of experimentally-determined three-dimensional structures of RNA-ligand complexes, that hinder the application of structure-based drug-design practices and, most importantly, the rationalization of key features underlying RNA-small molecule interaction. Thus, a thorough characterization of this interaction is highly desirable, and could boost the advancement towards real possibilities in the design of small-molecule drugs targeting RNAs.

At present, several obstacles have been recognized against the identification of promising ligands. The high charge density in the RNA backbone, which limits the capability of forming hydrophobic binding sites[19], together with its lower chemical variability compared to proteins[27], challenge the idea of designing binders possessing high affinity and specificity. Nevertheless, there are examples of highly folded RNA molecules with complex tertiary structure that are capable to form cavities reminiscent of protein binding pockets, and that appear suitable to host ligands[12,28]. Additionally, structural rearrangements upon ligand binding, similar to those often observed in targeted proteins, were also reported for RNA molecules[19,29]. Therefore, focusing on these types of RNAs, including for instance riboswitches and ribozymes, may be more promising. Furthermore, RNA molecules display marked structural flexibility and complex conformational dynamics[30]. For this reason, they are better described as conformational ensembles, which however are challenging to characterize through the experimental methods commonly used in structural biology endeavours. In this respect, the combination of experiments and computer simulations has proved effective in many cases[31].

Nowadays, computational methods have become of routine application in drug discovery programs. In particular, molecular dynamics (MD) and molecular docking have historically been widely applied to identify small-molecule binders to protein targets[32]. Nevertheless, their application to RNA targets is still in its infancy and under investigation[33,34]. Molecular docking allows predicting the three dimensional binding mode of small-molecule ligands to biomolecular targets. While this computational approach was originally developed and optimized for protein targets, earlier attempts to extend its scope to ligand discovery for RNA targets, in particular riboswitches, gave promising results[35–37]. More recently, adaptation of the existing docking software and scoring functions, as well as the development of novel RNA-tailored ones, are

reflecting the growing interest in RNA-oriented drug discovery[33,34]. While molecular docking provides a static picture of the binding process, MD simulations are particularly suitable to capture dynamical features of biomolecules, and have proven remarkably useful for drug discovery scopes in many applications[32], ranging from the estimate of binding free energies and the mechanistic description of binding processes, to the identification of binding sites for orthosteric and allosteric ligands, and the characterization of protein conformational changes associated with drug binding[38–51]. In the RNA field, MD simulations have been mainly employed to get fundamental insights into RNA structural dynamics by accurately reconstructing conformational ensembles and to detect and improve force field deficiencies[31,52], while few examples of application in drug discovery contexts exist[53–55]. Furthermore, it remains unclear to what extent structure-based methodologies, which have evolved over the years for speeding up the drug discovery process with protein targets, can be applied to the case of RNA[33,34].

In this study, we aim to explore key properties underlying ligand binding in protein and RNA targets using MD simulations. In particular, we apply computational approaches that are established to analyze MD trajectory data of proteins and that have been less applied, if at all, to RNAs. To set on a common ground, we consider as a case study the riboflavin ligand, which is able to bind both protein and RNA targets, namely the riboflavin (RF) kinase enzyme and the flavin mononucleotide (FMN) riboswitch (Figure 1). To gain insight into target-ligand interactions in these systems, we inspect essential aspects, including the conformational space explored by the biomolecular targets, the role of electrostatic interactions in binding, and the dynamics of ligand and water molecules in the binding sites.

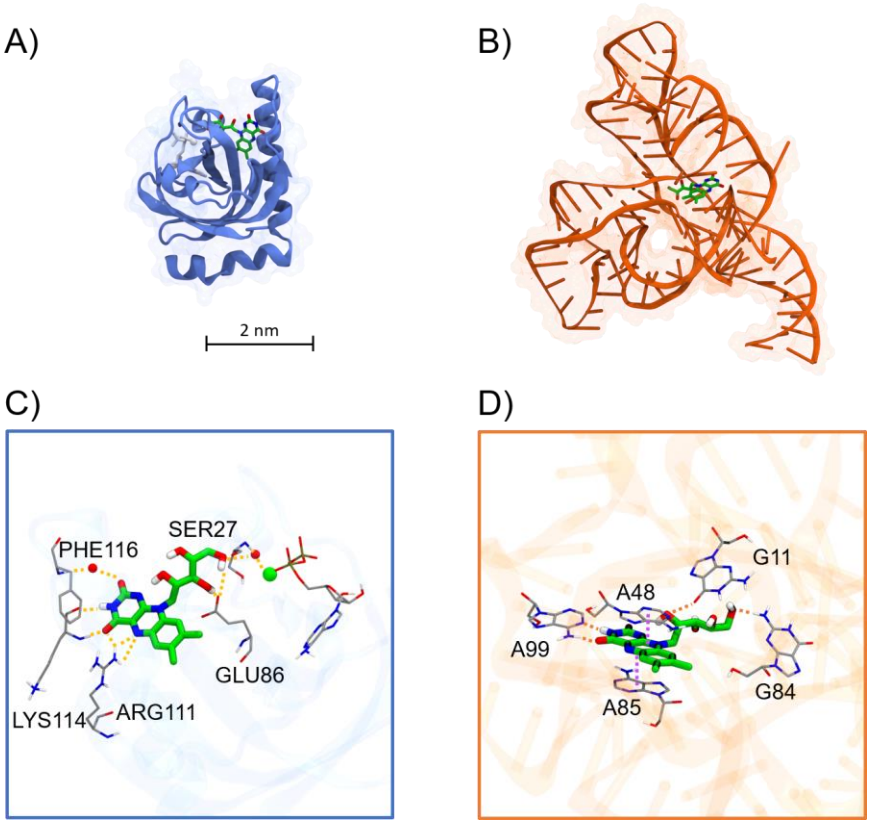

Figure 1. The systems considered in this study: (A) the riboflavin kinase (PDB ID 1NB9[56]) and (B) the FMN riboswitch (PDB ID 3F4G[57]). The protein and RNA molecules are shown in the NewCartoon representation of the Visual Molecular Dynamics (VMD) software[58], while the riboflavin ligand is displayed in licorice (with ligand carbon atoms highlighted in green). The two biomolecules are shown in scale. In the corresponding panels below, we report details of the (C) protein- and (D) RNA-ligand interactions in the crystal structures, with hydrogen bonds and π-π stackings shown as dashed lines in yellow and purple, respectively. The interacting residues from the target are labelled and are shown in licorice. In the RF kinase catalytic site (C), the ADP is shown in licorice, while the magnesium ion is shown in green.

METHODS

System setup and simulation protocol

Initial structures for each biomolecule were obtained from the protein data bank (PBD ID 1NB9[56] and 3F4G[57] for the kinase and riboswitch structure, respectively). The unbound states were obtained by removing the original ligand. All the structures were preliminarily processed using the Protein Preparation Wizard module of the Schrodinger suite[59,60]. In particular, missing sidechains were filled with Prime[61] and the protonation and tautomeric states of residues at pH 7.0 were assigned with Epik[60,62]. Simulation boxes were constructed using the standard Ambertools[63] and GROMACS modules by solvating the systems with OPC waters[64] and adding a 100 mM concentration of KCl. The KCl concentration was chosen according to the experimental conditions that are often used to study riboswitches[35,65–67], and that were used to measure the affinity (K_D) of the riboflavin ligand for the FMN riboswitch[57]. The same was used for the RF kinase for consistency.

The AMBER force field was used for protein (ff19sb)[68] and RNA (OL3)[69], the OPC[64] model was used for water molecules, and the parameters for magnesium were taken from Villa et al[70]. The RESP scheme[71] was applied to parametrize both the riboflavin in its neutral state, as predicted by Epik, and the ADP, using Gaussian 16[72] for computing the molecular electrostatic potential. Van Der Waals and electrostatic interactions were truncated at 12 Å, while long-range electrostatic interactions were treated with the particle-mesh Ewald method using a grid spacing of 1.6 Å.

Minimization of each system was performed with 5000 steps of steepest descent. The equilibration was carried out for each of the systems for a total of 1.15 ns, comprising gradual increase of the temperature to 300 K and equilibration of pressure at 1 bar. Production runs were performed in the NPT ensemble. The GROMACS 2021.3[73–75] MD engine was used for all the simulations with a timestep of 2 fs and the LINCS algorithm to costrain bonds[76]. The Nosé-Hoover

thermostat[77] was used to control the temperature at the reference value with a time costant for coupling of 1 ps, while constant pressure was maintained at 1 bar using the C-rescale barostat[78] with a time costant of 0.5 ps. For each of the two systems, three independent runs of 1 μs were performed in the bound and unbound states, amounting to a total simulation time of 12 μs.

Analysis of trajectories

The root mean square distance (RMSD)[79], the survival probability, and the hydrogen bond analyses were performed using modules available in the MDAnalysis[80,81] and MDtraj[82] python libraries. Dimensionality reduction on minimum distances between ligand and binding pocket residues was carried out through Principal Component Analysis (PCA) and t-SNE using the scikit-learn[83] implementations. To this end, binding pocket residues were identified by their proximity with the ligand during the simulation. Specifically, we selected residues having at least one heavy atom within 6 Å from the ligand for more than 20% of the simulation time. Minimum distances between binding pocket residues and ligand heavy atoms were then computed using MDTraj[82]. The POcket Volume MEasurer (POVME) algorithm version 2.2[49] was used to determine pocket volumes. The program fills a user-defined region with a grid of equidistant points. The algorithm then removes points that are in close proximity to atoms of the biomolecule and proceeds to calculate the volume of the remaining points. The Matplotlib[84] and Seaborn[85] libraries were used for visualization of the results and to produce all the plots in the manuscript. We make available a Jupiter notebook at github.com/CompMedChemLab/project_riboflavin, through which users can reproduce all of our analyses, results, and the plots reported in this work. The input files to perform the MD simulations in this work as well as the output simulation trajectories that we generated are freely available and deposited in Zenodo at http://doi.org/10.5281/zenodo.10797206.

Dimensionality reduction

Dimensionality reduction algorithms allow simplifying and summarizing features of complex datasets, and can be effectively used for analyzing and interpreting the highly dimensional information stored in MD trajectories. These algorithms transform high-dimensional data into a lower-dimensional space while preserving as much of the relevant information as possible. In the context of MD-generated data, they are used to capture most of the interesting features of the dynamics with few variables[86,87]. In this work, we used both the simplest linear method for dimensionality reduction, namely Principal Component Analysis (PCA), as well as the t-Distributed Stochastic Neighbor Embedding (t-SNE) method, a non-linear approach that is gaining increasing popularity and has been recently successfully applied to MD trajectory data[87]. We here summarize briefly the two methods.

The aim of PCA is to identify a set of orthogonal vectors (the Principal Components, PCs) along the D-dimensional feature space of the dataset $\mathbf{X}$ in which the variance of the data is maximized. The underlying steps can be summarized as follows: i) centering the data in order to ensure translational invariance of the projections thus obtaining the matrix $\widehat{\mathbf{X}}$, ii) computing the covariance matrix $\mathbf{C}$ as:.

$$\mathbf{C} = \frac{1}{N}\widehat{\mathbf{X}}^T\widehat{\mathbf{X}} \quad (1)$$

where N is the number of samples, and iii) extracting the eigenvectors of $\mathbf{C}$ by solving the eigenproblem. Eventually, dimensionality reduction is carried out through PCA by choosing a limited set of components $d \ll D$, and projecting the original data onto these vectors:

$$\mathbf{T} = \mathbf{XV} \quad (2)$$

where $\mathbf{V}$ is a matrix containing the d selected eigenvectors of $\mathbf{C}$.

Differently from PCA, t-SNE estimates the probability of each point to be a neighbor of every other point in the low-dimensional space using the distances in the high-dimensional one. The

algorithm then minimizes the divergence between these neighborhood probabilities in the original high-dimensional space and the corresponding probabilities in a lower-dimensional space, resulting in a set of projected coordinates that preserve the distribution of the data. The probability that point j is a neighbor of point i is calculated as:

$$P_{ij} = \frac{K_{ij}}{\sum_{k \neq j} K_{ik}} \quad (3)$$

where K_{ij} is a Gaussian kernel. The number of nearest neighbors is a free parameter that needs to be set by the user, and is referred to as the perplexity. The probabilities P_{ij} are transformed into a joint distribution by symmetrizing the matrix:

$$\bar{P}_{ij} = \frac{P_{ij} + P_{ji}}{2N} \quad (4)$$

The distribution of the points in the low-dimensional space is obtained through a Student's t-distribution with one degree of freedom, by computing the similarity between two points in the low-dimensional space as:

$$\bar{Q}_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq j} (1 + \|y_i - y_k\|^2)^{-1}} \quad (5)$$

The feature vectors y in the above equation can be initialized in different ways. In this work we exploited a PCA initialization, as implemented in scikit-learn[88]. The obtained low-dimensional distribution is compared with the original one by measuring their Kullback-Leibler divergence:

$$KL(\bar{P} \| \bar{Q}) = \sum_{j \neq i} \bar{P}_{ij} \, log \frac{\bar{P}_{ij}}{\bar{Q}_{ij}} \quad (6)$$

and this loss function is minimized through the steepest descent algorithm.

Water relaxation time

We estimated water residence time in the riboflavin binding sites using a survival probability (SP) function, which quantifies the relaxation time of the water molecules inside binding pockets. Specifically, SP measures the probability that a group of particles, herein the water molecules, will

remain in a specified region, herein the hydration shell of the binding cavity, for a certain amount of time t, and is defined as:

$$SP(t;r) = \frac{1}{T-t}\sum_{j=1}^{N_w(r)}\sum_{\tau=0}^{T-t} P_j(t_0, t_0+t; t^*) \quad (7)$$

In the above formula, T is the total simulation time, while $N_w(r)$ is the total number of water molecules, which is in turn a function of the radius of the hydration shell r[89–92] centered on the binding site. $P_j(t_0, t_0+t; t^*)$ is an indicator function that equals 1 if the jth water molecule found within the hydration shell at time t_0 has not left after a certain lag time t (i.e. in the time period between t_0 and t_0+t), except for a time $t^* \ll (t_0+t)$, and equals 0 otherwise. In our analyses, we considered a hydration shell with radius 10 Å centered in the binding pockets of both systems. To capture fine water dynamics with high statistical significance, we computed SP on MD trajectories saved every 5 ps. The calculation of SP was achieved via the Survival Probability module of the MDAnalysis package. For the calculation, we explored lag time values up to 1000 ps, and set t^*, referred to as intermittency in the MDAnalysis implementation, to 10 ps.

The trend of the SP function gives information about the local dynamics of water in the macromolecules' hydration shell, which herein encloses the binding sites. In particular, fitting the data with exponential functions allows computing the residence time τ of the water molecules[89,93,94]. Depending on the nature of water dynamics in the hydration shell, the time decay of the SP can be better approximated by a single exponential:

$$SP(t;r) = A\mathrm{e}^{-t/\tau} \quad (8)$$

or a double exponential function:

$$SP(t;r) = A\mathrm{e}^{-t/\tau_{\mathrm{fast}}} + B\mathrm{e}^{-t/\tau_{\mathrm{slow}}} \quad (9)$$

In the latter case, τ_{fast} and τ_{slow} are short and long time decay constants, indicating a dual lifetime regime for water molecules, such as a fraction of solvent molecules staying within the hydration

shell for prolonged periods of time and another entering and egressing rapidly[89,90]. A stretched exponential function (or Kohlrausch–Williams–Watts, KWW, function) can also be used [89,94]:

$$SP(t;r) = A\mathrm{e}^{-(t/\tau)^{\beta}} \quad (10)$$

where the stretching parameter β reflects the extent of nonexponential nature of the phenomenon. However, the interpretation is not straightforward as that of a single or double exponential. Nevertheless, it can provide a valuable estimate of the time scale over which the process evolves. Indeed, the diffusion of solvent molecules can be more complex than the description via single or double exponentials, as it may include multiple relaxation rates and may be better characterized by a distribution of rate constants.

Electrostatic contribution to the binding free energy

We estimated the electrostatic contribution to the binding free energy of the riboflavin ligand in both the RF kinase and FMN riboswitch systems. The electrostatic component to the binding free energy for a target-ligand system in a water buffer ($\Delta G_{\mathrm{binding}}^{\mathrm{ele},solv}$) can be computed according to a thermodynamic cycle as in Figure 2.

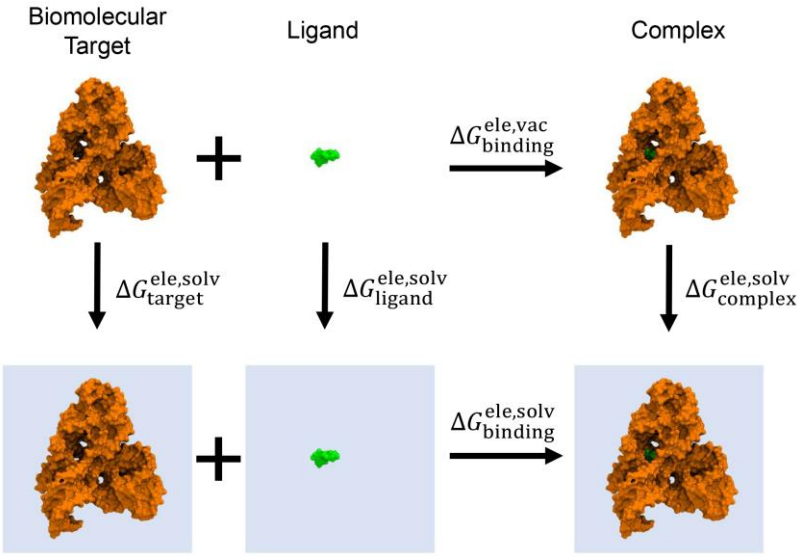

Figure 2. Schematic representation of the thermodynamic cycle considered for the calculation of the electrostatic contribution to the binding free energy. A generic ligand and biomolecular target are shown, with the top row indicating the binding process in vacuum, while the bottom row represents the event in solution. The arrows are labeled with the corresponding term.

Following the scheme of the thermodynamic cycle, we can write:

$$\Delta G_{\text{target}}^{\text{ele,solv}} + \Delta G_{\text{ligand}}^{\text{ele,solv}} + \Delta G_{\text{binding}}^{\text{ele},solv} - \Delta G_{\text{complex}}^{\text{ele,solv}} - \Delta G_{\text{binding}}^{\text{ele,vac}} = 0 \quad (11)$$

from which we can determine the electrostatic component to the binding free energy for a target-ligand system in a water buffer as:

$$\Delta G_{\text{binding}}^{\text{ele},solv} = \Delta G_{\text{complex}}^{\text{ele,solv}} + \Delta G_{\text{binding}}^{\text{ele,vac}} - (\Delta G_{\text{target}}^{\text{ele,solv}} + \Delta G_{\text{ligand}}^{\text{ele,solv}}) \quad (12)$$

In this way, the calculation of $\Delta G_{\text{binding}}^{\text{ele},solv}$ is decomposed in three main steps, where i) the isolated target and ligand in vacuum are transferred in solution ($\Delta G_{\text{target}}^{\text{ele,solv}} + \Delta G_{\text{ligand}}^{\text{ele,solv}}$), ii) the formed target-ligand complex in vacuum is transferred in solution ($\Delta G_{\text{complex}}^{\text{ele,solv}}$), and iii) the target-ligand complex is formed in vacuum ($\Delta G_{\text{binding}}^{\text{ele,vac}}$). Each of the ΔGs in the right-hand side of the above equation can be determined from differences between electrostatic contributions in the different media (vacuum or water buffer):

$$\Delta G_{\text{binding}}^{\text{ele},solv} = \left(G_{\text{complex}}^{\text{ele},solv} - G_{\text{complex}}^{\text{ele,vac}}\right) + \left(G_{\text{complex}}^{\text{ele,vac}} - \left(G_{\text{target}}^{\text{ele,vac}} + G_{\text{ligand}}^{\text{ele,vac}}\right)\right) - \left(\left(G_{\text{target}}^{\text{ele},solv} - G_{\text{target}}^{\text{ele,vac}}\right) + \left(G_{\text{ligand}}^{\text{ele},solv} - G_{\text{ligand}}^{\text{ele,vac}}\right)\right) \quad (13)$$

where $G^{\text{ele},solv}$ is the electrostatic energy of the species (target, ligand, or complex) evaluated in a water medium, and $G^{\text{ele,vac}}$ is the electrostatic energy of the species (target, ligand, or complex) evaluated in vacuum. Simplifying the equation, $\Delta G_{\text{binding}}^{\text{ele},solv}$ can be obtained as:

$$\Delta G_{\text{binding}}^{\text{ele},solv} = G_{\text{complex}}^{\text{ele},solv} - (G_{\text{target}}^{\text{ele},solv} + G_{\text{ligand}}^{\text{ele},solv}) \quad (14)$$

Each of these three terms was estimated on a pool of structures sampled from the MD simulations, selected through a cluster analysis procedure to represent as much as possible the conformational space explored. To this end, the k-means clustering algorithm was performed on a low dimensional embedding (see Dimensionality reduction section of the Methods) derived from the distances between residues in the binding pockets (selected from being in close proximity to the ligands). Thus, the centroids from the first fifteen most populated clusters were used in the estimation of the electrostatic contribution to the binding free energy. All the electrostatic calculations were performed by solving the linearized and non-linearized Poisson-Boltzmann equation, for the protein and RNA systems respectively, through the software Delphi version V5.1[95] patched with NanoShaper[96]. Delphi takes as input PQR files (i.e. modified pdb files, that contain also information about atom partial charge, Q, and radius, R), which we prepared using the radii and the charges reported in the force fields used in the MD simulations (i.e. ff19sb for the protein, OL3 for RNA, and GAFF2 for the ligand). The dielectric constant describing the water medium was set to 80, while a value of 2 was used to describe the electronic polarizability of the protein interior. An ionic atmosphere of 100 mM of monovalent salt was considered, and the linearized and non-linearized Poisson-Boltzmann equations were solved at the temperature of 298 K. Averages for the two systems were computed considering the respective cluster populations, and associated errors were estimated via bootstrap over the 15 values with 500 iterations. The energy extracted from DelPhi calculation includes the polarization contribution due to the jump in dielectric constant from the solute and the solvent, and the ionic contribution, due to the ionic atmosphere surrounding the solute. In the case of the non-linearized equation, also the electrostatic stress and the osmotic pressure terms are accounted for. In the Poisson-Boltzmann model, the energy is a free energy since it derives from averaging over all the solvent degrees of freedom

assuming the canonical ensemble. A further averaging is then done over different conformations coming from the molecular dynamics simulation. A 2 Å Stern layer was used in the calculations.

RESULTS AND DISCUSSION

In this study, we compare central features for target-ligand binding in protein and RNA targets of pharmaceutical relevance. To this end, we perform MD simulations and apply computational analyses that are established in MD simulations of protein-ligand systems, but have found less to no application in the emerging realm of RNA molecules. We focus on major issues in the drug discovery context, including the conformational dynamics of the biomolecular targets, the role of electrostatics in target-ligand binding, and the dynamics of ligand and water molecules within the binding sites. We take the riboflavin (RF) kinase enzyme and the flavin mononucleotide (FMN) riboswitch as case studies of protein and RNA targets, respectively (Figure 1). The RF kinase belongs to the well-known class of protein kinases, a prototypical and historical protein family within the drug discovery community, which has long been approached to identify potent and selective small molecule modulators. Conversely, the FMN riboswitch is a highly conserved RNA motif in bacteria, which has gained remarkable attention as a promising pharmaceutical target for the development of antibiotics[97]. Given its highly complex and stable folding, that allowed characterization of the 3D structure via X-ray christallography[57], and its natural disposition to bind metabolites, this riboswitch displays ideal prerequisites for establishing rational drug discovery campaigns. Indeed, Merck has designed a small molecule targeting the riboflavin binding site in the FMN riboswitch, and the compound is currently at the preclinical stage[24,97]. Riboflavin is the natural ligand of both RF kinase and FMN riboswitch, thus providing a common ground for comparing relevant features underlying target-ligand binding in these diverse macromolecules.

Enhanced conformational flexibility in RNA without appreciable pocket volume variation.

A thorough characterization of the structural dynamics of biomolecules is of paramount importance to understand their functioning. Furthermore, in the case of pharmaceutical targets, this can be crucial to identify fundamental features associated with ligand binding. For instance, assessing structural changes, e.g. by means of the RMSD, between bound and unbound states of a drug-target system can be informative about stabilizing effects induced by the presence of the ligand. Herein, for both the RF kinase and FMN riboswitch we observed an average RMSD on the whole structure ranging between 2.5 and 3 Å in both the ligand-bound and unbound states (Figure 3A, see Figure S1 for timeseries). This indicates that the overall folding was maintained by both systems during the MD simulations in the μs timescale, independently from presence of the riboflavin ligand. Nevertheless, the individual profiles in the four simulated systems demonstrated some noticeable differences. In the case of riboflavin-bound RF kinase, the RMSD shows a narrow distribution, which becomes slightly wider and tends to bimodality in the unbound protein. This suggests that the presence of the ligand may promote stability of the overall structure, as also confirmed by inspecting the root mean square fluctuation (RMSF, Figure S2). Interestingly, the same behavior is conserved and more marked in the FMN riboswitch. However, regardless of the presence of the riboflavin, the distributions of the RMSD in the FMN riboswitch are appreciably broader, demonstrating a global conformational flexibility that is enhanced in the RNA molecule with respect to RF kinase. Looking at the timeseries, a peak could be observed in the bound FMN riboswitch between 400 and 500 ns, which occurred once in only a single replicate. This was due to a transient interaction between the ends of the strands (see Figure S1), further demonstrating the significant structural flexibility allowed to the RNA structure despite a high level of folding achieved through its complex tertiary structure (see 3D structure in Figure 1B).

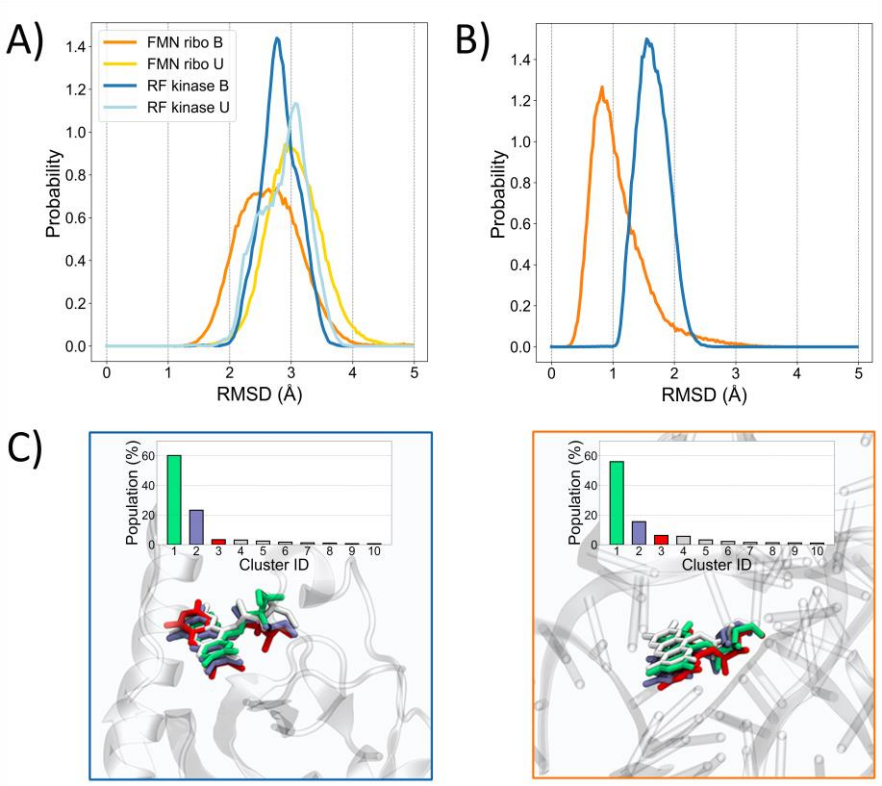

Figure 3. (A) RMSD of the RF kinase and the FMN riboswitch with respect to the starting structure of the MD simulations. The RMSDs were computed on the macromolecule heavy atoms and are shown as probability distributions by aggregating data from the three MD-simulation replicates performed on both ligand-bound (RF kinase B, FMN ribo B) and unbound systems (RF kinase U, FMN ribo U). (B) RMSD of riboflavin in the ligand-bound simulations with respect to the starting structure of the MD simulations. The RMSD, shown as probability density distribution, was computed after optimal alignment on binding pocket atoms with lowest root mean square fluctuation during the MD simulations. C) Superposition of the most populated configurations in the MD simulations (colored structures in green, purple and red, in decreasing cluster population order) with the corresponding one in the crystal (white structure, including the surrounding targets in transparent NewCartoon representation), for the riboflavin ligand bound to the RF kinase (left panel) and the FMN riboswitch (right panel).

Focusing on riboflavin in the target-ligand complexes, we observed a variable configurational space accessible to the ligands in the two biomolecular targets. In particular, we found remarkably higher RMSD values for the ligand in compex with the RF kinase compared to the complex with the FMN riboswitch (Figure 3B, see Figure S3 for timeseries). Interestingly, the RMSD distribution was broader in the RNA target, suggesting that a wider variety of configurations were accessible, but all closer to the reference, if compared with the RF kinase complex. In particular, looking at the timeseries, we noticed that the higher RMSD displayed in the protein complex was subsequent to the loss of interactions in the hydroxylic tail of the ligand with the magnesium atom in the binding pocket. This can also be noticed by visually comparing the most populated clusters (Figure 3C), where the centroid in the RF kinase complex exhibits major difference in the orientation of the hydroxylated tail, while the one in the FMN riboswitch has most of the difference due to rigid traslation of the ligand inside the pocket.

The structural dynamics of the residues comprised in the binding pocket can also play a remarkable role in the binding process. To highlight local conformational rearrangements in proximity of the riboflavin binding site, we focused on a subset of distances, namely those between residues belonging to the pockets (see Methods). Thus, we applied dimensionality reduction approaches to characterize the states visited in the MD simulations. We tested both linear and non-linear methods, namely the well-known principal component analysis (PCA) and the more recent t-distributed stochastic neighbor embedding (t-SNE), respectively. While PCA has found historical application in the analysis of MD trajectory data and has become of routine use nowadays[98], the exploitation of t-SNE on MD data is rather recent. The latter approach is particularly suited to disentangle complex manifolds associated with high-dimensional data, where correlations possibly vary in a non-linear manner. Thus, t-SNE might be particularly suited to deal with MD trajectory

data, such as MD simulations of biomolecules displaying high flexibility and complex structural dynamics. Indeed, very recently t-SNE has demonstrated effective in characterizing the conformational space explored by a selected set of intrinsically disordered proteins (IDPs), a notoriously challenging class of proteins in this respect[87]. Thus, we envisage promising potential in the study of RNA molecules, which pose comparable challenges from the computational standpoint[34]. We performed the dimensionality reductions on the MD simulation data of all the four systems (RF kinase and FMN riboswitch, both in the ligand-bound and unbound states), using as input features the distances between residues comprised in the binding pockets. In the RF kinase, the distribution in the low dimensional space obtained with PCA was highly crowded and partially overlapped between the the ligand-bound and unbound simulations (Figure S4). While this indicated that the conformational space sampled by the residues in the two cases was comparable to a certain extent, it was not possible to discriminate distinct conformations of the pocket. Conversely, in the FMN riboswitch there was a clearer separation between the simulations in the two states, and few clusters were visible (Figure S4). Interestingly, this indicated that for both the protein and RNA targets, the binding pockets explored distinct regions of configurational space in the ligand-bound and unbound states, and this behavior was consistent in the three replicates of simulation for each case. This suggested that presence of the riboflavin in the binding sites has the effect of noticeably reshaping the conformational preference of the pocket, with a sharper effect in the FMN riboswitch. Besides these considerations, on the whole, the results from the PCA were not particularly informative, as also quantified by the relatively low cumulative explained variance by the first two PCs (about 40% and 50% for the kinase and the riboswitch, respectively). Conversely, the crowding effect observed from the PCA analyses was remarkably attenuated by t-SNE, which had a better performance in separating clusters of conformations (Figure 4).

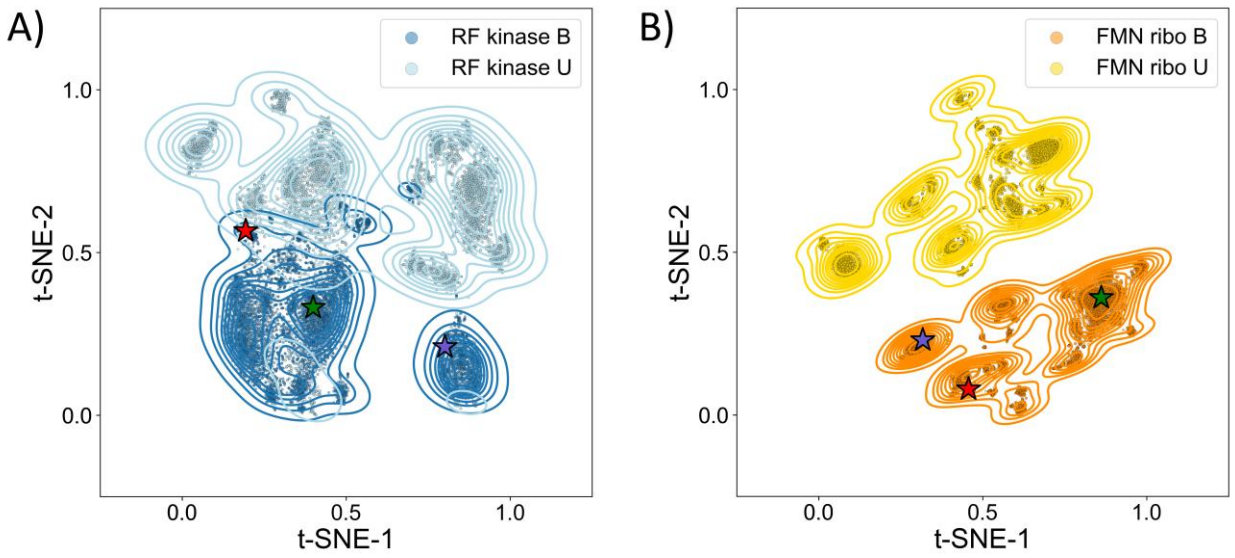

Figure 4. t-SNE space for (A) the RF kinase and (B) the FMN riboswitch, both in the ligand-bound and unbound states. The low dimensionality spaces were generated using as input features the minimum distances between residues within the protein and RNA binding sites. Contour lines are the kernel density estimation of the point distributions and better highlight regions of overlap between the ligand-bound and unbound systems. Colored stars indicate the location of ligand configurations in the clusters shown in Figure 3, using the same color code.

The projections obtained with t-SNE have the property to preserve the local neighborhood information, thus favoring a more straightforward interpretation of the conformational heterogeneity in each cluster, and possibly highlighting variations associated with ligand binding. Indeed, the overall picture reported by PCA was maintained, but the separation between different conformations of the pockets was remarkably improved through t-SNE. Interestingly, by highlighting the three most populated clusters of the ligand configurations in the binding site reveals that these belong to clearly distinct regions in the t-SNE space. In other words, different ligand configurations stabilize diverse pocket conformations, and the t-SNE approach was

sufficiently sensitive to the input features (distances between residues) to detect such different states of the pocket.

It is interesting to note that the sharp separation in the t-SNE space between the ligand-bound and unbound conformations of the pocket in the FMN riboswitch does not correspond to significant variation in the pocket volume (Figure 5B,D), which has a rather peaked and overlapping distribution (mean=637 Å^3 and 590 Å^3, and stdev=78 Å^3 and 104 Å^3 for the bound and unbound states, respectively).

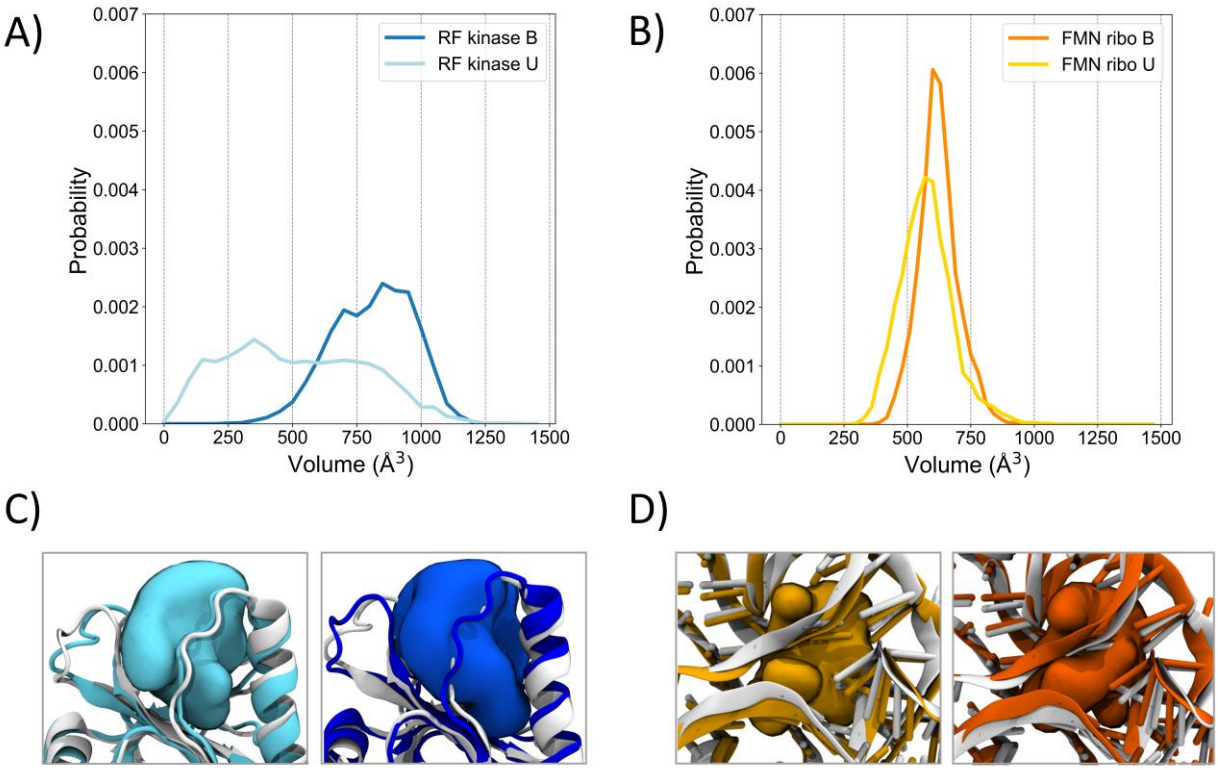

Figure 5. Distributions of the pocket volumes explored during the MD simulations of the (A) RF kinase and (B) FMN riboswitch, both in the bound and unbound states. In the bottom panels C) and D), the surfaces of the pockets for the different systems are also shown, color coded consistently with the corresponding panels A) and B). In C), the left panel has colored target structure and pocket surface for the unbound state of the protein, with the target structure in the

bound state in white for comparison, while the right panel has colored target structure and pocket surface for the bound state, with the target structure in the unbound state in white. The same scheme is followed in D) for the RNA target. The displayed snapshots correspond to trajectory frames with pocket volumes closest to the average values observed in the simulations.

This indicates that the volume of the pocket is poorly sensitive to the conformations accessible by the binding site, and the RNA molecule is able to experience appreciable conformational rearrangements within the binding site while preserving the size and overall architecture of the cavity. Additionally, this is intuitively in contrast with the enhanced flexibility displayed by the structure of the whole RNA molecule (see RMSD distribution in Figure 3). Conversely, the volume range covered by the binding site in the RF kinase was significantly wider. In particular, the distribution of the volume was remarkably different in the ligand-bound and unbound cases, with the former adopting larger values on average (mean=836 $Å^3$ and 533 $Å^3$ and stdev=157 $Å^3$ and 269 $Å^3$ for the bound and unbound states respectively). While compared to the FMN riboswitch the volume distribution was broader in both cases, in the unbound RF kinase it was almost uniform over the spanned range, highlighting a conformational exploration with no particular preference. Differently, a well-defined, narrower and Guassian-like distribution was observed in the simulations of the ligand-bound system. This comparison is interesting, as it suggests how the presence of the ligand is able to restrict the conformational dynamics of the binding site and stabilize specific conformations.

Such a divergent behavior between RF kinase and the FMN riboswitch can be attributed to both the different location and the biological role of the pockets in the two investigated systems. On the one hand, the more conserved volume of the pocket observed in the case of FMN riboswitch with

respect to RF kinase underscores the ability of RNA structures to form sizeable pockets that might be druggable. On the other hand, the enhanced conformational space might represent a challenge for conventional computational methods used in drug discovery that take into account only partially the conformational flexibility of the target. The known flexibility of RNA structure is seen as a major challenge for RNA-targeted drug discovery, as it may be difficult to take into account, for instance by selecting representative RNA conformations. This reasonably holds true for smaller RNA molecules, such as single stranded RNAs or short stem-loop motifs, which are indeed highly challenging to treat with conventional rational drug design approaches (mostly because they typically do not form clearly identifiable and stable binding sites). Nevertheless, the case of the FMN riboswitch studied herein showed how, notwithstanding a clearly appreciable structural flexibility of the RNA as a whole, the architecture of the ligand binding site was overall rather maintaned, while experiencing noticeable conformational rearrangments between the ligand-bound and unbound states. We note that in this case the FMN riboswitch binding site demonstrated even a higher stability with respect to RF kinase, and protein kinases represent well-known and tractable systems, though challenging in terms of selectivity. While these aspects are not sufficient, they nevertheless represent necessary prerequisites to establish rational drug discovery campaigns. The considerations herein may be highly system specific, however they provide a first indication on the feasibility of RNA-targeted drug discovery via standard computational means.

Comparable electrostatic contribution to ligand binding. Having investigated the structural and dynamical behavior of the two systems, looking both at the overall structure and with emphasis on the binding site, we then inspect the electrostatic properties of target-ligand binding. In

particular, we focus both on the electrostatic features of the targets and on the electrostatic contribution to the binding free energy. Notably, electrostatics is a key factor in shaping short- and long-ranged interactions, and has been proven to provide a significant contribution to drug binding in several cases[99–102]. RNA molecules are rich in negatively charged phosphate groups, and RNA nucleotides display relatively limited chemical variability compared to protein amino acids. This may result in non-obvious electrostatic profiles when it comes to ligand binding. Thus, since the RF kinase and the FMN riboswitch are bound by the same ligand, riboflavin, we investigated the differences in their binding to the same small molecule. We therefore first computed the electrostatic potential of the targets in implicit solvent (Figure 6A and 6B), as implemented in the DelPhi software (see Methods). As the figure shows, we observed a more pronounced negative electrostatic potential, broadly distributed over the FMN riboswitch surface. This was expected, and it is primarily due to the presence of the phosphate groups on the RNA backbone. In contrast, the RF kinase exhibited an overall electrostatic potential that was more balanced, showing regions of both locally concentrated and depleted charge density. Interestingly and consistently, these considerations on the two systems held true also in the two binding-site regions.

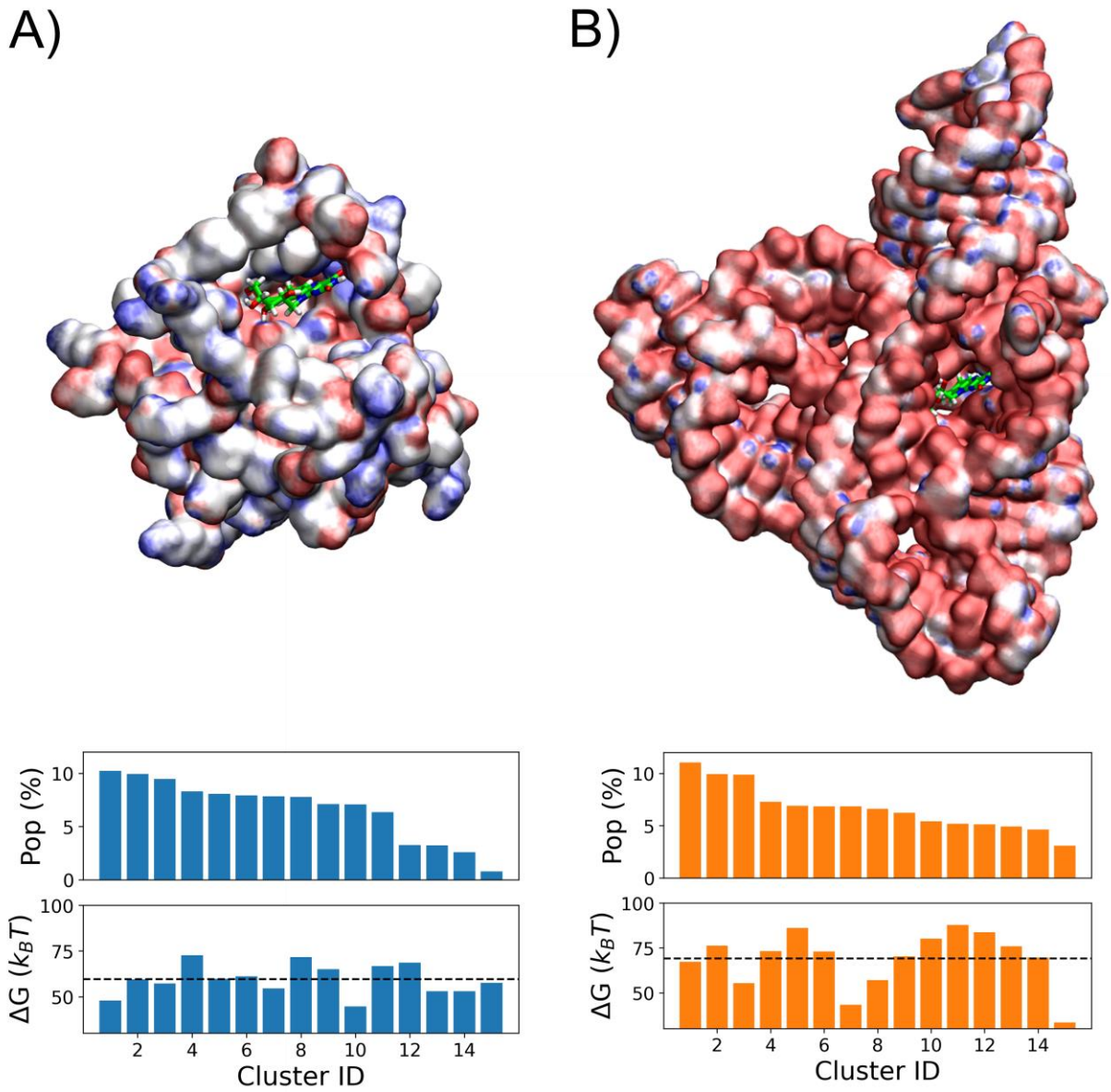

Figure 6. Electrostatic analysis on the RF kinase and the FMN riboswitch systems. The surface representation of the most populated cluster for the (A) protein and (B) RNA complexes is colored according to their electrostatic potential. The color scale is consistent for the two systems, and ranges from $-10\ k_{\mathrm{B}}T/e$ (red) to $+10\ k_{\mathrm{B}}T/e$ (blue). The bound riboflavin ligand is shown in licorice, with its carbon atoms highlighted in green. Below each system, we report the cluster population (upper panel) and the corresponding electrostatic component to the binding free energy in an implicit water buffer (lower panel, expressed in $k_{\mathrm{B}}T$).

To further obtain a more quantitative picture relative to ligand binding, we computed the electrostatic contribution to the binding free energy for riboflavin in complex with the two biomolecular targets (Figure 6, lower panels). To include possible local rearrangements of both the ligand and the surrounding residues in the targets' binding pockets, we repeated the calculation on representative conformations from our MD simulations, which we selected through a cluster analysis procedure (see Methods). The results of our calculations averaged over these representative conformations are 59.6 ± 2.2 and 69.1 ± 3.0 $k_{\mathrm{B}}T$ for the RF kinase and the FMN riboswitch, respectively, considering the respective cluster populations. Noatbly, the two systems displayed an overall similar contribution. This was unexpected, particularly considering the different electrostatic potentials displayed by the two target macromolecules. The value of the electrostatic contribution was slightly lower for the RF kinase system. Interestingly, this can be reconciled with the different affinity reported for the ligand towards the two biomolecular targets. Previous studies in the literature report K_{D}≈40 μM for riboflavin binding to the FMN riboswitch [57], and a K_{D} ≈1 μM for binding to RF kinase[103], indicating that the ligand displays higher affinity for the protein. We note that the latter K_{D} was measured on a bacterial RF kinase, while the present study focused on the human enzyme. Nevertheless, the two forms display high similarity in the RF kinase binding site (74%, see SI), suggesting that, in first approximation, we can reasonably consider this affinity approximately valid also for the human enzyme. Another remarkable aspect was that the electrostatic contribution was unfavorable in both cases. This suggests that the driving force for riboflavin binding to the two targets should be found elsewhere, for instance in non-polar interactions and/or indirect polar interactions.

Distinct pattern of hydrogen bonds and ligand conformations. The conformational space accessible to ligands inside binding sites and the possibility to establish multiple interactions with pockets' residues are remarkable aspects that can affect both thermodynamics and kinetics of the binding process. Within such a picture, the role played by water molecules can be of paramount importance, ranging from ligand and pocket (de)solvation to mediation of target-ligand specific interactions. As already discussed, a comparison of the conformational flexibility displayed by the riboflavin ligand in the RF kinase and FMN riboswitch binding sites from our MD simulations revealed that the ligand was less stable inside the FMN riboswitch binding pocket (broader RMSD distribution, see Figure 3B), despite remaining closer to its initial configuration with respect to the RNA. This observation can be further corroborated by inspecting the detailed network of ligand-target contacts and the internal degrees of freedom explored by the riboflavin ligand. Among the possible interactions, we focus in particular on hydrogen bonds, which explain most of the enthalpic contribution of drug-target binding in many cases. In the RF kinase the formation of hydrogen bonds was distributed over the whole structure of the ligand, and occurred for a significant fraction of simulation time, with most of the ligand heteroatoms mainly acting as acceptors (Figure 7A, leftmost panel).

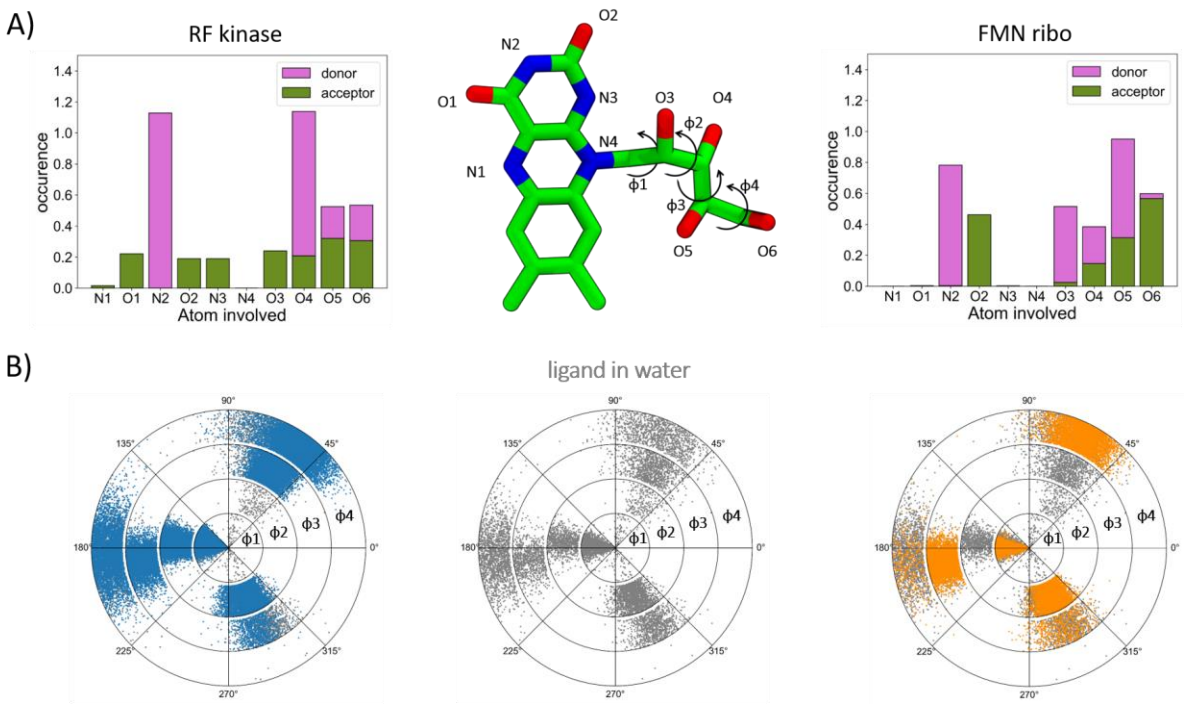

Figure 7. (A) Occurrence of hydrogen bonds formed by the riboflavin ligand with the RF kinase (leftmost panel) and the FMN riboswitch (rightmost panel) during the MD simulations. For each riboflavin heteroatom (central structure, only heavy atoms displayed and heteroatoms annotated with atom names), the two plots report the percentage occurrence of hydrogen bonds, highliting the fraction where the atom acts as a donor or acceptor. (B) Polar plots of the dihedrals in the riboflavin flexible tail (see the annotated structure in (A) for schematics). The central panel reports dihedral values computed from the MD simulation of riboflavin in water (grey points). In the leftmost and rightmost panels, these values are overlapped with the ones computed from the simulations of riboflavin in complex with the RF kinase (blue points) and FMN riboswitch (orange points), respectively.

Besides N2 in the aromatic moiety and O4 in the hydroxylated tail, that acted as donors for a significant fraction of the simulation, the other heteroatoms appear to be transiently involved in

hydrogen bonds. Comparing the results from the RF kinase to the ones of the RNA revealed that the occurrence of direct hydrogen bonds with the binding pocket was comparable on average (0.42 and 0.37, respectively). A closer inspection showed that the number of ligand heteroatoms involved in hydrogen bonding was lower when bound to the FMN riboswitch (Figure 7A: 6 atoms in the RNA, 8 in the protein). However, they all displayed significant occurrence, rather balanced between donors and acceptors. Interestingly, most of these hydrogen bonds involved the hydroxylic tail of the molecule (Figure 7A, right panel). While this can be highly case specific, it can be related to the different nature of the targets. The FMN riboswitch binding site is enriched with nucleobases, able to establish π-π interactions, and through which the ligand is able to intercalate its aromatic moiety (Figure 1D). Such intercalating behavior is common in known ligands binding to RNA targets[104,105], and the current space of RNA ligands is enriched with aromatic rings[106,107].

Interestingly, these interaction patterns observed in the two systems led to differential conformational preference of the ligand in the binding pockets. This behavior can be further inspected in terms of the different values of relevant dihedral angles explored by the ligand in the bound states (Figure 7B, blue and orange points for RF kinase and FMN riboswitch, respectively) during the simulations compared to those sampled in water (Figure 7B, grey points). Dihedral Φ1, which is the main determinant in the orientation of the tail with respect to the flavin moiety, shows only minor deviations, with the same dominant rotameric state displayed in all the simulations. Conversely, dihedral Φ2 undergoes a rotameric restriction upon binding, resulting in two almost equipopulated rotamers in the case of RF kinase and a single one in the FMN riboswitch, thus more marked for the RNA. Finally, in the RF kinase Φ3 and Φ4 dihedrals retain the possibility to cover the whole dihedral space even in the bound state, indicating an overall wider conformational

accessibility at the end of the tail. This can also be appreciated observing the most populated clusters of ligand configurations inside the pocket (Figure 3C, left panel). Thus, notwithstanding the higher degree of conformational variability of the riboflavin's tail in the RF kinase pocket, this did not preclude the establishment of hydrogen bonds. Differently, in the FMN riboswitch, while Φ4 retains the possibility to cover the entire dihedral space, although to a lower extent compared to RF kinase, the Φ3 dihedral experiences a significant rotameric restriction (Figure 7B, rightmost panel). Thus, the internal conformational space accessible to the ligand was more limited in the binding site of the FMN riboswitch. Here, a closer inspection of the ligand interaction network in the simulations revealed an interaction between the O4 in the ligand tail and a Mg^{2+} ions, mediated by two bridging water molecules, which was transiently maintained in the simulations (Figure S5), narrowing down even further the conformational space accessible to the ligand hydroxylated tail. Finally, this picture in the RNA system was compatible with the ligand RMSD with respect to the binding pocket, showing a broad distribution but closer to the reference (Figure 3B), demonstrating a semi-rigid rototranslational motion of the riboflavin in the FMN riboswitch binding site. Here, topological constraints associated with different size and shape of the pockets undoubtedly play a role. The FMN riboswitch adopts a highly complex tertiary structure, with an elaborated folding that is comparable to what can be observed in globular proteins. In this RNA, the ligand-occupied binding site is rather deep and of relatively low volume (637 $Å^3$), while the one in the RF kinase is more solvent-exposed, and adopts on average a larger volume (836 $Å^3$). This may be related to the biological role and functioning of the RF kinase, where the binding pocket is a catalytic site, and substrate binding needs to occur with high turnover.

Given the above considerations, in the case of the FMN riboswitch it is reasonable to expect a larger penalty contribution to the binding free energy due to a higher decrease of conformational

entropy upon binding compared to the RF kinase. At the same time, the riboflavin as a whole experiences larger fluctuations in terms of conformational variability within the FMN riboswitch binding pocket (Figure 3B, std=0.45 Å in the FMN riboswitch vs std=0.25 Å in the RF kinase). We expect that the entropic conformational penalty can be partially or entirely compensated by the higher residual rototranslational mobility (lower rototranslational entropic penalty), leading to a comparable net entropic contribution to binding for the two investigated systems.

Faster exchange rate of water in the bound protein binding site. Water molecules are central characters in drug-target binding and can affect both thermodynamics and kinetic aspects of this process. They play a key role in both ligand and binding site (de)solvation, and their contribution to favor ligand binding can be decisive in some cases[108–111]. We investigated the dynamics of the solvent molecules both in the bound and unbound states of the two biomolecular targets. In particular, we monitored the survival probability (SP) of waters in the binding sites, i.e. the amount of simulation time spent by water molecules in close proximity to the pockets' residues prior to being exchanged with the external buffer. The trend of the SP function provides information about the local dynamics of water in the considered region and allows estimating water residence time. Comparing the SP curves for the RF kinase and FMN riboswitch systems (Figure 8), we can clearly distinguish two main trends.

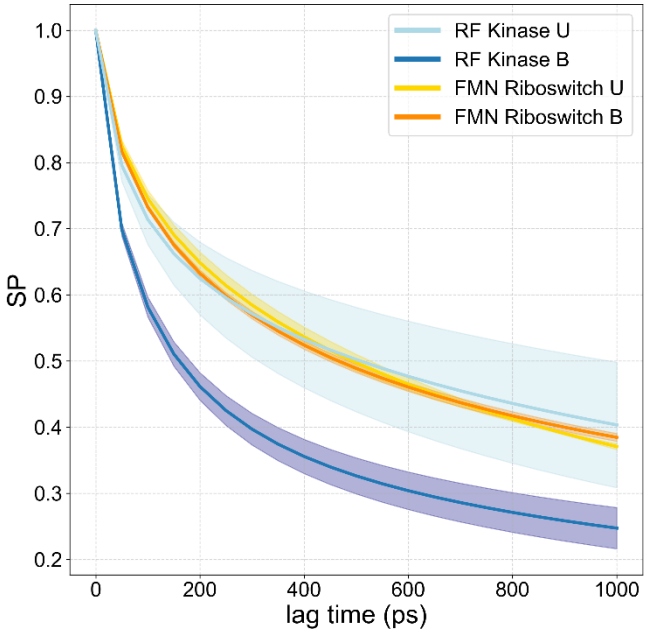

Figure 8. Survival probability of the water molecules in the binding site region of the RF kinase and FMN riboswitch systems from the MD simulations. Different colors are used for the two macromolecular systems and to distinguish simulations in complex with riboflavin (B, bound) or without the riboflavin ligand (U, unbound).

In particular, the RF kinase in the bound state displayed the fastest time decay, indicating a rapid and efficient exchange of water molecules in the binding region with the outer solvent buffer. Interestingly, compared to its unbound state, the riboflavin-bound RF kinase displayed a larger pocket volume on average (Figure 5A), which was also larger that the FMN riboswitch's one. Thus, the ligand presence stabilizes the binding pocket in conformations (see peaked distribution of the volume in Figure 5A) that offer wider pocket-solvent interface for exchange of water molecules (Figure S6A), which may result in the observed faster decay. Conversely, the FMN riboswitch showed a similar time decay for water SP in the bound and unbound cases, which is consistent with the similar pocket volume observed in the two forms (Figure 5B). Indeed, as

previously discussed, we have observed internal rearrangements of the pocket residues in the RNA (Figure 4B), while maintaining the overall volume of the binding site (Figure 5B).

The mean residence time of the water molecules can be estimated through fitting of the SP curves. We initially explored a single exponential function, which gave acceptable results (Figure S7A). However, when using a double exponential, the agreement was remarkably improved in all cases (Figure S7B). This is indicative of complex water dynamics, in a dual lifetime regime, for the water molecules within or in close proximity of the binding site. In particular, we could distinguish fast (τ_{fast}) and slow (τ_{slow}) residence times (Table 1).

Table 1. Residence times estimated through double exponential fitting of the SP function for the RF kinase and FMN riboswitch, both in the riboflavin-bound and unbound states.

System	τ_{fast} (ps)	τ_{slow} (ps)
RF kinase U	51.0	531.8
RF kinase B	48.1	344.2
FMN ribo U	56.3	582.4
FMN ribo B	55.6	469.8

The faster regime was characterized by values of τ_{fast} of about 50 ps in all cases, thus independent of the presence of the riboflavin ligand in the binding site. This observation suggests a rapid exchange between interfacial waters of the solvation shell considered with the surrounding buffer. Indeed, such timescale is compatible with residence time of mobile water molecules close to the biomolecule surface[112]. Conversely, the slower regime ranged approximately the 350-600 ps timescale, approaching residence time values ascribed to more buried waters in protein binding

sites[112,113]. Notably, the RNA in the bound and unbound and the protein in the unbound states displayed τ_{slow} around 500 ps, while the bound protein around 350 ps. This was unexpected as, intuitively, the presence of the ligand would retard water exchanges from the binding pocket, due to possible hydrogen bonding with the ligand, or to ligand steric hindrance reducing the mobility of other species within the cavity. However, as already mentioned, this faster exchange rate in the slow regime, i.e. lower τ_{slow} compared to the other systems, may be associated to the widest volume explored on average by the RF kinase in presence of the ligand, that favors exchange of water molecules with the surrounding solvent buffer. A further inspection of τ_{slow} revealed that the same trend can be observed between the protein and the RNA, with larger values in the unbound systems (τ_{slow}=344.2 and 531.8 ps for the RF kinase and τ_{slow}=469.8 and 582.4 ps for the FMN riboswitch, in the bound and unbound states, respectively), although the discrepancy was lower in the RNA case. For the latter, a visual inspection of the trajectories revealed that the pocket in the unbound state explored conformations where the access to the cavity was narrower (Figure S6B), partially hindering access and egress of water molecules and possibly causing the estimated lower exchange rates. Such modest flexibility of the pocket is compatible with a slightly broadened distribution of the sampled pocket volumes (Figure 5B) and, in particular, with the shift to lower volume values, indicating that also in the RNA case the presence of the ligand had a stabilizing effect on the pocket conformations. Finally, we note that compared to the riboflavin-bound RF kinase, water molecules in the FMN riboswitch exhibited in general higher persistency, independently from the presence of the ligand (Figure 8). Therefore, we can speculate that a ligand might display higher difficulty in displacing them, and this can have important implications for ligand binding free energies and association/dissociation rate constants, which should be taken into account in drug design endeavors.

Summary of main findings. In the following, we provide a brief summary of the major observations from each section of our study:

- The RF kinase and FMN riboswitch maintained the global folding in the MD simulations, with the RNA displaying higher conformational flexibility. The presence of the ligand stabilized the overall structure, particularly in the FMN riboswitch. Focusing on the binding site, the ligand in the RF kinase accessed distinct configurations, mainly rearranging the hydroxylated tail, while in the FMN riboswitch the variety of configurations visited, closer to the native pose, were primarily due to semi-rigid roto-translation. Additionally, we observed different conformations for the binding pockets, particularly between the ligand-bound and unbound states. Notably, in the FMN riboswitch such conformational variations did not correspond to appreciable differences in pocket volume between bound and unbound states. Interestingly, this was intuitively in contrast with the global flexibility displayed by the RNA structure. Differently, in the RF kinase the presence of the ligand restricted the conformational dynamics of the binding site. Finally, in the bound state of both protein and RNA we observed how different ligand configurations stabilized different pocket conformations.
- The electrostatic potential on the surface of the FMN riboswitch was mainly negative, as expected given the presence of the phosphate groups in the RNA. Conversely, a more balanced picture was found for the electrostatic potential of the RF kinase. Unexpectedly, particularly considering such discrepancy, we found a similar electrostatic contribution to the binding free energy for the protein and RNA systems. Notably, we observed that this

contribution was unfavorable in sign for both systems, suggesting that the driving force for riboflavin ligand binding to these two targets should be found elsewhere.

- We observed different patterns of hydrogen bonding and ligand conformations in the protein and RNA targets. Hydrogen-bonding interactions with the RF kinase binding site were rather uniformly distributed over the whole structure of the riboflavin ligand, with most of its heteroatoms acting as acceptors. Conversely, in the FMN riboswitch hydrogen bonds were mainly formed through the hydroxylated tail of the ligand, with heteroatoms balanced between donors and acceptors, while the aromatic moiety was involved in stacking interactions with the RNA target. Inspection of dihedral angle distributions of the ligand hydroxylated tail inside the RF kinase binding site revealed how riboflavin was able to explore a wide internal conformational space, comparable to being in the unbound state in solution. Differently, when bound to the FMN riboswitch the ligand displayed semi-rigid roto-translational motion and its internal conformational space was more resctricted.
- The bound RF kinase displayed rapid and efficient exchange of water molecules between the binding site and the buffer. Conversely, the time decay for water SP was comparable in the bound and unbound states of the FMN riboswitch. Inspection of the mean residence times of the waters in the binding pockets revealed two regimes. The faster one, with similar residence times of about 50 ps for all systems, was associated with rapid exchange between interfacial waters of the solvation shell and the surrounding buffer. The slower regime, in the 500 ps timescale, was ascribed to exchanges of more buried waters in the binding sites. In the case of the bound RF kinase, the exchange in this slower regime was the fastest among the four systems, and was attributed to the widest volume explored on average by the RF kinase in presence of the ligand.

These results have been obtained by applying computational approaches that are established to analyze MD trajectory data of protein targets and that have been less applied, if at all, to RNA molecules. Therefore, from a pragmatic standpoint, our work demonstrates that comparable analyses can be performed on protein and RNA targets. Despite focused on a limited set of systems, we have initially explored the applicability of these methods to an RNA molecule with highly folded structure and pharmaceutical relevance. In this sense, our study may serve as a practical guide in the selection of suitable computational approaches that demonstrated to be able to catch relevant features in RNA-ligand binding. In light of the relatively limited knowledge on the topic, this may prove particularly valuable to accelerate the set-up of innovative RNA-targeted drug discovery campaigns. Additionally, we provided further insights into RNA-ligand recognition, at the atomistic level. In this respect, the use of MD simulations to characterize specific interaction patterns and conformational preferences of small molecules inside RNA binding sites is of outstanding relevance in light of optimizing ligand binding. For instance, here we found how the RNA-ligand system was able to establish polar and non-covalent interactions as much as in the protein-ligand system, suggesting that rational drug design frameworks can possibly be applied also for RNA targets. Alike, we were able to explore and characterize the pocket conformational dynamics, which may reveal peculiar conformations that can be exploited for RNA ligand design. We believe that advancing the knowledge on these aspects is of fundamental importance to accelerate concrete RNA-targeted drug discovery.

CONCLUSIONS

RNA molecules are gradually emerging as promising pharmaceutical targets, opening the way to novel therapeutic avenues. Nevertheless, compared to more established protein targets, the current knowledge about their interaction with small molecue partners is still lagging behind. In an effort to provide further insight in this respect, we compare relevant aspects for target-ligand binding in pharmaceutically-relevant protein and RNA molecules, namely the RF kinase and the FMN riboswitch. Leveraging MD simulations and computational analyses, we focus on central issues including the conformational dynamics of the biomolecular targets, the role of electrostatic interactions, and the dynamics of ligand and water molecules in the binding sites. Firstly, our study demonstrated that comparable computational analyses can be performed on the protein and RNA targets. Furthermore, by discussing similarities and differences in the results from the two systems, we provided, to the best of our knowledge, a first comparison of key ligand-target binding properties in protein and RNA targets. These not only highlighted the applicability of the employed approaches, but also allowed us making valuable observations and draw conclusions, in a comparable way as it can be done for protein targets. Interestingly, our analyses shed further light on the capability of the RNA-ligand system to establish polar, non-covalent interactions to an extent comparable to ligand-binding protein targets.

We stress that the observed properties and the domain of applicability were explored on specific test cases and may be highly system dependent. For instance, we are aware that different RNA molecules can display different degrees of structural complexity, and here we picked an example possessing a high level of folding with a structurally complex and stable binding site, representing a favorable prerequisite to engage structure-based rational drug design efforts. While this is certainly not representative of the entire class of RNA molecules, nevertheless others of this kind may exist in the wide realm of non-coding, functional RNAs that are increasingly being identified.

Thus, our study represents a first resource reporting on the feasibility of applying relatively standard computational approaches in computer-aided drug design to treat suitable RNA targets, laying the ground to the use of molecular simulations towards successful RNA-targeted drug discovery.

ASSOCIATED CONTENT

Supporting Information. RMSD timeseries and distributions of the ligand from the MD simulations, RMSD timeseries and distributions of the targets from the MD simulations, RMSF of the biomolecular targets, Principal Component Analysis on the internal distances of the binding pockets, fitting of the SP curves with single and double exponentials, snapshot of the main entrance to FMN riboswitch binding pocket, timeseries of the distance between Mg^{2+} and O4 in the FMN Riboswitch simulations and the snapshot of the frames with the closest distance value to the mean of replica 1 and 2, sequence alignment of the human and bacterial RF kinases.

DataAvailabilityStatement. We freely provide the input files (initial coordinates, topologies, and GROMACS mdp parameter file) to perform the MD simulations in this work, as well as the output simulation trajectories (xtc format) that we generated, and the cluster centroids (pdbqt format) selected for the electrostatic analyses. We also supply a Jupyter notebook to reproduce all of our analyses, results, and the plots reported in this work. The material is freely available in Zenodo at http://doi.org/10.5281/zenodo.10797206. The Jupiter notebook can be straightforwardly consulted, and downloaded, at github.com/CompMedChemLab/project_riboflavin. All MD simulations were performed with GROMACS 2021.3[73–75], the VMD software version 1.9.3[58] was used for visualization, and analyses were conducted with MDanalysis version 2.6.1[80,81], MDtraj version 1.9.4[82], scikit-learn version 1.2[83], POVME version 2.2[49,114], and Delphi version version 5.1[95].

AUTHORS INFORMATION

CORRESPONDING AUTHOR

Mattia Bernetti, Computational and Chemical Biology, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, I-16163 Genova, Italy. Email: mattia.bernetti@iit.it

PRESENT ADDRESSES

AUTHOR CONTRIBUTIONS

S.B. performed the MD simulations, and conducted all the analyses together with M.B. The project was conceived by M.B. and developed under the supervision of M.B. and M.M. The electrostatic analyses were taken care of by W.R., who also provided constructive suggestions for the work. A first draft of the manuscript was written by S.B. and M.B., while the final version was produced through equal contribution of all authors. All authors have given approval to the final version of the manuscript.

FUNDING SOURCES

M.B. acknowledges funding by the project "National Center for Gene Therapy and Drugs based on RNA Technology" (CN00000041), financed by NextGenerationEU PNRR MUR – M4C2 – Action 1.4 - Call "Potenziamento strutture di ricerca e di campioni nazionali di R&S" (CUP: J33C22001130001).

NOTES

ACKNOWLEDGMENT

We acknowledge the Data Science and Computation Facility at IIT for computing time and support on the Franklin HPC system.

ABBREVIATIONS

REFERENCES

(1) Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **2014**, *157* (1), 77–94. https://doi.org/10.1016/j.cell.2014.03.008.

(2) Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A. Insights into RNA Structure and Function from Genome-Wide Studies. *Nat Rev Genet* **2014**, *15* (7), 469–479. https://doi.org/10.1038/nrg3681.

(3) Yao, R.-W.; Wang, Y.; Chen, L.-L. Cellular Functions of Long Noncoding RNAs. *Nat Cell Biol* **2019**, *21* (5), 542–551. https://doi.org/10.1038/s41556-019-0311-8.

(4) Rinn, J. L.; Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **2012**, *81* (1), 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902.

(5) Matsui, M.; Corey, D. R. Non-Coding RNAs as Drug Targets. *Nat Rev Drug Discov* **2017**, *16* (3), 167–179. https://doi.org/10.1038/nrd.2016.117.

(6) Winkle, M.; El-Daly, S. M.; Fabbri, M.; Calin, G. A. Noncoding RNA Therapeutics — Challenges and Potential Solutions. *Nat Rev Drug Discov* **2021**, *20* (8), 629–651. https://doi.org/10.1038/s41573-021-00219-z.

(7) Zhu, Y.; Zhu, L.; Wang, X.; Jin, H. RNA-Based Therapeutics: An Overview and Prospectus. *Cell Death Dis* **2022**, *13* (7), 1–15. https://doi.org/10.1038/s41419-022-05075-2.

(8) Wang, F.; Zuroske, T.; Watts, J. K. RNA Therapeutics on the Rise. *Nature Reviews Drug Discovery* **2020**, *19* (7), 441–442. https://doi.org/10.1038/d41573-020-00078-0.

(9) Damase, T. R.; Sukhovershin, R.; Boada, C.; Taraballi, F.; Pettigrew, R. I.; Cooke, J. P. The Limitless Future of RNA Therapeutics. *Frontiers in Bioengineering and Biotechnology* **2021**, *9*.

(10) Khvorova, A.; Watts, J. K. The Chemical Evolution of Oligonucleotide Therapies of Clinical Utility. *Nat Biotechnol* **2017**, *35* (3), 238–248. https://doi.org/10.1038/nbt.3765.

(11) Smith, C. I. E.; Zain, R. Therapeutic Oligonucleotides: State of the Art. *Annu Rev Pharmacol Toxicol* **2019**, *59*, 605–630. https://doi.org/10.1146/annurev-pharmtox-010818-021050.

(12) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat Rev Drug Discov* **2018**, *17* (8), 547–558. https://doi.org/10.1038/nrd.2018.93.

(13) Neil, E. E.; Bisaccia, E. K. Nusinersen: A Novel Antisense Oligonucleotide for the Treatment of Spinal Muscular Atrophy. *The Journal of Pediatric Pharmacology and Therapeutics* **2019**, *24* (3), 194–203. https://doi.org/10.5863/1551-6776-24.3.194.

(14) Ledford, H. Gene-Silencing Technology Gets First Drug Approval after 20-Year Wait. *Nature* **2018**, *560* (7718), 291–292. https://doi.org/10.1038/d41586-018-05867-7.

(15) Umuhire Juru, A.; Hargrove, A. E. Frameworks for Targeting RNA with Small Molecules. *J Biol Chem* **2021**, *296*, 100191. https://doi.org/10.1074/jbc.REV120.015203.

(16) Yu, A.-M.; Choi, Y. H.; Tu, M.-J. RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. *Pharmacol Rev* **2020**, *72* (4), 862–898. https://doi.org/10.1124/pr.120.019554.

(17) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chemical Biology* **2016**, *23* (9), 1077–1090. https://doi.org/10.1016/j.chembiol.2016.05.021.

(18) Falese, J. P.; Donlic, A.; Hargrove, A. E. Targeting RNA with Small Molecules: From Fundamental Principles towards the Clinic. *Chem. Soc. Rev.* **2021**, *50* (4), 2224–2243. https://doi.org/10.1039/D0CS01261K.

(19) Menichelli, E.; Lam, B. J.; Wang, Y.; Wang, V. S.; Shaffer, J.; Tjhung, K. F.; Bursulaya, B.; Nguyen, T. N.; Vo, T.; Alper, P. B.; McAllister, C. S.; Jones, D. H.; Spraggon, G.; Michellys, P.-Y.; Joslin, J.; Joyce, G. F.; Rogers, J. Discovery of Small Molecules That Target a Tertiary-Structured RNA. *Proceedings of the National Academy of Sciences* **2022**, *119* (48), e2213117119. https://doi.org/10.1073/pnas.2213117119.

(20) Suresh, B. M.; Taghavi, A.; Childs-Disney, J. L.; Disney, M. D. Fragment-Based Approaches to Identify RNA Binders. *J. Med. Chem.* **2023**, *66* (10), 6523–6541. https://doi.org/10.1021/acs.jmedchem.3c00034.

(21) Costales, M. G.; Childs-Disney, J. L.; Haniff, H. S.; Disney, M. D. How We Think about Targeting RNA with Small Molecules. *J. Med. Chem.* **2020**, *63* (17), 8880–8900. https://doi.org/10.1021/acs.jmedchem.9b01927.

(22) Fedorova, O.; Jagdmann, G. E.; Adams, R. L.; Yuan, L.; Van Zandt, M. C.; Pyle, A. M. Small Molecules That Target Group II Introns Are Potent Antifungal Agents. *Nat Chem Biol* **2018**, *14* (12), 1073–1078. https://doi.org/10.1038/s41589-018-0142-0.

(23) Di Giorgio, A.; Duca, M. Synthetic Small-Molecule RNA Ligands: Future Prospects as Therapeutic Agents. *Medchemcomm* **2019**, *10* (8), 1242–1255. https://doi.org/10.1039/c9md00195f.

(24) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M. R.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting RNA Structures with Small Molecules. *Nat Rev Drug Discov* **2022**, *21* (10), 736–762. https://doi.org/10.1038/s41573-022-00521-4.

(25) Sheridan, C. First Small-Molecule Drug Targeting RNA Gains Momentum. *Nature Biotechnology* **2021**, *39* (1), 6–8. https://doi.org/10.1038/s41587-020-00788-1.

(26) Leach, K. L.; Brickner, S. J.; Noe, M. C.; Miller, P. F. Linezolid, the First Oxazolidinone Antibacterial Agent. *Annals of the New York Academy of Sciences* **2011**, *1222* (1), 49–54. https://doi.org/10.1111/j.1749-6632.2011.05962.x.

(27) Bagnolini, G.; Luu, T. B.; Hargrove, A. E. Recognizing the Power of Machine Learning and Other Computational Methods to Accelerate Progress in Small Molecule Targeting of RNA. *RNA* **2023**, *29* (4), 473–488. https://doi.org/10.1261/rna.079497.122.

(28) Hewitt, W. M.; Calabrese, D. R.; Schneekloth, J. S. Evidence for Ligandable Sites in Structured RNA throughout the Protein Data Bank. *Bioorganic & Medicinal Chemistry* **2019**, *27* (11), 2253–2260. https://doi.org/10.1016/j.bmc.2019.04.010.

(29) Campagne, S.; Boigner, S.; Rüdisser, S.; Moursy, A.; Gillioz, L.; Knörlein, A.; Hall, J.; Ratni, H.; Cléry, A.; Allain, F. H.-T. Structural Basis of a Small Molecule Targeting RNA for a Specific Splicing Correction. *Nat Chem Biol* **2019**, *15* (12), 1191–1198. https://doi.org/10.1038/s41589-019-0384-5.

(30) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The Roles of Structural Dynamics in the Cellular Functions of RNAs. *Nat Rev Mol Cell Biol* **2019**, *20* (8), 474–489. https://doi.org/10.1038/s41580-019-0136-0.

(31) Bernetti, M.; Bussi, G. Integrating Experimental Data with Molecular Simulations to Investigate RNA Structural Dynamics. *Current Opinion in Structural Biology* **2023**, *78*, 102503. https://doi.org/10.1016/j.sbi.2022.102503.

(32) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684.

(33) Manigrasso, J.; Marcia, M.; De Vivo, M. Computer-Aided Design of RNA-Targeted Small Molecules: A Growing Need in Drug Discovery. *Chem* **2021**, *7* (11), 2965–2988. https://doi.org/10.1016/j.chempr.2021.05.021.

(34) Bernetti, M.; Aguti, R.; Bosio, S.; Recanatini, M.; Masetti, M.; Cavalli, A. Computational Drug Discovery under RNA Times. *QRB Discovery* **2022**, *3*, e22. https://doi.org/10.1017/qrd.2022.20.

(35) Daldrop, P.; Reyes, F. E.; Robinson, D. A.; Hammond, C. M.; Lilley, D. M.; Batey, R. T.; Brenk, R. Novel Ligands for a Purine Riboswitch Discovered by RNA-Ligand Docking. *Chemistry & Biology* **2011**, *18* (3), 324–335. https://doi.org/10.1016/j.chembiol.2010.12.020.

(36) Daldrop, P.; Brenk, R. Structure-Based Virtual Screening for the Identification of RNA-Binding Ligands. In *Therapeutic Applications of Ribozymes and Riboswitches: Methods and Protocols*; Lafontaine, D., Dubé, A., Eds.; Humana Press: Totowa, NJ, 2014; pp 127–139. https://doi.org/10.1007/978-1-62703-730-3_10.

(37) Colizzi, F.; Lamontagne, A.-M.; Lafontaine, D. A.; Bussi, G. Probing Riboswitch Binding Sites with Molecular Docking, Focused Libraries, and In-Line Probing Assays. In *Therapeutic Applications of Ribozymes and Riboswitches: Methods and Protocols*; Lafontaine, D., Dubé, A., Eds.; Humana Press: Totowa, NJ, 2014; pp 141–151. https://doi.org/10.1007/978-1-62703-730-3_11.

(38) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703. https://doi.org/10.1021/ja512751q.

(39) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127* (8), 2600–2607. https://doi.org/10.1021/ja0445950.

(40) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular Basis of Cyclooxygenase Enzymes (COXs) Selective Inhibition. *Proc Natl Acad Sci U S A* **2010**, *107* (12), 5411–5416. https://doi.org/10.1073/pnas.0913377107.

(41) Favia, A. D.; Masetti, M.; Recanatini, M.; Cavalli, A. Substrate Binding Process and Mechanistic Functioning of Type 1 11β-Hydroxysteroid Dehydrogenase from Enhanced Sampling Methods. *PLOS ONE* **2011**, *6* (9), e25375. https://doi.org/10.1371/journal.pone.0025375.

(42) Provasi, D.; Bortolato, A.; Filizola, M. Exploring Molecular Mechanisms of Ligand Recognition by Opioid Receptors with Metadynamics. *Biochemistry* **2009**, *48* (42), 10020–10029. https://doi.org/10.1021/bi901494n.

(43) Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V. Mechanistic Insight into Ligand Binding to G-Quadruplex DNA. *Nucleic Acids Res* **2014**, *42* (9), 5447–5455. https://doi.org/10.1093/nar/gku247.

(44) Bernetti, M.; Masetti, M.; Recanatini, M.; Amaro, R. E.; Cavalli, A. An Integrated Markov State Model and Path Metadynamics Approach To Characterize Drug Binding Processes. *J. Chem. Theory Comput.* **2019**, *15* (10), 5689–5702. https://doi.org/10.1021/acs.jctc.9b00450.

(45) Bertazzo, M.; Bernetti, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Fully Flexible Docking via Reaction-Coordinate-Independent Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2018**, *58* (2), 490–500. https://doi.org/10.1021/acs.jcim.7b00674.

(46) Kuzmanic, A.; Bowman, G. R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Acc Chem Res* **2020**, *53* (3), 654–661. https://doi.org/10.1021/acs.accounts.9b00613.

(47) Aguti, R.; Bernetti, M.; Bosio, S.; Decherchi, S.; Cavalli, A. On the Allosteric Puzzle and Pocket Crosstalk through Computational Means. *The Journal of Chemical Physics* **2023**, *158* (16), 165101. https://doi.org/10.1063/5.0145364.

(48) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: Open-Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27* (23), 3276–3285. https://doi.org/10.1093/bioinformatics/btr550.

(49) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. POVME: An Algorithm for Measuring Binding-Pocket Volumes. *Journal of Molecular Graphics and Modelling* **2011**, *29* (5), 773–776. https://doi.org/10.1016/j.jmgm.2010.10.007.

(50) La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **2017**, *3* (9), 949–960. https://doi.org/10.1021/acscentsci.7b00211.

(51) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proceedings of the National Academy of Sciences* **2013**, *110* (16), 6358–6363. https://doi.org/10.1073/pnas.1303186110.

(52) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.

(53) Stelzer, A. C.; Frank, A. T.; Kratz, J. D.; Swanson, M. D.; Gonzalez-Hernandez, M. J.; Lee, J.; Andricioaei, I.; Markovitz, D. M.; Al-Hashimi, H. M. Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble. *Nat Chem Biol* **2011**, *7* (8), 553–559. https://doi.org/10.1038/nchembio.596.

(54) Ganser, L. R.; Lee, J.; Rangadurai, A.; Merriman, D. K.; Kelly, M. L.; Kansal, A. D.; Sathyamoorthy, B.; Al-Hashimi, H. M. High-Performance Virtual Screening by Targeting a High-Resolution RNA Dynamic Ensemble. *Nat Struct Mol Biol* **2018**, *25* (5), 425–434. https://doi.org/10.1038/s41594-018-0062-4.

(55) Levintov, L.; Vashisth, H. Ligand Recognition in Viral RNA Necessitates Rare Conformational Transitions. *J. Phys. Chem. Lett.* **2020**, *11* (14), 5426–5432. https://doi.org/10.1021/acs.jpclett.0c01390.

(56) Karthikeyan, S.; Zhou, Q.; Mseeh, F.; Grishin, N. V.; Osterman, A. L.; Zhang, H. Crystal Structure of Human Riboflavin Kinase Reveals a β Barrel Fold and a Novel Active Site Arch. *Structure* **2003**, *11* (3), 265–273. https://doi.org/10.1016/S0969-2126(03)00024-8.

(57) Serganov, A.; Huang, L.; Patel, D. J. Coenzyme Recognition and Gene Regulation by a Flavin Mononucleotide Riboswitch. *Nature* **2009**, *458* (7235), 233–237. https://doi.org/10.1038/nature07642.

(58) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(59) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J Comput Aided Mol Des* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(60) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for pKaprediction and Protonation State Generation for Drug-like Molecules. *J Comput Aided Mol Des* **2007**, *21* (12), 681–691. https://doi.org/10.1007/s10822-007-9133-z.

(61) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *Journal of Molecular Biology* **2002**, *320* (3), 597–608. https://doi.org/10.1016/S0022-2836(02)00470-9.

(62) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J Comput Aided Mol Des* **2010**, *24* (6), 591–604. https://doi.org/10.1007/s10822-010-9349-1.

(63) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E. I.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. Jr. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183–6191. https://doi.org/10.1021/acs.jcim.3c01153.

(64) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J Phys Chem Lett* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.

(65) Connelly, C. M.; Numata, T.; Boer, R. E.; Moon, M. H.; Sinniah, R. S.; Barchi, J. J.; Ferré-D'Amaré, A. R.; Schneekloth, J. S. Synthetic Ligands for PreQ1 Riboswitches Provide Structural and Mechanistic Insights into Targeting RNA Tertiary Structure. *Nat Commun* **2019**, *10* (1), 1501. https://doi.org/10.1038/s41467-019-09493-3.

(66) White, N.; Sadeeshkumar, H.; Sun, A.; Sudarsan, N.; Breaker, R. R. Na+ Riboswitches Regulate Genes for Diverse Physiological Processes in Bacteria. *Nat Chem Biol* **2022**, *18* (8), 878–885. https://doi.org/10.1038/s41589-022-01086-4.

(67) White, N.; Sadeeshkumar, H.; Sun, A.; Sudarsan, N.; Breaker, R. R. Lithium-Sensing Riboswitch Classes Regulate Expression of Bacterial Cation Transporter Genes. *Sci Rep* **2022**, *12* (1), 19145. https://doi.org/10.1038/s41598-022-20695-6.

(68) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(69) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(70) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8* (4), 1493–1502. https://doi.org/10.1021/ct3000734.

(71) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. https://doi.org/10.1021/j100142a004.

(72) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Rev. A.03, 2016.

(73) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; van Drunen, R.; van der Spoel, D.; Sijbers, A.; Keegstra, H.; Renardus, M. Gromacs: A Parallel Computer for Molecular Dynamics Simulations. *PHYSICS COMPUTING '92* **1993**, 252–256.

(74) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(75) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(76) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(77) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069–4074. https://doi.org/10.1063/1.449071.

(78) Bernetti, M.; Bussi, G. Pressure Control Using Stochastic Cell Rescaling. *The Journal of Chemical Physics* **2020**, *153* (11), 114107. https://doi.org/10.1063/5.0020514.

(79) Theobald, D. L. Rapid Calculation of RMSDs Using a Quaternion-Based Characteristic Polynomial. *Acta Crystallogr A* **2005**, *61* (Pt 4), 478–480. https://doi.org/10.1107/S0108767305015266.

(80) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(81) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105. https://doi.org/10.25080/Majora-629e541a-00e.

(82) 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* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.

(83) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12* (85), 2825–2830.

(84) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(85) Waskom, M. L. Seaborn: Statistical Data Visualization. *Journal of Open Source Software* **2021**, *6* (60), 3021. https://doi.org/10.21105/joss.03021.

(86) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121* (16), 9722–9758. https://doi.org/10.1021/acs.chemrev.0c01195.

(87) Appadurai, R.; Koneru, J. K.; Bonomi, M.; Robustelli, P.; Srivastava, A. Clustering Heterogeneous Conformational Ensembles of Intrinsically Disordered Proteins with T-Distributed Stochastic Neighbor Embedding. *J. Chem. Theory Comput.* **2023**, acs.jctc.3c00224. https://doi.org/10.1021/acs.jctc.3c00224.

(88) Maaten, L. van der; Hinton, G. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* **2008**, *9* (86), 2579–2605.

(89) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein−Solvent Interface. *J. Phys. Chem. B* **2002**, *106* (26), 6617–6633. https://doi.org/10.1021/jp020100m.

(90) Masetti, M.; Falchi, F.; Recanatini, M. Protein Dynamics of the HIF-2α PAS-B Domain upon Heterodimerization and Ligand Binding. *PLOS ONE* **2014**, *9* (4), e94986. https://doi.org/10.1371/journal.pone.0094986.

(91) Halle, B.; Persson, F. Analysis of Protein Dynamics Simulations by a Stochastic Point Process Approach. *J. Chem. Theory Comput.* **2013**, *9* (6), 2838–2848. https://doi.org/10.1021/ct400161u.

(92) Sánchez, H. R. Residence Times from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2022**, *126* (43), 8804–8812. https://doi.org/10.1021/acs.jpcb.2c03756.

(93) Bizzarri, A. R.; Cannistraro, S. Anomalous and Anisotropic Diffusion of Plastocyanin Hydration Water. *EPL* **1997**, *37* (3), 201. https://doi.org/10.1209/epl/i1997-00128-3.

(94) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys J* **2000**, *79* (6), 2966–2974.

(95) Rocchia, W.; Alexov, E.; Honig, B. Extending the Applicability of the Nonlinear Poisson−Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001**, *105* (28), 6507–6514. https://doi.org/10.1021/jp010454y.

(96) Decherchi, S.; Rocchia, W. A General and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale. *PLOS ONE* **2013**, *8* (4), e59744. https://doi.org/10.1371/journal.pone.0059744.

(97) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T. Selective Small-Molecule Inhibition of an RNA Structural Element. *Nature* **2015**, *526* (7575), 672–677. https://doi.org/10.1038/nature15542.

(98) Bernetti, M.; Bertazzo, M.; Masetti, M. Data-Driven Molecular Dynamics: A Multifaceted Challenge. *Pharmaceuticals (Basel)* **2020**, *13* (9), 253. https://doi.org/10.3390/ph13090253.

(99) Sivanesan, D.; Basu, G.; Go, N. The Role of Electrostatics in Discrimination of Adenine and Guanine by Proteins. *Genome Informatics* **2002**, *13*, 316–317. https://doi.org/10.11234/gi1990.13.316.

(100) Radić, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. Electrostatic Influence on the Kinetics of Ligand Binding to Acetylcholinesterase: DISTINCTIONS BETWEEN ACTIVE CENTER

LIGANDS AND FASCICULIN *. *Journal of Biological Chemistry* **1997**, *272* (37), 23265–23277. https://doi.org/10.1074/jbc.272.37.23265.

(101) Mehler, E. L.; Gerhards, J. Electronic Determinants of the Anti-Inflammatory Action of Benzoic and Salicylic Acids. *Mol Pharmacol* **1987**, *31* (3), 284–293.

(102) Wade, R. C.; Gabdoulline, R. R.; Lüdemann, S. K.; Lounnas, V. Electrostatic Steering and Ionic Tethering in Enzyme–Ligand Binding: Insights from Simulations. *Proceedings of the National Academy of Sciences* **1998**, *95* (11), 5942–5949. https://doi.org/10.1073/pnas.95.11.5942.

(103) Sebastián, M.; Serrano, A.; Velázquez-Campoy, A.; Medina, M. Kinetics and Thermodynamics of the Protein-Ligand Interactions in the Riboflavin Kinase Activity of the FAD Synthetase from Corynebacterium Ammoniagenes. *Sci Rep* **2017**, *7* (1), 7281. https://doi.org/10.1038/s41598-017-07875-5.

(104) Thomas, J. R.; Hergenrother, P. J. Targeting RNA with Small Molecules. *Chem. Rev.* **2008**, *108* (4), 1171–1224. https://doi.org/10.1021/cr0681546.

(105) Guan, L.; Disney, M. D. Recent Advances in Developing Small Molecules Targeting RNA. *ACS Chem. Biol.* **2012**, *7* (1), 73–86. https://doi.org/10.1021/cb200447r.

(106) Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove, A. E. R-BIND: An Interactive Database for Exploring and Developing RNA-Targeted Chemical Probes. *ACS Chem. Biol.* **2019**, *14* (12), 2691–2700. https://doi.org/10.1021/acschembio.9b00631.

(107) Haniff, H. S.; Knerr, L.; Liu, X.; Crynen, G.; Boström, J.; Abegg, D.; Adibekian, A.; Lekah, E.; Wang, K. W.; Cameron, M. D.; Yildirim, I.; Lemurell, M.; Disney, M. D. Design of a Small Molecule That Stimulates Vascular Endothelial Growth Factor A Enabled by Screening RNA Fold-Small Molecule Interactions. *Nat Chem* **2020**, *12* (10), 952–961. https://doi.org/10.1038/s41557-020-0514-4.

(108) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chemistry & Biology* **1996**, *3* (12), 973–980. https://doi.org/10.1016/S1074-5521(96)90164-7.

(109) Hummer, G. Under Water's Influence. *Nature Chem* **2010**, *2* (11), 906–907. https://doi.org/10.1038/nchem.885.

(110) Baron, R.; Setny, P.; Andrew McCammon, J. Water in Cavity−Ligand Recognition. *J Am Chem Soc* **2010**, *132* (34), 12091–12097. https://doi.org/10.1021/ja1050082.

(111) Darby, J. F.; Hopkins, A. P.; Shimizu, S.; Roberts, S. M.; Brannigan, J. A.; Turkenburg, J. P.; Thomas, G. H.; Hubbard, R. E.; Fischer, M. Water Networks Can Determine the Affinity of Ligand Binding to Proteins. *J. Am. Chem. Soc.* **2019**, *141* (40), 15818–15826. https://doi.org/10.1021/jacs.9b06275.

(112) Henchman, R. H.; Tai, K.; Shen, T.; McCammon, J. A. Properties of Water Molecules in the Active Site Gorge of Acetylcholinesterase from Computer Simulation. *Biophysical Journal* **2002**, *82* (5), 2671–2682. https://doi.org/10.1016/S0006-3495(02)75609-9.

(113) García, A. E.; Stiller, L. Computation of the Mean Residence Time of Water in the Hydration Shells of Biomolecules. *Journal of Computational Chemistry* **1993**, *14* (11), 1396–1406. https://doi.org/10.1002/jcc.540141116.

(114) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, *10* (11), 5047–5056. https://doi.org/10.1021/ct500381c.

For Table of Contents Only