

RESEARCH ARTICLE

Molecular basis for the increased affinity of an RNA recognition motif with re-engineered specificity: A molecular dynamics and enhanced sampling simulations study

Anna Bochicchio¹^{¶a}, Miroslav Krepl²^{¶*}, Fan Yang^{3¶b}, Gabriele Varani³, Jiri Sponer^{2,4}[¶], Paolo Carloni^{1,5*}

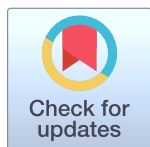
1 Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, Jülich, Germany, **2** Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic, **3** Department of Chemistry, University of Washington, Seattle, Washington, United States of America, **4** Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University Olomouc, Olomouc, Czech Republic, **5** JARA-HPC, Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, Jülich, Germany

 These authors contributed equally to this work.

^{¶a} Current address: Computational Biology, Department of Biology, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

^{¶b} Current address: School of Life Science and Technology, Harbin University of Technology, Nan Gang District, Harbin, China

* krepl@ibp.cz (MK); p.carloni@fz-juelich.de (PC)



OPEN ACCESS

Citation: Bochicchio A, Krepl M, Yang F, Varani G, Sponer J, Carloni P (2018) Molecular basis for the increased affinity of an RNA recognition motif with re-engineered specificity: A molecular dynamics and enhanced sampling simulations study. PLoS Comput Biol 14(12): e1006642. <https://doi.org/10.1371/journal.pcbi.1006642>

Editor: Shi-Jie Chen, University of Missouri, UNITED STATES

Received: July 11, 2018

Accepted: November 13, 2018

Published: December 6, 2018

Copyright: © 2018 Bochicchio et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files and available from the Zenodo database (accession number [http://doi.org/10.5281/zenodo.1297931](https://doi.org/10.5281/zenodo.1297931), <https://doi.org/10.5281/zenodo.1307420>, <https://doi.org/10.5281/zenodo.1310053>, <https://doi.org/10.5281/zenodo.1465433>, <https://doi.org/10.5281/zenodo.1310054>, <https://doi.org/10.5281/zenodo.1307484>, <https://doi.org/10.5281/zenodo.1307182>).

Abstract

The RNA recognition motif (RRM) is the most common RNA binding domain across eukaryotic proteins. It is therefore of great value to engineer its specificity to target RNAs of arbitrary sequence. This was recently achieved for the RRM in Rbfox protein, where four mutations R118D, E147R, N151S, and E152T were designed to target the precursor to the oncogenic miRNA 21. Here, we used a variety of molecular dynamics-based approaches to predict specific interactions at the binding interface. Overall, we have run approximately 50 microseconds of enhanced sampling and plain molecular dynamics simulations on the engineered complex as well as on the wild-type Rbfox-pre-miRNA 20b from which the mutated systems were designed. Comparison with the available NMR data on the wild type molecules (protein, RNA, and their complex) served to establish the accuracy of the calculations.

Free energy calculations suggest that further improvements in affinity and selectivity are achieved by the S151T replacement.

Author summary

RNA is an outstanding target for oncological intervention. Engineering the most common RNA binding motif in human proteins (called RRM) so as to bind to a specific RNA has an enormous pharmacological potential. Yet, it is highly non trivial to design RRM-bearing protein variants with RNA selectivity and affinity sufficiently high for clinical applications. Here we present an extensive molecular simulation study which shed light on the exquisite molecular recognition of the empirically-engineered complex between the

1308045, <https://doi.org/10.5281/zenodo.1464923>.

Funding: MK and JS were supported by the Czech Science Foundation [grant number P305/12/G034] and the project LO1305 of the Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme II. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

RRM-bearing protein Rbfox and its RNA target pre-miR21. The simulations allow predicting a variant, the S151T, which may lead to further enhancement of selectivity and affinity for pre-miR21.

Introduction

The RNA recognition motif (RRM) is the largest family of eukaryotic RNA-binding proteins [1], involved in virtually all post-transcriptional regulatory events [2]. RRM bind a wide-range of single-stranded RNAs [3], stem-loops and other RNA structures [2, 4–6]. Therefore, engineering RRM binding interfaces to target specific RNAs may create widely applicable tools for regulating gene expression [7, 8]. Yet, a variety of factors have hampered such efforts, including the complexities of the protein-RNA interactions, a poor understanding of the structural and biophysical basis for specificity, and the idiosyncratic way in which various RRM domains bind to RNA [9, 10].

Recently, some of us were able to engineer the conserved RRM domain of the human Rbfox protein by modulating its specificity for a target RNA [8]. The protein is part of a small family of tissue-specific alternative splicing regulators. It was chosen for its ability to bind with high sequence specificity and affinity—in the low nM range—to the r-GCAUG sequence in specific RNAs. These are the single-stranded RNAs and the hairpin microRNA precursors that code for miR107 and miR20b (referred to as pre-miR20b, hereafter, see Fig 1) [3, 6]. The r-G₂₉AAUC₃₃ sequence in the terminal loop of the chosen RNA target, the oncogenic precursor miRNA 21 (pre-miR21) [11], bears two nucleotide changes (at positions 30 and 33) from the r-GCAUG sequence. These mutations are sufficient to nearly abolish Rbfox binding [8]. The successfully engineered R118D-E147R-N151S-E152T quadruple mutant (Rbfox* hereafter, Fig 1) binds tightly to the pre-miR21 terminal loop sequence ($K_d \sim 13$ nM) [8], but also to pre-miR20b, with a dissociation constant only ~10 fold higher ($K_d \sim 150$ nM) [8]. Further improvements in binding specificity could be facilitated by understanding of the structural dynamics of key interactions at the protein-RNA interface at atomic level of description.

Molecular Dynamics (MD) simulations in explicit solvent are a useful tool to dissect the nature of interactions and specificity in biomolecular complexes [12, 13], providing information beyond what can be obtained experimentally. In particular, MD nicely complements NMR experiments on RNA interactions with RRM class of binding domains [14–16] by providing insights into specific interactions that are not revealed by experiments.

In this manuscript, we report the use of molecular simulation approaches to predict the structural determinants of the Rbfox*•pre-miR21 complex. After performing standard simulations, we use free-energy calculations to investigate a new mutant (S151T Rbfox*) that is predicted to improve selectivity towards the pre-miR21 target RNA relative to Rbfox*. The accuracy of our simulations is established by a comparison with the available NMR data and structure of the Rbfox•pre-miR20b complex [6, 17].

Results and discussion

As a first task (i), we tested our computational setup by performing extensive MD simulations, extended to the microsecond timescale, on the Rbfox•pre-miR20b complex in explicit solvent. Note that the RNA hairpin loop is remodeled by the protein compared to the free RNA; in the complex the hairpin segment is larger to accommodate the protein [6]. The RNA/protein interface is stabilized by a number of intermolecular stacking interactions and hydrogen bonds, which provide tight sequence specificity for the nucleotide sequence–G₂₉C₃₀A₃₁U₃₂G₃₃–. The MD results reproduce the available NMR structural data well and describe accurately the

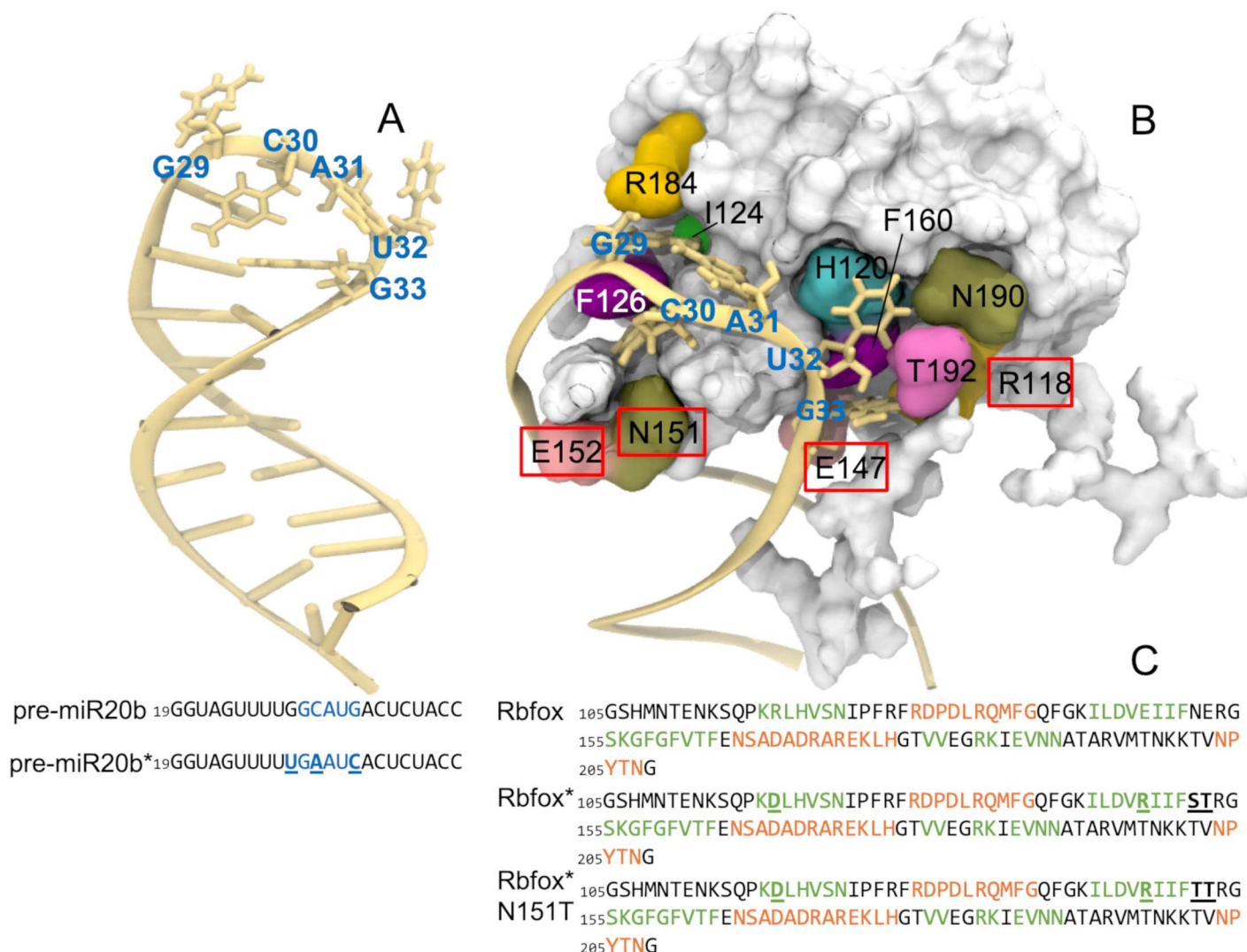


Fig 1. (A) NMR structure of pre-miR20b RNA (pdb 2n7x). (B) NMR structure of the Rbfox•pre-miR20b complex (pdb 2n82). The residues at the binding interface are highlighted. The amino acids labelled with a red square correspond to the mutated residues in Rbfox* and in Rbfox* S151T. (C) Left: nucleotide sequence of the pre-miR20b and of the mutant pre-miR20b*. Right: amino acids sequence of the Rbfox, Rbfox* and Rbfox* S151T mutants. Highlighted in green and in orange are the amino acids corresponding to β strands and α helices, respectively. The mutated nucleotides and amino acids are underlined.

<https://doi.org/10.1371/journal.pcbi.1006642.g001>

interactions at the binding interface. Comparison with simulations of the two isolated molecules (protein and RNA) suggests significant changes of the protein flexibility upon complex formation. Next (ii), we constructed a model of the binding interface of the Rbfox*•pre-miR21 complex by replacing G₂₈, C₃₀ and G₃₃ with U, A and C, respectively, and by substituting R118, E147, N151 and E152 residues with D, R, S, and T, respectively. We performed a series of explicit-solvent simulations on the resulting model. To ensure adequate sampling of the conformational space of the mutated complex, we used an enhanced sampling method (Replica Exchange with solute scaling -REST2- [17] simulations). The results were consistent with affinity data.

(iii) To further cross-validate the simulations predictions on the Rbfox*•pre-miR21 binding interface, we performed two additional simulations on the Rbfox•pre-miR21 and Rbfox*•pre-miR20b systems. The molecular description of the interactions at the two binding interfaces is in qualitative agreement with the experimental binding affinity data.

Finally (iv), we used the simulated model of the Rbfox*•pre-miR21 complex to design a mutant with predicted higher affinity and selectivity.

The studied protein-RNA complexes are characterized by a complex interplay between the sequence and structural dynamics. Therefore, quantitative analysis of the simulation trajectories is not trivial. To do so, we have employed a wide range of different descriptors to characterize the protein conformational dynamics and plasticity (RMSD and PAD [18]), RNA structural variation (ϵ RMSD [19] and structural parameters: torsion angles, base-pair and base-pair steps parameters), and to quantify the change in protein stability upon mutations and RNA binding (conformational entropy).

Validation of simulation setup (Simulations 1–13 in Table 1)

Recent MD studies on diverse protein-RNA complexes provided indication that the standard equilibration protocols, usually sufficient to equilibrate isolated medium-size RNA or protein molecules, might be inadequate for simulations of protein-RNA complexes [20]. Therefore, we performed multiple microsecond-long simulations, and exploited the existing NMR information as restraints at the early stages of most simulations (details provided in Materials and Method section).

The properties described below are calculated on the unrestrained parts of the initially restrained simulations of the system (Table 1, sim. 3–7 and 9–13), and on the fully unrestrained runs (Table 1, sim. 2 and 8) for comparisons. The individual trajectories sampled a similar conformational space (S1 Fig). Consequently, the average structural and dynamic properties calculated over the entire MD ensemble (all trajectories merged) do not significantly differ from those determined over the individual trajectories.

Rbfox protein flexibility in simulations (Simulations 1 and 8–13 in Table 1)

The root mean square deviation (RMSD) of the RRM domain (residues 117–193) in the Rbfox•pre-miR20b complex and in the free state in aqueous solution fluctuate around an average of 0.17 ± 0.02 nm, and 0.30 ± 0.02 nm, respectively, after only 200 ns (S2 Fig). This suggests that the systems are well equilibrated for most of the dynamic runs. These RMSD values are within the uncertainty of the NMR ensemble.

As expected [6], the protein RRM domain (residues 117–193) becomes generally more rigid upon RNA binding. Indeed, we calculate a substantial decrease in the *per residue* conformational entropy upon binding (S3 Fig).

Finally, the backbone flexibility, described here in terms of the so-called Protein Angular dispersion for the Ramachandran angles (PAD) [18], is larger in the free state than in the Rbfox•pre-miR20b complex. The larger the PAD value, the more flexible the protein backbone. The same analysis also allows identification of conformational transitions of the backbone during simulations [18]. These involve residues belonging to the β_2 and β_3 strands and to the $\beta_2\beta_3$ loop. This region is inserted into pre-miR20b terminal loop and anchors the RNA to the protein surface (Fig 2) in a manner reminiscent of the structure of the U1A complex (PDB 1URN, [23]). However, unlike U1A, the Rbfox RRM binds much more strongly to a single stranded RNA compared to a stem-loop with the same binding sequence [6]. The pronounced flexibility of the $\beta_2\beta_3$ loop might not be optimal for binding to structured RNAs [24, 25].

The RNA is structurally stable in simulations of the complex (Simulations 8–13) but not in isolation (Simulations 2–7 and 1–3 (χ_{OL3} -CP-OPC) in Table 1).

The RNA conformational ensemble in the Rbfox•pre-miR20b complex simulations is compatible with that of the NMR ensemble (S5 Fig). In particular, the conformational flexibility of nucleotide U₂₇, not bound to the protein, is relatively large both in the NMR ensemble and in

Table 1. Simulations performed in this work.

Simulated systems and simulation numbers	Length
Rbfox	
1	1 μ s
Pre-miR20b	
2	1 μ s
3 ^a	1 μ s
4 ^a	1 μ s
5 ^a	1 μ s
6 ^a	1 μ s
7 ^a	1 μ s
1 (χ_{OL3} -CP-OPC) ^b	1 μ s
2 (χ_{OL3} -CP-OPC) ^b	1 μ s
3 (χ_{OL3} -CP-OPC) ^b	1 μ s
Rbfox•pre-miR20b	
8	1 μ s
9 ^a	1 μ s
10 ^a	1 μ s
11 ^a	1 μ s
12 ^a	1 μ s
13 ^a	1 μ s
Rbfox*	
14	1 μ s
Rbfox*•pre-miR20b*	
15 (REST2 PS) ^c	8 $\times$ 2 μ s
16 (REST2)	16 $\times$ 1 μ s
17	1 μ s
18	1 μ s
S151T Rbfox*•pre-miR20b*	
19	0.9 μ s
20	0.5 μ s
21 (FE)	10 $\times$ 100 $\times$ 200 ps ^d
22 (FE)	10 $\times$ 100 $\times$ 1 ns ^d
Rbfox•pre-miR20b*	
23	1 μ s
Rbfox*•pre-miR20b	
24	1 μ s

^aNMR restrains applied for the first 120 ns

^b χ_{OL3} -CP-OPC refers to MD simulations performed with the χ_{OL3} force field supplemented by Case modified phosphate's oxygens van der Waals parameters [21] and using the OPC water model [22].

^cREST2 PS refers to REST2 simulations with *partial scaling* of the solute atoms.

^dHere, we report the number of independent runs, the number of forward and backward transformations, and the length of transformation, respectively.

<https://doi.org/10.1371/journal.pcbi.1006642.t001>

simulations. It dominantly contributes to the observed relatively large eRMSD values (S4 Fig). If calculated on the loop nucleotides directly bound to the protein (G₂₈GCAUG₃₃), the average eRMSD value is only 0.76 ± 0.14 .

The backbone torsion angles values of the loop region (nucleotides 28–33) and the stem base pair parameters remain in agreement with those calculated for the structures of the NMR ensemble (S6 Fig).

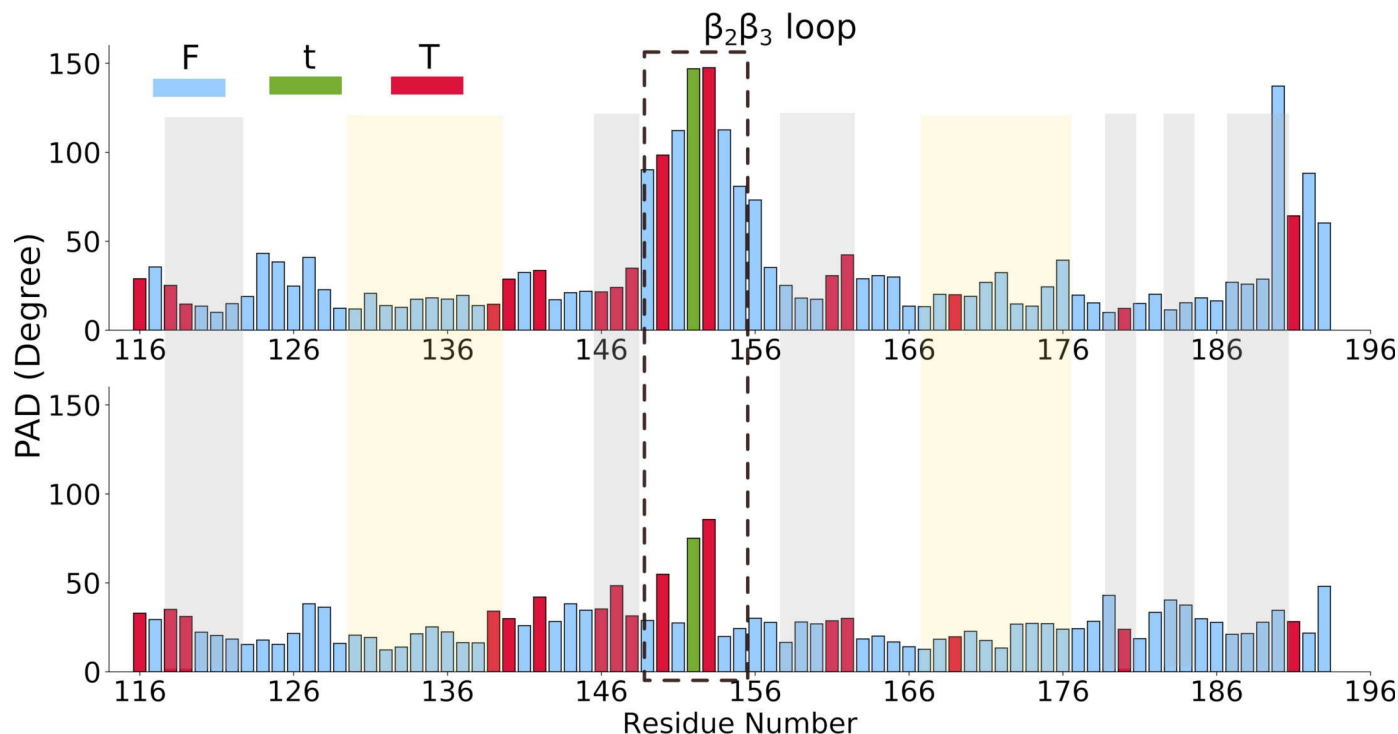


Fig 2. Fluctuations of the PAD angle in the simulations of free (top; see Table 1, sim. 1) and bound Rbfox protein (bottom; see Table 1, sim. 9–13). F indicates fluctuations; t short transitions and T long transitions. The structured regions of the protein are highlighted: the grey and yellow regions correspond to the β -strands and helices, respectively.

<https://doi.org/10.1371/journal.pcbi.1006642.g002>

In contrast, the simulations show immediate and large-scale conformational changes within the apical loop of the free pre-miR20b RNA (we reiterate that the RNA structure in isolation differs significantly from the complex). Unsatisfactory behavior of simulations of RNA hairpin loops has been widely analysed in literature [13, 26]. It is ascribed to accumulation of various inaccuracies in the force fields, such as an overstabilization of non-native base-phosphate and/or sugar-phosphate interactions, underestimated stability of the hydrogen bonding interaction in base pairing and various difficulties in describing the sugar-phosphate backbone substates [13]. Therefore, the description of the apical loop of the free pre-miR20b RNA can be expected to be less accurate than that of the RNA in complex with the protein. Indeed, in all simulations of the free pre-miR20b RNA (Table 1, sim. 2–7), performed with the standard AMBER force field (χ_{OL3} ; described in Methods), the $U_{27}GGCAU_{32}$ loop is rearranged, and the original NMR conformation is never recovered afterwards (Fig 3, S7 Fig). Note that in the NMR structure, the loop is rigidly ordered and characterized by a $U_{27}-G_{28}$ stacking interaction and a $G_{28}-U_{32}$ type 3 base-phosphate (3BPh) interaction [27] (Fig 3).

To illustrate the changes in the loop conformations, we show the overlap of frames from each simulation and the NMR starting structure in S7 Fig. We observe high mobility of the G_{29} base, which flips around its χ torsion from *anti* to *syn* to stack with the G_{28} base (Fig 3), followed by the loss of the $G_{28}-U_{32}$ 3BPh interaction. The C_{30} base is either bulged out or forms stacking interactions with A_{31} . The high mobility of C_{30} and U_{32} results in a significant distortion of the backbone torsion angles (Fig 3). Due to our inability to reproduce the NMR structure of the isolated RNA hairpin, we did not attempt any simulations of the isolated RNA starting from its conformation seen in the RNA/protein complex.

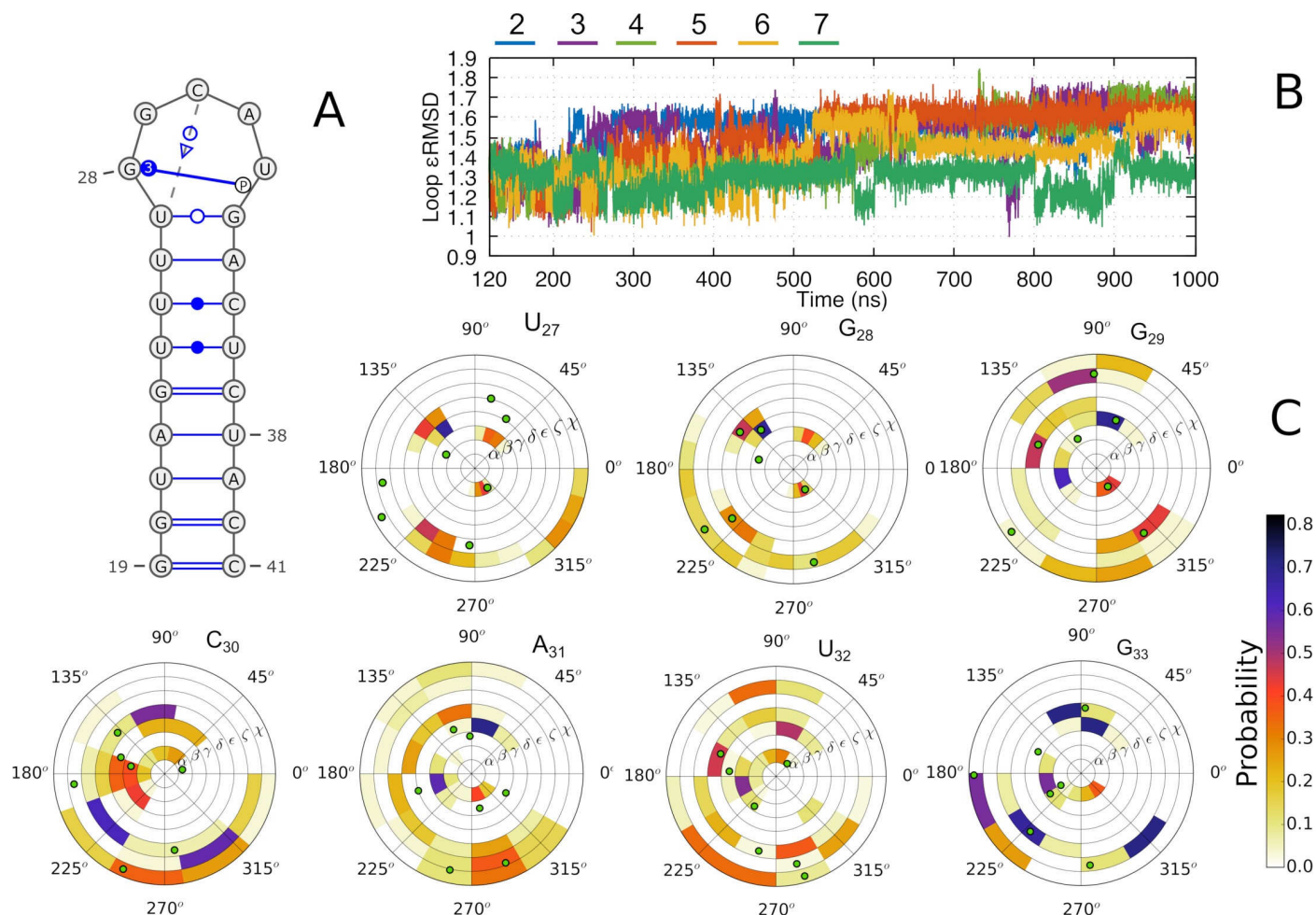


Fig 3. (A) 2D representation of the free pre-miR20b stem-loop structure. (B) Time development of the ϵ RMSD of the pre-miR20b loop (r-U₂₇GGCAUG₃₃) in the six χ_{OL3} MD simulations (Table 1, sim. 2–7). (C) RNA backbone dihedral angle histograms calculated over the aggregated simulations. The green dots indicate the values of the angles in the lowest energy structure of the NMR ensemble 2n7x from which the simulations were started.

<https://doi.org/10.1371/journal.pcbi.1006642.g003>

The pronounced flexibility of the terminal loop, however, does not affect the pre-miR20b stem dynamics, as shown by a comparison between the base pair and base-pair steps parameters of the simulated and NMR-solved ensembles of structures (S8 Fig).

Likewise, the inability to reproduce the experimental conformation of the apical loop in simulations of the free pre-miR20b RNA should not affect our investigations of the protein-RNA complex where the loop is sufficiently stabilized by the protein, in addition, the splayed RNA conformation is likely less strained than in the isolated hairpin loop as suggested by the experimental data [6].

It has been demonstrated that using the revised van der Waals phosphate's oxygen parameters reported in ref. [21], along with the 3-charge, 4-point OPC water model [22] partially improves simulations of RNA tetranucleotides by stabilizing the native A-form like conformations [28, 29]; this protocol is referred as χ_{OL3} -CP-OPC force-field combination, hereafter. However, for the pre-miR20b loop, the χ_{OL3} -CP-OPC force field-based simulations did not achieve better accuracy than those based on the parent χ_{OL3} force field (Table 1, sim. 1–3 (χ_{OL3} -CP-OPC)); a similar unsatisfactory outcome was reported also for other hairpin-loop systems investigated recently [26]. Indeed, the pre-miR20b χ_{OL3} -CP-OPC simulations showed

syn/anti nucleobase flips and alternative stacking conformations for all the terminal loop nucleotides (S6 Fig). Since no improvement was detected with this protocol, we did not attempt χ_{OL3} -CP-OPC simulations of the protein-RNA complexes. The use of the modified phosphate parameters, while improving the A-form single-strand simulations, might destabilize for example some native BPh interactions in folded RNAs. Note that simulations of free RNA hairpin loops remain a fundamental challenge for all currently available RNA force fields [13][30].

Analysis of average ion occupancies [31] revealed an average local concentration of approximately 1 M sodium near the U_{27}/G_{33} base pair in simulations of the free RNA. This is a significantly elevated ion concentration, well above the bulk value. Here, the Na^+ ions interact with the (U_{27}) O4 atom with a residency time of few ns, at a distance of ~ 0.25 nm (S9 Fig). This ion-binding site is completely absent in the complex structure since the U_{27}/G_{33} base pair is disrupted by protein binding. Instead, in the complex, the Na^+ concentration is strongly localized at the dinucleotide step 34–35 (S9 Fig).

The RNA/protein interface

The interface structure sampled in MD simulations is similar to that of the NMR-resolved ensemble (Fig 4). In particular, G_{29} stacks with F126 and R184. The $G_{29}/R184$ stacking interaction is always observed in the simulations (S10 Fig) even though it is absent in some frames of the NMR ensemble. The network of interactions is further stabilized by a bifurcated hydrogen bond involving the I124 backbone and the G_{29} base (Table 2). As in the NMR ensemble, G_{29} and A_{31} form a *trans* Watson Crick/Shallow groove (tWS) base-pair [32] in the simulations.

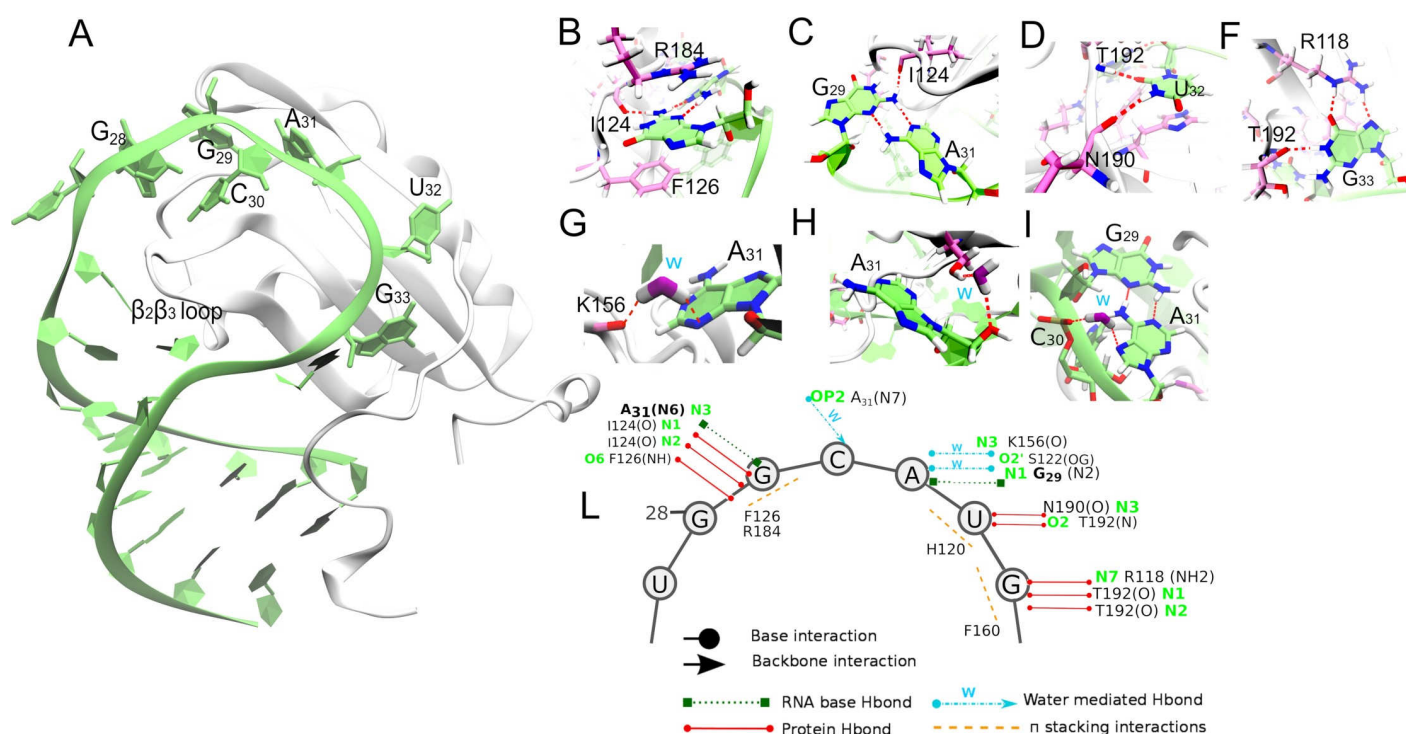


Fig 4. (A) Bird's-eye view of the Rbfox•pre-miR20b complex. (B)-(H) Close-up views of the interactions observed at the binding interface during MD simulations. The H-bonds are indicated by dotted red lines. The depicted snapshots belong to the representative structures of the 20 clusters ("MD-adapted structure ensemble"), which have the highest agreement with NMR NOE data. (I) Scheme of the interactions. Circle and arrowheads depict interaction with RNA bases or phosphate groups, respectively.

<https://doi.org/10.1371/journal.pcbi.1006642.g004>

Table 2. Hydrogen bonds at the binding interface of the Rbfox•pre-miR20b complex observed during the MD simulations (Table 1, sim. 8–13). The average donor-acceptor distances and angles are calculated over the entire simulation ensemble for the trajectory frames in which the individual H-bonds are observed. The interactions are further characterized by their occupancy in the individual simulation trajectories 8–13.

Acceptor	Donor	Distance (nm)	Angle (°)	8	9	10	11	12	13	NMR Ensemble
G ₂₉ (O6)	F126 (N)	0.31 ± 0.02	152 ± 1	60%	83%	72%	72%	72%	69%	75%
I124 (O)	G ₂₉ (N1)	0.28 ± 0.01	148 ± 1	85%	96%	96%	96%	92%	96%	100%
I124 (O)	G ₂₉ (N2)	0.29 ± 0.02	147 ± 2	90%	89%	88%	89%	90%	88%	100%
G ₂₉ (N3)	A ₃₁ (N6)	0.31 ± 0.02	158 ± 1	70%	75%	74%	69%	73%	74%	90%
A ₃₁ (N1)	G ₂₉ (N2)	0.30 ± 0.02	166 ± 2	95%	99%	98%	98%	99%	98%	90%
N190 (O)	U ₃₂ (N3)	0.31 ± 0.02	162 ± 1	89%	93%	95%	96%	95%	95%	100%
U ₃₂ (O2)	T192 (N)	0.31 ± 0.01	159 ± 1	90%	94%	95%	94%	95%	95%	100%
G ₃₃ (N7)	R118 (NH2)	0.29 ± 0.02	161 ± 1	79%	94%	95%	94%	95%	94%	40%
T192 (O)	G ₃₃ (N1)	0.28 ± 0.01	155 ± 2	70%	96%	98%	97%	67%	65%	90%
T192 (O)	G ₃₃ (N2)	0.28 ± 0.02	150 ± 1	73%	69%	74%	96%	68%	95%	100%

<https://doi.org/10.1371/journal.pcbi.1006642.t002>

The G₂₉/A₃₁ interaction is further stabilized by a water molecule, coordinated by the A₃₁ N7 atom and C₃₀ phosphate group (Fig 4). The adenine base forms water-mediated hydrogen bonds with the S122 and K156 side chains. The residence time of water molecules in these interactions are of a few tens of ns, sensibly longer than the common time-scale (50–500 ps) of short-residency hydration sites around RNA molecules [33–35]. These results are fully consistent with the earlier report of structured hydration sites in the simulations of the Rbfox RRM in complex with single-stranded RNA [15].

In our simulations, the U₃₂ base forms stacking interactions with H120 (S10 Fig) and hydrogen bonds with the backbone of N190 and T192 (Table 2, Fig 4). G₃₃ base is in *syn* conformation and forms stacking interactions with F160 (Fig 4, S10 Fig) and hydrogen bonds with T192 backbone and R118 side chain (Table 2, Fig 4). The latter interaction, always observed with good hydrogen bond geometry (Table 2) in the simulations, is present in eight conformers of the NMR ensemble, while in the other twelve the residue is solvent-exposed, perhaps due to insufficiently clear experimental information. Notably, the R118 side chain forms a very similar hydrogen-bonding interaction in the NMR structure of the Rbfox-RRM bound to the single-stranded RNA r-U₁G₂C₃A₄U₅G₆U₇ [3].

Next, we present several comparisons with the primary NMR data, which is a very stringent way to judge the accuracy of simulations [14].

On average, the back-calculated Chemical Shifts (CSs) for ¹³C and ¹H atoms of free and bound pre-miR20b, along with the ¹³C', ¹³C_a, ¹³C_b, and ¹⁵N atoms of Rbfox in its complex with pre-miR20b are, within the accuracy and the limits of the empirical methods (LARMOR^d [36] and SHIFTX+ [37]) used for the predictions (S11 Fig), in fair agreement with experimental observations.

The agreement between observed and calculated chemical shifts is only fair because of a variety of reasons. These include the fact that SHIFTX+ is expected not to be able to accurately predict shifts of residues in close proximity to RNA. Indeed, the characteristic ring current and charge for non-protein like molecules are not included in the SHIFTX2 parameterization [38]. Larmor^D suffers from the same drawback as it was parameterized by excluding RNA' structures in complex with proteins or other ligands in its training data set [36]. Moreover, the apparent agreement (S11 Fig, Pearson correlation coefficients R = 0.99 for ¹³C and R = 0.97 for ¹H) with the measured chemical shifts for the free pre-miR20b RNA (S11 Fig), which shows large conformational changes in simulations, suggest that Larmor^D sensitivity to structural changes might be limited.

However, the SHIFTX+ predictions were still sensitive to spurious motions of F150 and F158 as shown by the calculated distributions of the N nuclei, which are characterized by multiple peaks (S13 Fig). This might be related to temporary flips of the χ_1 dihedral angles from gauche(+) to gauche(-) during the simulations (S13 Fig). A similar pronounced flexibility for phenylalanine and tyrosine side chains was observed also in our previous simulations using the ff14SB force field [15, 16]. This potentially erroneous behavior can be related to the energy barrier for side chain rotation around the χ_2 angle [15], which might be too low in the employed force field.

The chemical shift predictions for the C_α nuclei of residues V146, E147, and V151, located on β_2 strand, deviate from the experimental values beyond two standard deviations (S12 Fig). In this case, the divergence reflects the dependence of the carbon shifts on backbone φ and ψ angles, for which we already described enhanced fluctuations during the simulations (Fig 2).

None of the back-calculated CSs of the bound RNA show significant differences compared to the experimental values (S15 Fig).

On average, ~84% of the NOE upper bounds are satisfied for the isolated protein and RNA and only ~3% of the violated NOE have violations greater than 0.05 nm (S1 Table). Not unexpectedly, in free RNA, the larger NOE violations are typically localized to the G₂₈, G₂₉ and U₃₂ nucleotides of the apical loop, which exhibit major conformational changes in the simulations (see above).

For the complex: on average, the percentage of satisfied NOE upper bounds is about 76%. Interestingly, the single trajectory generated without the initial application of the NMR restraints (13), shows the largest number of violations (Table 3). This indicates that the application of the experimental restraints in the early part of production trajectory leads to visibly better agreement with the inter-molecular NOEs in the subsequent unrestrained trajectories. However, this assessment should be taken with some caution, because it is made based on only a single unrestrained trajectory. The greatest distance violations (>0.3 nm) are observed for the inter-molecular NOEs involving amino acids F150 and F158 and nucleotides U₂₅, and U₃₂, possibly because of the relatively high flexibility of the two phenylalanine side chains noted above.

Altogether, these analyses demonstrate that the MD-derived conformational ensemble of structures reproduces fairly well the experimentally sampled conformations for the Rbfox•pre-miR20b complex. These and previous [14] results suggest that our protocol can be employed to study the dynamic properties of the engineered Rbfox*•pre-miR20b* complex and to compare it with the wild type complex from which it was designed.

The Rbfox*•pre-miR20b* complex

One of our main goals was to perform atomistic simulations of the Rbfox*•pre-miR20b* complex, for which the experimental structure is not available, and characterize the

Table 3. Percentage of intra- and intermolecular NOE violations observed in the course of the simulations of the protein-RNA complex. Trajectory 8 (Table 1) has been obtained without applying the restraints in the initial stages of the simulation—see Methods.

Simulation	Intra RNA	Intra Protein	Inter Protein-RNA
8	19%	24%	48%
9	17%	22%	34%
10	18%	21%	32%
11	17%	22%	33%
12	19%	22%	32%
13	18%	21%	33%

<https://doi.org/10.1371/journal.pcbi.1006642.t003>

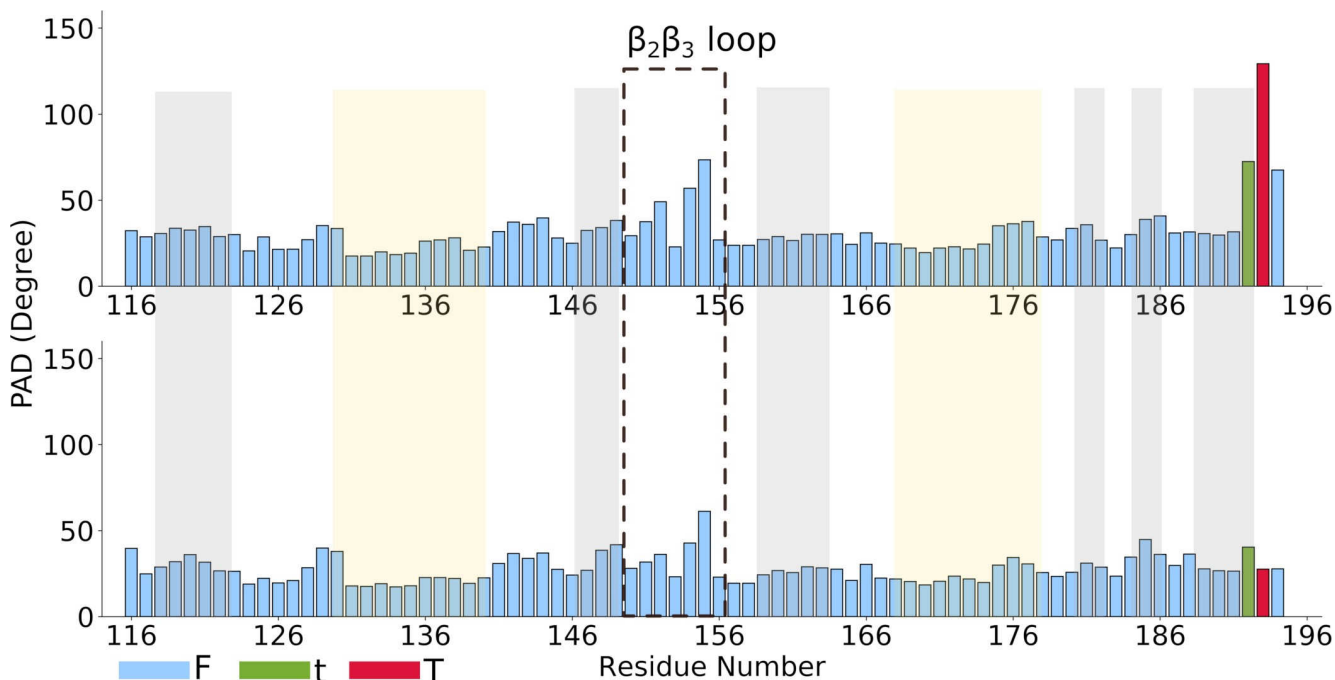


Fig 5. PAD and tag analyses for Rbfox* free (top) and bound to pre-miR20b* RNA (bottom). “F” indicates fluctuations; “t” short transitions and “T” long transitions. The secondary structure regions of the proteins are highlighted as in Fig 1.

<https://doi.org/10.1371/journal.pcbi.1006642.g005>

molecular interactions at its binding interface. The complex features R118D, E147R, N151S and E152T mutations on Rbfox, as well as G₂₈U, C₃₀A and G₃₃C mutations in the pre-miR20b RNA. To achieve wide exploration of the engineered binding interface conformational space, we used enhanced sampling methods, and specifically Hamiltonian replica exchange (HREX) MD simulations. These have become a common way to elucidate conformational ensembles of proteins [39–41] and nucleic acids [42, 43]. HREX simulations should eliminate any bias caused by the initial building up of the mutated structure. A wealth of recipes for HREX has been proposed in the last years (among others, [39, 44–50]). One of the most successful approaches is the so-called replica exchange solute tempering in its REST2 variant [17], in which only the solute Hamiltonian is scaled. Being mainly interested in the properties of the Rbfox*•pre-miR20b* binding interface, we have used a promising cost-saving variant of REST2, where only this part of the solute is scaled (Methods, REST2 PS, Table 1, sim. 15). However, a standard REST2 simulation was also performed (Methods) for comparison (Table 1, sim. 16). A discussion of convergence issues of these types of simulations is offered in the SI.

Overall, the structure of the mutated complex remains very similar to that of the wild type, while the flexibility is reduced (Fig 5 and Table 4). Unlike in the wild-type, no backbone conformational transitions are observed for the $\beta_2\beta_3$ loop in the MD simulations of either the free (Table 1, sim. 14) or the bound (Table 1, sim. 17–18) Rbfox* (Fig 5).

The results of this analysis are consistent with a considerable loss of per-residue conformational entropy of the Rbfox* $\beta_2\beta_3$ loop residues (Table 4 and S4 Fig). Upon RNA binding, the protein and the $\beta_2\beta_3$ loop become even less mobile. This is shown by a calculation of the so-called PAD values, which provide a measure of proteins’ backbone conformational flexibility and of the conformational entropy differences (Fig 5 and Table 4).

Table 4. Conformational entropy differences associated with Rbfox* protein for residues belonging to the $\beta_2\beta_3$ loop. ΔS values are calculated as $S_{Rbfox^*F} - S_{Rbfox^*F}$ (ΔS^a) and $S_{Rbfox^*F} - S_{Rbfox^*C}$ (ΔS^b), over simulations 1 (Rbfox free), 14 (Rbfox* free), and 17 (Rbfox* bound) as listed in Table 1. The subscripts F and C refer to the free and bound proteins, respectively. These values are obtained with very approximate methods, and they should be taken only for qualitative comparisons.

Residues	ΔS^a (kcal/mol)	ΔS^b (kcal/mol)
F150	0.97 ± 0.04	0.57 ± 0.02
N151S	0.73 ± 0.02	0.16 ± 0.01
E152T	1.23 ± 0.01	0.13 ± 0.01
R153	1.22 ± 0.02	0.45 ± 0.03
G154	0.72 ± 0.01	0.3 ± 0.02

<https://doi.org/10.1371/journal.pcbi.1006642.t004>

This increase in rigidity of the protein structure, and in particular of the loop—whose stiffening might provide a better steric fit for RNA binding [23]—might contribute to the 2-fold increase of binding affinity for E152T relative to the wild type protein [8].

At the RNA/protein interface of the mutant, the hydrogen bonds, and stacking interactions of the G₂₉, A₃₁ and U₃₂ bases (parts of the RNA/protein interface which are not mutated) observed in the wild-type simulations (Fig 4) are also preserved in the mutant complex (Fig 6, Table 5, S16 Fig). Interestingly, the water molecules coordinated by A₃₁, S122 and K156 side chains as in the wild-type complex exhibit slow exchange with bulk solvent, with residence time of tens of ns (Figs 4 and 6). This leads us to suggest that these hydration sites could be indeed important in stabilizing the binding interface [15].

Overall, the simulations convincingly suggest that the system is able to structurally tolerate the mutations without altering the overall Rbfox*•pre-miR20b* binding mode.

A₃₀ forms, about 30% of the time, a hydrogen bond with the S151 backbone (Table 5). This observation is consistent with the experimental data which shows only slight improvement of the mutant binding affinity relative to the wild type protein•pre-miR21 complex upon the N151S mutation [8]. Note that the S151 side chain mostly interacts with the N6 atom of A₃₀ in simulations, instead of the N1 atom, as was originally suggested (8). When not present, the S151/A₃₀ hydrogen bond is most often replaced by an intramolecular S151/G154 interaction (Table 5), which contributes to stabilizing the $\beta_2\beta_3$ loop.

C₃₃, in the *anti* conformation (G₃₃ in Rbfox•pre-miR20b complex was in *syn*), establishes either direct or water-mediated hydrogen bonds with the R147 guanidinium group (Fig 6). There are also hydrogen bonds with D118 and T192 side chains in most of the simulations (Table 5, Fig 6). This is consistent with the larger binding affinity gain of the R118D-E147R mutant for pre-miR21 relative to that of the wild type protein [8]. Indeed, these two mutations most significantly, increased the binding affinity of the mutant protein to pre-miR21 (by $\sim 10^2$ fold) compared to the wild type protein•pre-miR21 complex [8]. Lastly, we note that the A₃₁ forms an intermolecular stacking interaction with residue R153 within the $\beta_2\beta_3$ loop. This interaction is absent in the wild type complex. We suggest that the network of intermolecular interactions shown by our simulation is qualitatively consistent with the experimentally measured affinities and the rationale behind the design of the mutations [8].

The Rbfox•pre-miR20b* and the Rbfox*•pre-miR20b complexes

To further investigate the accuracy of our predicted interactions at the engineered complex binding interface, we performed two 1-microsecond long simulations (Table 1, sim. 23 and 24). The first focuses on the Rbfox•pre-miR20b* complex and shows that the interface is solely maintained by the stacking interactions of G₂₉ with F126 and R184 along with a hydrogen

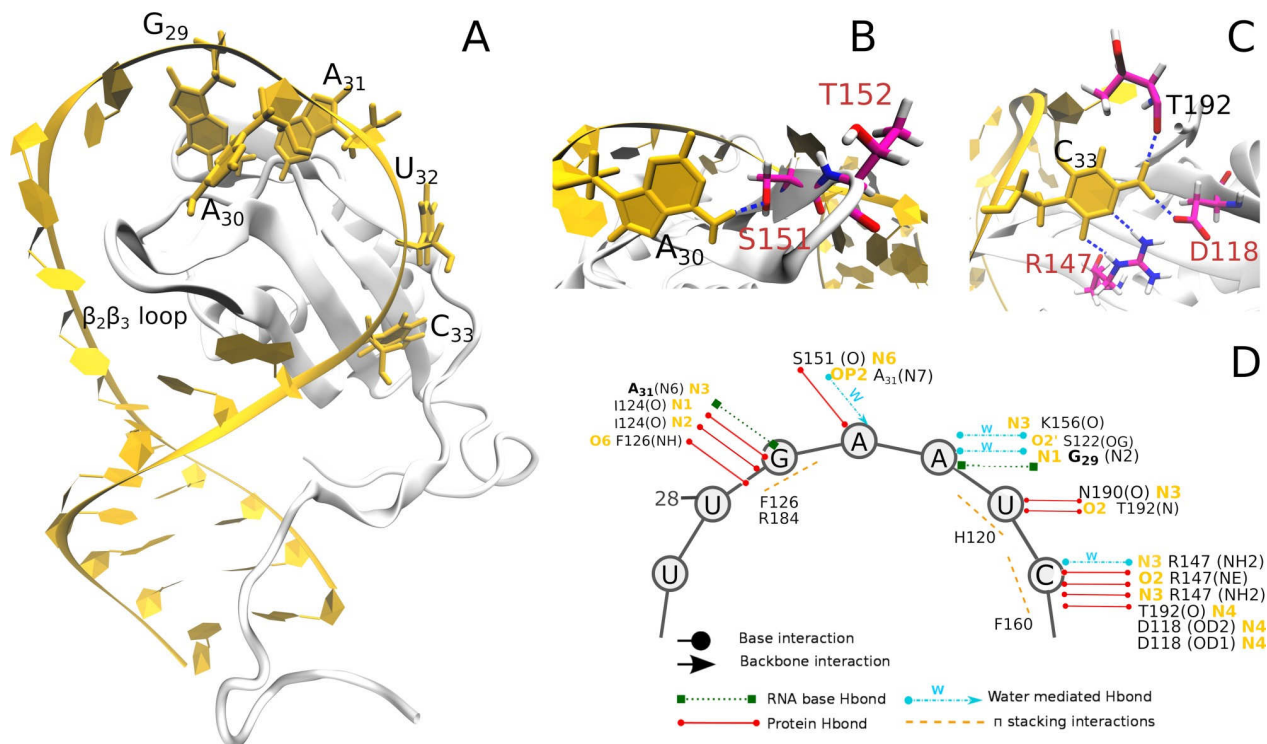


Fig 6. (A) Bird's-eye view of the Rbfox*•pre-miR20b* complex. (B) and (C) Close-up views of the interactions established by the mutated residues with A₃₀ and C₃₃, respectively. (D) Scheme of the interactions between pre-miR20b* and Rbfox* observed in the MD simulations. Circle and arrowheads depict interaction with RNA bases or phosphate groups, respectively.

<https://doi.org/10.1371/journal.pcbi.1006642.g006>

bond between I124 backbone and the G₂₉ base (Table 6 and S18 Fig). This finding is qualitatively consistent with electrophoretic mobility shift assay experiments, which indicate an extremely weak binding for these substitutions [8]. The second simulation is carried out on

Table 5. Hydrogen bonds at the binding interface of the Rbfox*•pre-miR20b* complex during simulations. The average donor-acceptor distances and angles are calculated over the entire simulation ensemble for the trajectory frames in which the individual H-bonds are observed and the interactions are further characterized by their occupancy in the individual simulation trajectories. The trajectories are numbered as in Table 1.

Acceptor	Donor	Distance (nm)	Angle (degree)	15 (REST2 PS)	16 (REST2)	17	18
I124 (O)	G ₂₉ (N1)	0.29 ± 0.01	149 ± 1	96%	95%	96%	96%
I124 (O)	G ₂₉ (N2)	0.29 ± 0.02	148 ± 1	93%	95%	94%	94%
G ₂₉ (O6)	F126 (N)	0.31 ± 0.02	154 ± 2	70%	75%	72%	73%
G ₂₉ (N3)	A ₃₁ (N6)	0.31 ± 0.02	160 ± 1	75%	70%	74%	69%
A ₃₁ (N1)	G ₂₉ (N2)	0.30 ± 0.02	166 ± 2	99%	95%	98%	98%
N190 (O)	U ₃₂ (N3)	0.31 ± 0.02	162 ± 1	93%	94%	96%	96%
S151 (O)	A ₃₀ (N6)	0.29 ± 0.03	151 ± 1	28%	27%	30%	29%
U ₃₂ (O2)	T192 (N)	0.31 ± 0.01	154 ± 1	91%	90%	95%	90%
C ₃₃ (N3)	R147 (NH2)	0.29 ± 0.01	163 ± 1	99%	99%	97%	99%
C ₃₃ (O2)	R147 (NE)	0.29 ± 0.01	159 ± 2	98%	99%	97%	98%
T192 (O)	C ₃₃ (N4)	0.29 ± 0.01	160 ± 2	96%	97%	98%	97%
D118 (OD1/OD2)	C ₃₃ (N4)	0.29 ± 0.01	159 ± 1	99%	99%	97%	98%
S151 (OG)	G154 (O)	0.27 ± 0.01	159 ± 1	47%	45%	50%	48%

<https://doi.org/10.1371/journal.pcbi.1006642.t005>

Table 6. Hydrogen bonds at the binding interfaces of the Rbfox•pre-miR20b* and Rbfox*•pre-miR20b complexes in MD simulations. The average donor-acceptor distances and angles are calculated for the trajectory frames in which the individual H-bonds are observed and the interactions are further characterized by their occupancy.

Acceptor	Donor	Distance (nm)	Angle (degree)	Occupancy
Rbfox•pre-miR20b*				
I124 (O)	G ₂₉ (N1)	0.29 ± 0.02	150 ± 1	90%
I124 (O)	G ₂₉ (N2)	0.3 ± 0.01	146 ± 1	80%
G ₂₉ (O6)	F126 (N)	0.32 ± 0.01	153 ± 1	66%
Rbfox*•pre-miR20b				
I124 (O)	G ₂₉ (N1)	0.3 ± 0.01	149 ± 1	95%
I124 (O)	G ₂₉ (N2)	0.3 ± 0.02	147 ± 1	94%
G ₂₉ (O6)	F126 (N)	0.31 ± 0.02	151 ± 2	60%
G ₂₉ (N3)	A ₃₁ (N6)	0.30 ± 0.02	160 ± 1	74%
A ₃₁ (N1)	G ₂₉ (N2)	0.31 ± 0.02	167 ± 2	90%
N190 (O)	U ₃₂ (N3)	0.3 ± 0.02	162 ± 1	95%
U ₃₂ (O2)	T192 (N)	0.31 ± 0.01	157 ± 1	94%
G ₃₃ (N7)	R147 (NE)	0.29 ± 0.01	160 ± 1	75%
T192 (O)	G ₃₃ (N1)	0.29 ± 0.01	149 ± 1	76%
R147 (NH2)	D118 (OD2/OD1)	0.31 ± 0.02	150 ± 2	70%

<https://doi.org/10.1371/journal.pcbi.1006642.t006>

the Rbfox*•pre-miR20b complex. The binding interface resulting from our simulation features equivalent hydrogen bonds and stacking interactions (Table 6) as observed in the wild type (Table 2 and Fig 4) and the Rbfox*•pre-miR20b* (Table 5 and Fig 6) complexes. These interactions involve the G₂₉, A₃₁ and U₃₂ bases. The RNA does not interact with the protein's β₂β₃ loop, and, in particular, with the mutated S151 and T152. Most notably, G₃₃ maintains its *syn* conformation (as in the Rbfox•pre-miR20b complex) and forms a hydrogen bond with the mutated R147 side chain. The latter in turn forms a hydrogen bond with the mutated D118 (S19 Fig). Hence, the simulation suggests a strong compensatory effect upon amino acids substitution at this site as new interactions are formed to maintain complex stability. This may be consistent with the good binding affinity of the mutated Rbfox* protein features for the pre-miR20b terminal loop [8] (see Introduction).

These results, although based on single trajectories, further establish the simulations predictive power through their qualitative agreement with the experimental binding assays.

***In silico* engineering a mutation with higher affinity and selectivity**

Based on overall consistency between predictions and available experimental data, we sought to identify a mutation, which would further improve affinity and selectivity for the target pre-miR21b RNA. Specifically, our simulations show that the N151S substitution as suggested by the experiments [8] does not lead to significant interactions with the RNA, possibly because of the intrinsic flexibility of the protein β₂β₃ loop. We therefore reasoned that placing a bulkier group, such as Thr, in position 151 would be advantageous. Both S151 and T151 are capable of forming the same H-bonds with the N1 and/or the N6 atoms of A₃₀. However, the bulkier side chain of T151 might influence the dynamics of the β₂β₃ loop. We therefore investigated the structure of S151T Rbfox*•pre-miR20b* by MD simulations (Table 1, sim. 19–20) and the change in affinity upon the S151T mutation by alchemical calculations using non-equilibrium approach (see Methods) [51]. The method has been successfully applied to a variety of protein mutants[51], and more recently, to protein–DNA–mutant complexes [52], providing accurate free energy estimates [52]. We refer to other works for a detailed comparison between the alchemical non-equilibrium and the equilibrium free energy calculations [53, 54].

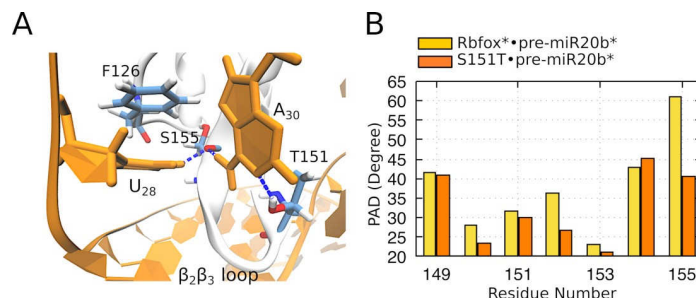


Fig 7. (A) Close up view on the interactions established by A₃₀ with T151 and S155 and by U₂₈ with S155 and F126. (B) Comparison between $\beta_2\beta_3$ loop and adjacent residues PAD values of the Rbfox* and S151T mutant in complex with pre-miR20b* RNA. The PAD values relate to the flexibility of the complex.

<https://doi.org/10.1371/journal.pcbi.1006642.g007>

While the basic structure of the complex is overall unaffected, the A₃₀(N6) forms hydrogen bond with the S155 backbone oxygen, while the A₃₀(N1) forms a hydrogen bond with the T151 hydroxyl group (Fig 7). These interactions might decrease the flexibility of the $\beta_2\beta_3$ loop compared to the previous mutant (Fig 7), and lead to an indirect stabilization of the position of the U₂₈ base, which is able to form a stable H-bond with the S155 backbone oxygen, and a stacking interaction with F126 (Fig 7). None of these interactions are observed in the Rbfox*•pre-miR20b* complex, where the U₂₈ base is always solvent-exposed. Note that identical binding pattern for the U₂₈ was also observed in earlier MD simulation studies of the wild type Rbfox complexed with a single-stranded RNA [14, 15]. The T151 thus might be better in overall accommodation of the pre-miR20b* RNA than the S151.

The free energy change associated with the S151T mutation, calculated using computational alchemy over two different simulation time windows (see Materials and Methods), is either -1.2 ± 0.3 kcal mol⁻¹ or -1.3 ± 0.1 kcal mol⁻¹. Hence, this estimation appears to be well converged and suggests that the mutation increases to a small, yet significant extent the affinity of the complex.

The presence of the U₂₈ /protein interactions might also improve the selectivity of this mutant for the r-U₂₈GAAUC₃₃ sequence in pre-miR20b* RNA over r-G₂₈GCAUG₃₃ found in pre-miR20b. Indeed, in the wild type complex, G₂₈ (equivalent to U₂₈) exhibits pronounced flexibility in simulations and does not form any contact with the protein, ([6] and this work). Hence, our simulations suggest that the proposed mutation would alter the preference of the binding interface for the pre-miR20b* sequence over the pre-miR20b RNA, improving both the affinity and the selectivity of the engineered protein for the target pre-miR21 RNA.

Conclusions

MD simulations of protein-RNA complexes remain somewhat limited by practical considerations of sampling (i.e. simulation time-scale) and inaccuracies resulting from force-field limitations [12, 20], yet they can supply important insight that often cannot be obtained by experiments, specifically on free-energy contributions and persistence of intermolecular contacts. The MD simulations in explicit solvent conducted here, covering overall about 50 microseconds of simulation data, including several state-of-the-art simulation techniques and validated by their full consistency with experimental data, provide a detailed atomistic picture of the effect of mutations in the Rbfox*•pre-miR20b* interface. The simulations also suggest a new mutant, S151T, which is predicted to be more selective and have higher affinity for the pre-miR-21 sequence than the S151 suggested in the original design.

Methods

Structures building and force-field selections

We used the lowest energy structures of the NMR ensembles 2cq3 [55], 2n7x [6] and 2n82 [6] as starting structures for the simulations of the free Rbfox protein, pre-miR20b RNA and Rbfox-pre-miR20b complex, respectively.

The starting structure of Rbfox* was prepared by introducing R118D, E147R, N151S and E152T mutations into Rbfox (in both free and bound states) using the Swiss MODEL software (available at <https://spdbv.vital-it.ch/>) [56–59]. The starting model of pre-miR20b* was obtained by replacing the G₂₈, C₃₀ and G₃₃ in the pre-miR20b structure with U, A and C, respectively, using the *tleap* module of Amber 16 (available as AmberTools16 at <http://ambermd.org/AmberTools16-get.html>) [60]. The pre-miR20b* sequence (r-G₁₉GUAGUUUUU₂₈GAAUC₃₃ACUCUACC₄₁) is equivalent to that of the pre-miR21 only in the terminal loop (nucleotides 28–33: UGAAUC). This is part of the protein-RNA interface. The remainder of the sequence does not interact with the protein and was therefore left unchanged and identical to the pre-miR20b (r-G₁₉GUAGUUUUG₂₈GCAUG₃₃ACUCUACC₄₁).

The molecules were solvated in truncated octahedral water boxes with minimal distance of 0.10 nm between solutes and the box border. The solutes were neutralized with sodium ions followed by addition of a sufficient number of Na⁺/Cl[−] ion pairs to reach the excess salt concentration of 80 mM. Similar solvent conditions were shown to work well for other protein-RNA systems [14, 15, 20]. The topology and coordinate files for the simulations were prepared using the *tleap* module of Amber 16 [60].

TIP3P [61], Joung and Cheatham [62], and the amber ff14SB [63], and χ_{OL3} [64] force fields were used for water, ions, proteins, and RNA respectively. This combination has shown satisfactory behavior with other protein-RNA complexes [14].

We performed also a second set of MD simulations of the free pre-miR20b RNA. These were carried out in exactly the same way as the first set except that they included the recently suggested modification of van der Waals oxygen radii for organic phosphates (atom types O2, OH, and OS)[21], along with the OPC water model [22].

MD simulations protocol

All systems were subjected to energy minimization, and equilibration using a standard equilibration protocol [20]. In order to reduce the likelihood of instabilities in the production runs [14], NMR restraints, when available, were applied in the early stages of the majority of the simulations of the pre-miR20b RNA (Table 1, sim. 3–7) and of the Rbfox•pre-miR20b complex simulations (Table 1, sim. 9–13). Specifically, after the initial standard equilibration, the systems were simulated in the following way: 0–100 ns—all available NMR hydrogen restraints (both inter- and intra-molecular NOE interactions) were utilized, 100–120 ns—only protein–RNA (intermolecular NOE) restraints were utilized, and after 120 ns—entirely unrestrained simulations were conducted. The aim of the procedure is to guarantee a sufficient equilibration of the systems before data is gathered. Since the restraints are lifted in the later stages of the simulations, they do not bias the results. Only the primary NMR data (NOE distance restraints) were used, and were introduced with a flat-well potential [14]. Earlier, this approach was shown to be able to prevent the abrupt structural disruptions which can otherwise occur in beginning of MD simulations of protein-RNA complexes. By giving the structures more time to relax without immediate deviations from the NMR ensemble, it is possible to achieve more stable simulations of protein-RNA complexes [14]. Some simulations were also performed without the initial use of NMR restraints (Table 1, sim. 2 and 8). For detailed discussion of this protocol see [14].

Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [65]. Periodic boundary conditions and a 2 fs integration step were employed. The particle mesh Ewald (PME) approach [66] was used for handling electrostatic interactions. The cut-off distance of the non-bonded Lennard-Jones interactions was 0.9 nm. We used the Nose–Hoover thermostat [67] and Andersen–Parrinello–Rahman barostat [68] to maintain the systems at a temperature of 298 K and pressure of 1 bar, respectively. The completely unrestrained simulations were performed using GROMACS 5.1 (<http://www.gromacs.org>) [69]. Simulation runs initially using the NMR restraints were performed with the pmemd module of AMBER 14 (<http://ambermd.org>) [70].

Replica Exchange simulations of the Rbfox*•pre-miR20b* complex

In order to provide proper sampling of the Rbfox*•pre-miR20b* binding interface conformational space, we performed two distinct Replica Exchange with Solute Scaling (REST2) [17] simulations. The method is based on a modification of the potential energy, so that the interactions between solute atoms are scaled by a factor λ , solvent–solvent interactions remain unscaled, and solute–solvent interactions are scaled by $\lambda^{1/2}$. Scaling the energy by a factor λ is equivalent to scaling of the temperature by $1/\lambda$. Thus, in the case of REST2, only the solute atoms are effectively heated up in REST2. Solvent–solvent interactions that typically contribute the most to the energy differences between replicas, do not contribute to exchanges, allowing to effectively reduce the number of replicas and the computational cost [17].

In a first simulation run (Table 1, sim. 15 (REST2 PS)), we explored the possibility to enhance sampling of the mutated binding interface only, by rescaling the force field parameters of the nucleotides A₃₀ and G₃₃ along with their flanking phosphates and the protein residues within 0.5 nm of those nucleotides (the complete list of included atoms is reported in S1 Table). Eight replicas were simulated with scaling factor λ ranging from 1 (reference replica) to 0.6, according to a geometric distribution, and leading to an average acceptance rate of 22%. Each replica was simulated for 2 μ s, giving a cumulative time of $8 \times 2 \mu\text{s} = 16 \mu\text{s}$. For this simulation, an in-house modified Amber 16 version was used and the same simulation setup described above was adopted.

The above-proposed simulation protocol requires decoupling the degrees of freedom of the binding interface from rest of the system, but this procedure might affect fundamental molecular properties such as electrostatics and hydrophobicity [71]. Therefore, to test the accuracy of the calculations, a second REST2 simulation was conducted using a standard protocol, namely rescaling the force-field parameters of the entire solute. In this case, the Hamiltonian Replica Exchange (H-REX) code [71] as implemented in the Plumed-HREX patch of Gromacs 5.1 (<https://plumed.github.io/doc-v2.3/user-doc/html/hrex.html>) [71] was used. Sixteen replicas of the system were simulated, with the setup described above. A geometrical distribution of sixteen λ values ranging from 1 to 0.7 was used, which resulted in an average acceptance rate of ~20%. Each replica was simulated for 1 μ s (Table 1, sim. 16 (REST 2)).

A cluster analysis was performed to identify the most populated conformers in the total simulated ensemble. In order to ensure that the clusters found would be consistent across both REST2 runs, clustering was performed on the combined trajectory obtained from the two reference (unbiased) trajectories. The *k-means* clustering algorithm implemented in cptraj module [72] of Amber 16 [60] was used based on the Root Mean Square Deviation (RMSD) of the interface of the protein-RNA complex (nucleotides 28–33 and the amino-acids residues within 0.45 nm from those nucleotides). The combined clustering results were also parsed to obtain results for each individual simulation [73, 74].

A representative structure for each cluster was identified as the centroid conformer of the cluster (i.e., the trajectory frame with the lowest cumulative RMSD distance to every other point in the cluster). Subsequently, two additional unbiased MD simulations (Table 1, sim. 17–18) were started from the representative structures of the two most populated clusters (accounting for ~44% of all structures). Here we used Gromacs 5.1 [69] and the same protocol described above.

To compare the conformational space sampled by the two REST2 simulations and their efficiency with respect to conventional MD, we estimated the probability density $p(x)$ of observing the system in a state x using a Gaussian kernel density estimate [75] (Gaussian KDE) along two collective variables (CV) [76].

Overall changes are described by the differences in the distribution of reciprocal inter-atomic distances (DRID)[77] with respect to the representative structure of the most populated cluster. The distribution is evaluated from the inverse intra-molecular distances between all the C_a and the P atoms of the protein and RNA. For each C_a and P inverse distance distribution, three features are extracted: the mean, the square root of the central moment, and the cube root of the third central moment. This assigns a feature matrix $\mathbf{v}_n \in \mathbb{R}^{3 \times N}$ to each configuration n . The difference between configuration n and the reference structure is then

$$\text{DRID} = \frac{1}{3N} \sum_{i=1}^N \|\mathbf{v}_n^{(i)} - \mathbf{v}_0^{(i)}\| \quad (1)$$

where N is the number of residues, $\mathbf{v}_n^{(i)}$ denotes the feature vector of the i th C_a or P atom in $\mathbf{v}_n$, and $\mathbf{v}_0$ is the feature matrix of the reference configuration.

Local changes were captured from the fraction of conserved contacts Q between the protein and the RNA at the binding interface. Q is defined via a list of contact pairs between the heavy atoms i of residues 28–33 of the RNA loop and the heavy atoms j of the protein residues:

$$Q(x) = \frac{1}{N} \sum_{(ij)} \frac{1}{1 + \exp[\beta(r_{ij} - \lambda r_{ij}^0)]} \quad (2)$$

where the sum runs over N pairs of contacts (ij) , $r_{ij}(x)$ is the distance between i and j in configuration x , r_{ij}^0 is the distance between i and j in the reference conformation, β is a smoothing parameter taken to be 0.5 nm and the factor λ , taken to be 1.8 as default [78], accounts for fluctuations when the contact is formed.

The DRID feature vector and the fraction of native contact were obtained using the MDtraj code (<http://mdtraj.org/>) [79].

MD and Free energy simulations of the S151T Rbfox*•pre-miR20b* complex

A model of the S151T Rbfox*•pre-miR20b* complex was prepared from the representative structure of the most populated cluster of the Rbfox*•pre-miR20b* complex simulations (see above for details). A threonine residue at position 151 was introduced using the Swiss MODEL software [56–59] and two standard independent MD simulations (Table 1, sim. 19–20) were conducted using the same protocol as described above.

The free energy difference associated with the S151T mutation ($\Delta\Delta G$) was computed according to the thermodynamics cycle equation: $\Delta\Delta G = \Delta G_{\text{co}} - \Delta G_{\text{s}} = \Delta G^{\text{S151T}} - \Delta G^{\text{S151T}}$. The ΔG_{co} and ΔG_{s} represent the results of the non-equilibrium alchemical calculations[52] of the S151T protein-RNA complex and of the free protein state, respectively. The ΔG^{S151T} and ΔG^{S151T} are the dissociation energy of the Rbfox*•pre-miR20b* and of the S151T Rbfox*•pre-miR20b* complex, respectively.

The free energy calculations were conducted in a following fashion: From the equilibrium production simulations of the Rbfox*•pre-miR20b* complex (Table 1, sim. 17) and of the Rbfox* protein (Table 1, sim. 14), 10 conformations were extracted equidistantly in time, and, for every configuration, a hybrid structure/topology for the S151T mutation was generated using the pmx utilities (<http://pmx.mpibpc.mpg.de/>) [51, 80]. Subsequently, a 1 ns MD simulation for every configuration was performed to equilibrate the velocities on the introduced dummy atoms.

From each equilibrium simulation, 100 snapshots were extracted equidistantly in time, and finally, a 200 ps (Table 1, sim. 21 (FE)) or 1 ns (Table 1, sim. 22 (FE)) alchemical transition was initiated to morph the system from one physical state to the other. The alchemical transformations were performed in both directions: S151 to S151T and vice versa. A soft-core function with the default parameters ($\alpha = 0.3$, $\sigma = 0.25$, $p = 1$) [51, 81] was used for the non-bonded interactions during the non-equilibrium transitions. The work values from the non-equilibrium transitions were used to calculate free energy differences based on the Crooks theorem [82] utilizing the maximum likelihood estimator [83]. The protocol described above was applied to all the alchemical simulations.

Table 1 reports the complete list of the simulations performed in this work (overall more than 50 μ s of molecular simulation).

Simulation analysis

Hydrogen bonds were analyzed using the cpptraj [72] module of AMBER 16 (available as AmberTools16 at <http://ambermd.org/AmberTools16-get.html>) [60]. We used a distance cut-off of 0.35 nm between the relevant heavy atoms and an angle cut-off of 135° for the intervening hydrogen atom. These interactions are characterized by the percentage of the trajectory during which they are observed (i.e. occupancy). Aromatic amino acids and nucleobases were considered to form stacking interactions if the distance between their centers of mass was less than 0.5 nm and the angle between the two planes was less than 30°.

RNAs base pair, base-pair steps and the ion distributions around the RNA helical axes in the simulated systems were analyzed with the Curves+ program [31] and the Canal and Canion utilities (available at https://bisi.ibcp.fr/tools/curves_plus/). Average ion molarities were calculated by setting the groove limit at a radius of 0.11 nm from the RNA helical axis, while the angular limits were determined by the average position of the sugar C1' atoms.

Deviations relative to the initial RNA structure were calculated using the ϵ RMSD metric [19], a recently suggested RNA-specific structural metrics that is considered more robust than the notoriously insensitive and ambiguous RMSD [84, 85]. Two structures with ϵ RMSD of 0.7 or lower are considered to be significantly similar [19]. The ϵ RMSD was calculated using the baRNABA package (available at <https://github.com/srnas/barnaba>).

The protein's deviations from the initial structure were analyzed in terms of the RMSD, calculated using the cpptraj [72] module of Amber 16 [60]. The protein backbone conformational plasticity was calculated in terms of PAD $_{\omega}$ angle from the T-PAD analysis (freely available upon request) [18]. The latter provides a quantitative analysis of local plasticity of individual residues in terms of the angular dispersion ω , which is the sum of the Ramachandran angles Φ and ψ . Moreover, it allows distinction between backbone local fluctuations and conformational transitions (from one region of the Ramachandran plot to another) even when they occur with the same amplitude [18]: the tag “F” is assigned to fluctuations, “T” to long transitions (i.e., contributing more than 30% of the simulation time) and “t” to short transitions (i.e., contributing less than 30% of the simulation time). This analysis has been successfully used in the past to evaluate proteins backbone fluctuations from MD simulation trajectories and NMR structures [86].

The conformational entropy has been estimated by calculating the chain's conformational entropy from the distribution of the backbone (ϕ , ψ) and side-chains rotameric angles, $[\chi_n]$ following ref. [87]. The calculation has been performed on the trajectories of Rbfox and Rbfox* free and in complex with RNA.

NMR observables

Protein's chemical shifts (CS) were predicted using SHIFTX2 v. 1.07 (<http://www.shiftx2.ca/>) [37, 38]. LARMOR^D software (<https://brooks.chem.lsa.umich.edu/>) [36] was used for RNA. In the SHIFTX2 program the sequence information is not used in the prediction, so that the predictions are identical to those of the SHIFTX+ program (<http://www.shiftx2.ca/performance.html>). We run SHIFTX2 and LARMOR^D on each frame extracted from the un-restrained simulations every 10 ps of the free pre-miR20b and of the Rbfox•pre-miR20b complex, for which experimental CS are available. The chemical shifts predictions for these 48,000 pre-miR20b and 48,000 Rbfox•pre-miR20b conformers were then linearly averaged to make a final prediction for the protein's $^{13}\text{C}_\alpha$, $^{13}\text{C}'$, $^{13}\text{C}_\beta$, ^{15}N and for the RNA's ^{13}C and ^1H CS.

For the set of experimental upper bound distance restraint r_{NOE} , the simulated NOE's $\langle r_{ij} \rangle$ were calculated according to:

$$\langle r_{ij} \rangle = \left(\frac{1}{N_f} \sum (r_{ij})^{-6} \right)^{-\frac{1}{6}} \quad (3)$$

where $r_{i,j}$ is the interatomic distance between atoms i and j , and the sum runs over the N_f trajectories frames. The average distance violation was defined as:

$$\frac{1}{N_{\text{NOE}}} \sum (r_{\text{NOE}} - \langle r_{ij} \rangle) \quad \text{if } r_{\text{NOE}} < \langle r_{ij} \rangle \quad (4)$$

where the sum runs over all reported intermolecular NOE-based distance restraints. The conformers with best match with the NOEs upper bounds were then selected to produce an “MD-adapted structure ensemble” using the same protocol as in [14]. In particular, we used the combined simulation trajectories of the Rbfox•pre-miR20b complex and from each we selected 10% of frames with fewest NOE violations. *K-means* clustering algorithm was used to cluster this group of frames based on the RMSD of the complex. The representative structures of the 20 clusters obtained constitute the “MD-adapted structure ensemble”: sets of atomic coordinates (deposited as PDB files at <https://doi.org/10.5281/zenodo.1297931>) that capture the flexibility and the conformers suggested by MD simulations while still retaining the highest possible level of agreement with the primary NMR data.

Supporting information

S1 Text. Are similar types of motion sampled similarly across different MD simulations?. (PDF)

S2 Text. Convergence of REST2 and REST2 PS simulations. (PDF)

S1 Fig. Overlap of principle components (PCs) for independent simulations. Histograms from PCs analysis in Cartesian space calculated from the trajectories with independent projection of the PCs on the separate trajectories of the pre-miR20b (A, Table 1, sim. 2–6) and (B) of the Rbfox•pre-miR20b (Table 1, sim. 9–13). (PDF)

S2 Fig. Root mean square deviation (RMSD). calculated over heavy atoms with respect to the initial structure of the Rbfox protein in simulations of (A) the free state (Table 1, sim.1) and (B) bound to pre-miR20b RNA (Table 1, sim. 2–7).

(PDF)

S3 Fig. Entropy differences between Rbfox* and Rbfox, free (F) and in complex with RNA (C). Details of the calculations are reported in the Materials and Methods section.

(PDF)

S4 Fig. (A) RNA backbone dihedral angles calculated over the aggregated simulations of the pre-miR20b Rbfox complex (Table 1, sim. 8–13). The green dots indicate the values of the angles in the lowest energy structure of the NMR ensemble of the Rbfox•pre-miR20b complex from which the simulations were started. (B) ϵ RMSD of the pre-miR20b loop sequences U₂₇GGCAUG₃₃ (left) and G₂₈GCAUG₃₃ (right) in complex with Rbfox versus time in the six MD simulations performed (Table 1, sim. 8–13).

(PDF)

S5 Fig. Base pair (bp) and base pair steps (bps) of pre-miR20b in complex with Rbfox. (A) Bp and (B) bps parameters for base pairs G₂₀-C₄₀, U₂₁-A₃₉, A₂₂-U₃₈, G₂₃-C₃₇, calculated over the aggregated simulations (Table 1, sim. 8–13; dark blue) and NMR ensemble (light blue).

(PDF)

S6 Fig. (A) 2D representation of the free pre-miR20b loop structure. (B) Time development of the ϵ RMSD of the pre-miR20b loop (r-U₂₇GGCAUG₃₃) in the three χ_{OL3} -CP-OPC MD simulations (Table 1, sim. 1–3 (χ_{OL3} -CP-OPC)). C. RNA backbone dihedral angle histograms calculated over the aggregated simulations. The green dots indicate the values of the angles in the lowest energy structure of the NMR ensemble 2n7x from which the simulations were started.

(PDF)

S7 Fig. Conformations of the pre-miR20b loop in simulations 2–7 (as listed in Table 1).

The loop conformation in the initial structure is shown as the grey overlay.

(PDF)

S8 Fig. Base pair (bp) and base pair steps (bps) of pre-miR20b RNA in the free state. (A) bp and (B) bps parameters for base pairs G₂₀-C₄₀, U₂₁-A₃₉, A₂₂-U₃₈, G₂₃-C₃₇, U₂₄-U₃₆, U₂₅-C₃₅, U₂₆-A₃₄, calculated over the entire MD (dark blue) and NMR (light blue) ensembles.

(PDF)

S9 Fig. Average Na⁺ distribution for the (A) free pre-miR20b (Table 1, simulations 2–7) and (B) its complex with the Rbfox protein (Table 1, simulations 8–13), as a function of the distance from the helical axis (R). The results are plotted as molarities as shown by the color bars, with blue to yellow scale indicating increasing values. The vertical white line indicates the radial position of the phosphorus atoms.

(PDF)

S10 Fig. Stacking interactions in the Rbfox•pre-miR20b complex. Stacking geometries are described by the center of mass distance d and the angle θ between the planes of the bases and amino acid side chain. Amino acid and nucleobase are considered stacked if $d < 0.5$ nm and $\theta < 30^\circ$. The distributions are calculated for G₂₉-R184, G₂₉-F126, U₃₂-H120 and G₃₃-F160 pairs in the individual unrestrained trajectories (Table 1, sim. 8–13).

(PDF)

S11 Fig. Calculated and experimentally measured [6] chemical shifts (CS) for pre-miR20b in the free state (A), for pre-miR20b in complex with the Rbfox (B) and for the Rbfox protein (C). The CS have been calculated using the SHIFTX+ [88] for the protein and LARMOR^D [36] for the RNA. Additional details are reported in the Materials and Methods section.
(PDF)

S12 Fig. Comparison between calculated and experimental [6] ^{13}C , ^{13}Ca , ^{13}Cb and ^{15}N CS of Rbfox in the free state and bound to pre-miR20b. The grey square represents the standard deviation. Additional details on the calculations are reported in the Materials and Methods section.
(PDF)

S13 Fig. Distributions of CS predicted by SHIFTX+ [88] for F150 and F158 ^{15}N from the MD simulations of the Rbfox in complex with pre-miR20b (see Materials and Methods for details) and values of the χ_1 angles of the same residues in trajectory 9. The blue dot represents the calculated average value; the red one corresponds to the experimental value. A similar behaviour is observed in the other simulations performed on the system.
(PDF)

S14 Fig. Comparison of calculated and experimental chemical shifts for the (A) ^{13}C and (B) ^1H atoms of pre-miR20b in the free state (Table 1, sim. 2–7). Representation as in S12 Fig.
(PDF)

S15 Fig. Comparison of calculated and experimental chemical shifts for the ^{13}C and ^1H atoms of pre-miR20b bound to Rbfox (Table 1, sim. 8–13). Representation as in S12 Fig.
(PDF)

S16 Fig. Stacking interactions in the Rbfox*•pre-miR20b* complex. The distributions are calculated for G₂₉-R184, G₂₉-F126, U₃₂-H120 and C₃₃-F160 pairs in replica exchange (Table 1, sims. 15–16) and unbiased MD (Table 1, sims. 17–18) simulations.
(PDF)

S17 Fig. Probability density of sampling the conformational landscape of the Rbfox*•pre-miR20b* complex in the two-dimensional space of Native Contacts and of DRID (see Method sections for details) for plain MD, conventional (standard) REST2 and REST2 with partial scaling (REST2 PS), respectively.
(PDF)

S18 Fig. Rbfox*•pre-miR20b* complex. (A) Top view of the structure in the simulation (Table 1, sim. 23). (B) Details of the stacking interactions of G₂₉ with F126 and R184.
(PDF)

S19 Fig. Rbfox*•pre-miR20b complex. (A) Top view of the structure in the simulation (Table 1, sim. 24). (B) Details of the H-bond interactions (dashed blue lines) involving G₃₃, R147 and D118.
(PDF)

S1 Table. Selected region for the “Partial Scaling” REST2 simulation of the Rbfox*•pre-miR20b* complex.
(PDF)

S2 Table. NOE-upper bounds violations (ν) for pre-miR20b in the free state (Table 1, sim. 2–7), Rbfox (Table 1, sim. 1) and Rbfox*•pre-miR20b complex (Table 1, sim. 8–13) calculated along all individual MD trajectories and on the “full ensemble” obtained by merging

all the unrestrained parts of the initially restrained trajectories.
(PDF)

Acknowledgments

AB and PC acknowledge the JARA-HPC computer time on the supercomputer JUQUEEN at Forschungszentrum Jülich, Jülich, Germany.

Author Contributions

Conceptualization: Anna Bochicchio, Miroslav Krepl, Gabriele Varani, Jiri Sponer, Paolo Carloni.

Data curation: Anna Bochicchio, Miroslav Krepl, Fan Yang.

Funding acquisition: Gabriele Varani, Jiri Sponer, Paolo Carloni.

Investigation: Anna Bochicchio, Miroslav Krepl, Fan Yang, Gabriele Varani, Jiri Sponer, Paolo Carloni.

Methodology: Anna Bochicchio, Miroslav Krepl, Jiri Sponer, Paolo Carloni.

Project administration: Gabriele Varani, Jiri Sponer, Paolo Carloni.

Resources: Gabriele Varani, Jiri Sponer, Paolo Carloni.

Supervision: Gabriele Varani, Jiri Sponer, Paolo Carloni.

Validation: Anna Bochicchio, Miroslav Krepl, Gabriele Varani, Jiri Sponer, Paolo Carloni.

Writing – original draft: Anna Bochicchio, Miroslav Krepl, Gabriele Varani, Jiri Sponer, Paolo Carloni.

Writing – review & editing: Anna Bochicchio, Miroslav Krepl, Fan Yang, Gabriele Varani, Jiri Sponer, Paolo Carloni.

References

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001; 291(5507):1304–51. <https://doi.org/10.1126/science.1058040> PMID: 11181995
2. Birney E, Kumar S, Krainer AR. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Research*. 1993; 21(25):5803–16. PMID: 8290338
3. Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *The EMBO Journal*. 2006; 25:163–73. <https://doi.org/10.1038/sj.emboj.7600918> PMID: 16362037
4. Burd C, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science*. 1994; 265:615–21. PMID: 8036511
5. Maris C, Dominguez C, Allain FH-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*. 2005; 272:2118–31. <https://doi.org/10.1111/j.1742-4658.2005.04653.x> PMID: 15853797
6. Chen Y, Zubovic L, Yang F, Godin K, Pavelitz T, Castellanos J, et al. Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Research*. 2016; 44(9):4381–95. <https://doi.org/10.1093/nar/gkw177> PMID: 27001519
7. Maris C, Dominguez C, Allain FHT. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*. 2005; 272:2118–31. <https://doi.org/10.1111/j.1742-4658.2005.04653.x> PMID: 15853797

8. Chen Y, Yang F, Zubovic L, Pavelitz T, Yang W, Godin K, et al. Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nature Chemical Biology*. 2016; 12(9):717–23. <https://doi.org/10.1038/nchembio.2128> PMID: 27428511
9. Cléry A, Blatter M, Allain FHT. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*. 2008; 18(3):290–8. <https://doi.org/10.1016/j.sbi.2008.04.002> PMID: 18515081
10. Afroz T, Cienikova Z, Cléry A, Allain FH. One, two, three, four! How multiple RRM reads the genome sequence. *Methods in enzymology*. 558: Elsevier; 2015. p. 235–78. <https://doi.org/10.1016/bs.mie.2015.01.015> PMID: 26068744
11. Krichevsky AM, Gabriely G. miR-21: a small multi-faceted RNA. *Journal of Cellular and Molecular Medicine*. 2009; 13(1):39–53. <https://doi.org/10.1111/j.1582-4934.2008.00556.x> PMID: 19175699
12. Šponer J, Krepl M, Banáš P, Kůhrová P, Zgarbová M, Jurečka P, et al. How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *Wiley Interdisciplinary Reviews: RNA*. 2017; 8(3).
13. Sponer J, Bussi G, Krepl M, Banas P, Bottaro S, Cunha RA, et al. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chemical reviews*. 2018; 118(8):4177–338. <https://doi.org/10.1021/acs.chemrev.7b00427> PMID: 29297679
14. Krepl M, Cléry A, Blatter M, Allain FHT, Sponer J. Synergy between NMR measurements and MD simulations of protein/RNA complexes: application to the RRM, the most common RNA recognition motifs. *Nucleic Acids Research*. 2016; 44:6452–70. <https://doi.org/10.1093/nar/gkw438> PMID: 27193998
15. Krepl M, Blatter M, Cléry A, Damberger FF, Allain FHT, Sponer J. Structural study of the Fox-1 RRM protein hydration reveals a role for key water molecules in RRM–RNA recognition. *Nucleic Acids Research*. 2017; 45(13):8046–63. <https://doi.org/10.1093/nar/gkx418> PMID: 28505313
16. dit Konte ND KM, Damberger FF, Ripin N, Duss O, Sponer J, Allain FH-T. Aromatic side-chain conformational switch on the surface of the RNA Recognition Motif enables RNA discrimination. *Nature Communication*. 2017;8.
17. Wang LaF, Richard A and Berne BJ. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *The Journal of Physical Chemistry B*. 2011; 115:9431. <https://doi.org/10.1021/jp204407d> PMID: 21714551
18. Caliendo R, Rossetti G., Carloni P. Local Fluctuations and Conformational Transitions in Protein. *Journal of Chemical Theory and Computation*. 2012; 8:10.
19. Bottaro S, Di Palma F, Bussi G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Research*. 2014; 42:13306–14. <https://doi.org/10.1093/nar/gku972> PMID: 25355509
20. Krepl M, Havrila M, Stadlbauer P, Banas P, Otyepka M, Pasulka J, et al. Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein–RNA Complexes? *Journal of Chemical Theory and Computation*. 2015; 11:1220–43. <https://doi.org/10.1021/ct5008108> PMID: 26579770
21. Steinbrecher T, Latzer J, Case DA. Revised AMBER Parameters for Bioorganic Phosphates. *Journal of Chemical Theory and Computation*. 2012; 8(11):4405–12. <https://doi.org/10.1021/ct300613v> PMID: 23264757
22. Izadi S, Anandakrishnan R, Onufriev AV. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters*. 2014; 5(21):3863–71. <https://doi.org/10.1021/jz501780a> PMID: 25400877
23. Oubridge C IN, Evans PR, Teo C-H, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*. 1994; 372:432–8. <https://doi.org/10.1038/372432a0> PMID: 7984237
24. Kurisaki I, Takayanagi M, Nagaoka M. Combined Mechanism of Conformational Selection and Induced Fit in U1A–RNA Molecular Recognition. *Biochemistry*. 2014; 53:3646–57. <https://doi.org/10.1021/bi401708q> PMID: 24828852
25. Blakaj DM, McConnell KJ, Beveridge DL, Baranger AM. Molecular Dynamics and Thermodynamics of Protein–RNA Interactions: Mutation of a Conserved Aromatic Residue Modifies Stacking Interactions and Structural Adaptation in the U1A–Stem Loop 2 RNA Complex. *Journal of the American Chemical Society*. 2001; 123:2548–51. PMID: 11456923
26. Kůhrová P, Best RB, Bottaro S, Bussi G, Sponer J, Otyepka M, et al. Computer folding of RNA tetraloops: identification of key force field deficiencies. *Journal of Chemical Theory and Computation*. 2016; 12(9):4534–48. <https://doi.org/10.1021/acs.jctc.6b00300> PMID: 27438572
27. Zirbel CL, Šponer JE, Šponer J, Stombaugh J, Leontis NB. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic acids research*. 2009; 37(15):4898–918. <https://doi.org/10.1093/nar/gkp468> PMID: 19528080
28. Bergonzo C, Cheatham TE. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *Journal of Chemical Theory and Computation*. 2015; 11(9):3969–72. <https://doi.org/10.1021/acs.jctc.5b00444> PMID: 26575892

29. Bottaro S, Bussi G, Kennedy SD, Turner DH, Lindorff-Larsen K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science Advances*. 2018; 4(5).
30. Kuhrova P, Mlynsky V, Zgarbova M, Krepl M, Bussi G, Best RB, et al. IMPROVING THE PERFORMANCE OF THE RNA AMBER FORCE FIELD BY TUNING THE HYDROGEN-BONDING INTERACTIONS. *bioRxiv*. 2018.
31. Lavery R, Maddocks JH, Pasi M, Zakrzewska K. Analyzing ion distributions around DNA. *Nucleic Acids Res*. 2014; 42(12):8138–49. <https://doi.org/10.1093/nar/gku504> PMID: 24906882
32. Leontis NB. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research*. 2002; 30:3497–531. PMID: 12177293
33. Auffinger P, Hashem Y. Nucleic acid solvation: from outside to insight. *Current Opinion in Structural Biology*. 2007; 17(3):325–33. <https://doi.org/10.1016/j.sbi.2007.05.008> PMID: 17574833
34. Auffinger P, Westhof E. Water and ion binding around RNA and DNA (C,G) oligomers11 Edited by I. Tinoco. *Journal of Molecular Biology*. 2000; 300(5):1113–31. <https://doi.org/10.1006/jmbi.2000.3894> PMID: 10903858
35. Krasovska MV, Sefcikova J, Réblová K, Schneider B, Walter NG, Šponer J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophysical Journal*. 2006; 91(2):626–38. <https://doi.org/10.1529/biophysj.105.079368> PMID: 16617077
36. Frank AT, Law SM, Brooks CL. A Simple and Fast Approach for Predicting ¹H and ¹³C Chemical Shifts: Toward Chemical Shift-Guided Simulations of RNA. *The Journal of Physical Chemistry B*. 2014; 118(42):12168–75. <https://doi.org/10.1021/jp508342x> PMID: 25255209
37. Neal S NA, Zhang H, Wishart DS. Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts. *Journal of Biomolecular NMR*. 2003; 26(3):215–40. PMID: 12766419
38. Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR*. 2011; 50(1):43. <https://doi.org/10.1007/s10858-011-9478-4> PMID: 21448735
39. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*. 2002; 116(20):9058–67.
40. Affentranger R, Tavernelli I, Di Iorio EE. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *Journal of Chemical Theory and Computation*. 2006; 2(2):217–28. <https://doi.org/10.1021/ct050250b> PMID: 26626508
41. Sanbonmatsu K, Garcia A. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Structure, Function, and Bioinformatics*. 2002; 46(2):225–34.
42. Henriksen NM, Roe DR, Cheatham TE. Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*. 2013; 117(15):4014–27. <https://doi.org/10.1021/jp400530e> PMID: 23477537
43. Bergonzo C, Henriksen NM, Roe DR, Cheatham TE. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*. 2015; 21(9):1578–90. <https://doi.org/10.1261/rna.051102.115> PMID: 26124199
44. Sugita Y, Okamoto Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chemical Physics Letters*. 2000; 329(3–4):261–70.
45. Itoh SG, Okumura H, Okamoto Y. Replica-exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules. *The Journal of Chemical Physics*. 2010; 132(13):134105. <https://doi.org/10.1063/1.3372767> PMID: 20387919
46. Liu P, Kim B, Friesner RA, Berne B. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(39):13749–54. <https://doi.org/10.1073/pnas.0506346102> PMID: 16172406
47. Fajer M, Hamelberg D, McCammon JA. Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic integration. *Journal of Chemical Theory and Computation*. 2008; 4(10):1565–9. <https://doi.org/10.1021/ct800250m> PMID: 19461870
48. Zacharias M. Combining elastic network analysis and molecular dynamics simulations by hamiltonian replica exchange. *Journal of chemical theory and computation*. 2008; 4(3):477–87. <https://doi.org/10.1021/ct7002258> PMID: 26620788
49. Vreede J, Wolf MG, de Leeuw SW, Bolhuis PG. Reordering hydrogen bonds using Hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems. *The Journal of Physical Chemistry B*. 2009; 113(18):6484–94. <https://doi.org/10.1021/jp809641j> PMID: 19358572

50. Lee MS, Olson MA. Protein folding simulations combining self-guided Langevin dynamics and temperature-based replica exchange. *Journal of Chemical Theory and Computation*. 2010; 6(8):2477–87. <https://doi.org/10.1021/ct100062b> PMID: 26613500
51. Gapsys V, Michielssens S, Seeliger D, de Groot BL. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angewandte Chemie*. 2016; 128(26):7490–4.
52. Gapsys V, de Groot BL. Alchemical Free Energy Calculations for Nucleotide Mutations in Protein–DNA Complexes. *Journal of Chemical Theory and Computation*. 2017; 13(12):6275–89. <https://doi.org/10.1021/acs.jctc.7b00849> PMID: 29125747
53. Seeliger D, Buelens FP, Goette M, de Groot BL, Grubmüller H. Towards computational specificity screening of DNA-binding proteins. *Nucleic acids research*. 2011; 39(19):8281–90. <https://doi.org/10.1093/nar/gkr531> PMID: 21737424
54. Goette M, Grubmüller H. Accuracy and convergence of free energy differences calculated from non-equilibrium switching processes. *Journal of computational chemistry*. 2009; 30(3):447–56. <https://doi.org/10.1002/jcc.21073> PMID: 18677708
55. Tsuda K MY, Inoue M, Kigawa T, Terada T, Shirouzu M, Yokoyama S. Solution structure of RNA binding domain in RNA binding motif protein 9. 2005.
56. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*. 2014; 42(W1):252–8.
57. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*. 2009; 37(Database issue):387–92.
58. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006; 22(2):195–201. <https://doi.org/10.1093/bioinformatics/bti770> PMID: 16301204
59. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*. 2009; 30(S1):162–73.
60. Case DA BR, Botello-Smith W, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Luchko T, Luo R, Madej B, Mermelstein D, Merz KM, Monard G, Nguyen H, Nguyen HT, Omelyan I, Onufriev A, Roe DR, Roitberg A, Sagui C, Simmerling CL, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM, Kollman PA. AMBER 2016. San Francisco: University of California; 2016.
61. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics*. 1983; 79(2):926–35.
62. Joung IS, Cheatham TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B*. 2008; 112(30):9020–41. <https://doi.org/10.1021/jp8001614> PMID: 18593145
63. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. 2015; 11:3696–713. <https://doi.org/10.1021/acs.jctc.5b00255> PMID: 26574453
64. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE 3rd, et al. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation*. 2011; 7(9):2886. <https://doi.org/10.1021/ct200162x> PMID: 21921995
65. Leimkuhler BJ, Skeel RD. Symplectic Numerical Integrators in Constrained Hamiltonian Systems. *Journal of Computational Physics*. 1994; 112(1):117–25.
66. Darden T, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993; 98(12):10089–92.
67. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*. 1985; 31(3):1695.
68. Parrinello M RA. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981; 52(12):7182.
69. Abraham MJ, Murtola T., Schulz R., Pall 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):6.
70. D.A. Case RMB, Botello-Smith W., Cerutti D.S., Cheatham T.E. III, Darden T.A., Duke R.E., Giese T.J., Gohlke H., Goetz A.W., Homeyer N., Izadi S., Janowski P., Kaus J., Kovalenko A., Lee T.S., LeGrand S., Li P., Lin C., Luchko T., Luo R., Madej B., Mermelstein D., Merz K.M., Monard G., Nguyen H., Nguyen H.T., Omelyan I., Onufriev A., Roe D.R., Roitberg A., Sagui C., Simmerling C.L., Swails J.,

- Walker R.C., Wang J., Wolf R.M., Wu X., Xiao L., York D.M. and Kollman P.A. AMBER 2014. San Francisco: University of California; 2014.
71. Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics*. 2014; 112(3–4):379–84.
72. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*. 2013; 9:3084–95. <https://doi.org/10.1021/ct400341p> PMID: 26583988
73. Roe DR, Bergonzo C., Cheatham T. E. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *The Journal of Physical Chemistry B*. 2014; 118(13):3543–52. <https://doi.org/10.1021/jp4125099> PMID: 24625009
74. Bergonzo C, Henriksen NM, Roe DR, Swails JM, Roitberg AE, Cheatham TE. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *Journal of Chemical Theory and Computation*. 2014; 10:492–9. <https://doi.org/10.1021/ct400862k> PMID: 24453949
75. Parzen E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*. 1962; 33(3):1065–76.
76. Westerlund AM, Harpole TJ, Blau C, Delemotte L. Inference of Calmodulin's Ca²⁺-Dependent Free Energy Landscapes via Gaussian Mixture Model Validation. *Journal of Chemical Theory and Computation*. 2018; 14(1):63–71. <https://doi.org/10.1021/acs.jctc.7b00346> PMID: 29144736
77. Zhou T, Caflisch A. Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *Journal of Chemical Theory and Computation*. 2012; 8(8):2930–7. <https://doi.org/10.1021/ct3003145> PMID: 26592131
78. Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*. 2013; 110(44):17874.
79. McGibbon Robert T, Beauchamp Kyle A, Harrigan Matthew P, Klein C, Swails Jason M, Hernández Carlos X, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*. 2015; 109(8):1528–32. <https://doi.org/10.1016/j.bpj.2015.08.015> PMID: 26488642
80. Gapsys V, Michielssens S, Seeliger D, de Groot BL. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry*. 2015; 36(5):348–54. <https://doi.org/10.1002/jcc.23804> PMID: 25487359
81. Gapsys V, Seeliger D, de Groot BL. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *Journal of Chemical Theory and Computation*. 2012; 8(7):2373–82. <https://doi.org/10.1021/ct300220p> PMID: 26588970
82. Crooks GE. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*. 1999; 60(3):2721–6.
83. Shirts MR, Bair E., Hooker G., Pande V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters*. 2003; 91(14):140601. <https://doi.org/10.1103/PhysRevLett.91.140601> PMID: 14611511
84. Cruz JA, Blanchet M-F, Boniecki M, Bujnicki JM, Chen S-J, Cao S, et al. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18(4):610–25. <https://doi.org/10.1261/rna.031054.111> PMID: 22361291
85. Parisien M CJ, Westhof E, Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*. 2009; 15(10):1875–85. <https://doi.org/10.1261/rna.1700409> PMID: 19710185
86. Dibenedetto D, Rossetti G, Caliendo R, Carloni P. A Molecular Dynamics Simulation-Based Interpretation of Nuclear Magnetic Resonance Multidimensional Heteronuclear Spectra of α -Synuclein-Dopamine Adducts. *Biochemistry*. 2013; 52(38):6672–83. <https://doi.org/10.1021/bi400367r> PMID: 23964651
87. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proceedings of the National Academy of Sciences*. 2014; 111(43):15396.
88. Neal S, Nip AM, Zhang H, Wishart DS. Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts. *Journal of Biomolecular NMR*. 2003; 26(3):215–40. PMID: 12766419