RNA-protein complexes and force field polarizability

*Hanna Baltrukevich[1], Piia Bartos[2]**

[1]Jagiellonian University in Krakow, Faculty of Pharmacy, ul. Gołębia 24, 31-007 Kraków, Poland

[2]University of Eastern Finland, School of Pharmacy, Faculty of Health Sciences, Yliopistonranta 1C, 70210 Kuopio, Finland

*To whom correspondence should be addressed: Piia.Bartos@uef.fi

ABSTRACT (max 200 words)

Molecular dynamic (MD) simulations offer a way to study biomolecular interactions and their dynamics at the atomistic level which is not possible with the current experimental methods. There are only a few studies of RNA-protein complexes in MD simulations, and here we wanted to study how force fields differ when simulating RNA-protein complexes. We tested three non-polarizable force fields: Amber protein force fields ff14SB and ff19SB with RNA force field OL3, and the all-atom OPLS4 force field. Due to the highly charged and polar nature of RNA, we also tested the polarizable AMOEBA force field. Our results show that the non-polarizable force fields overestimate the electrostatic interactions between the RNA and the protein which leads to compact and stable complexes. This effect is not seen in the polarizable force field, but it is computationally much more demanding. As a conclusion, all the tested force fields can be used to simulate RNA-protein complexes. If computational resources are not limited or if lengthy simulations are not needed for the studied problem, the polarizable force field AMOEBA is the best choice.

Introduction

Molecular dynamics (MD) simulations are routinely used to study the structure and dynamics of biomolecules at the atomistic level. Even though the models are by their very nature wrong in many ways, they are useful in showing us atomistic details of phenomena which cannot be directly observed experimentally.[1] MD simulations have led to advances in drug and enzyme design and material science, and they have greatly increased our understanding of the interactions of biomolecules at the atomistic level.

During the last few years, there have been some studies published about RNA-protein complex simulations.[2–15] While this field is starting to gain interest, it is unfortunate to see that there are no studies published which would compare different force fields in studying RNA-protein complexes. The selection of force field and other simulation parameters depends on the studied system,[16] and thus some time should be spent testing suitable options for each study case. One can argue that polarizability of a force field could help in simulations containing RNA, as strong electrostatic interactions are in dominant role in these systems.[17] This has made it difficult to optimize the parameters for the traditional fixed charge non-polarizable force fields for nucleic acids. In this study, we wanted to find if there are differences between traditional non-polarizable point-charge force fields and a polarizable force field. The test case used in this study was complex of argonaute-2 (Ago2) with two strands of RNA.

Ago2 is a crucial component of the RNA-induced silencing complex (RISC) which when bound to an RNA molecule inhibits gene expression.[18] The preprocessed structure of Ago2 in complex with a guide and target RNA strands which we used for the simulations is shown in Figure 1.[19] The guide RNA binds to Ago2 mainly from the 5' end which contains the seed sequence (nucleotides 2-7) which are crucial to the binding to the complementary mRNA strand.[20] The other end of the RNA, 3' end is bound to the most flexible part of Ago2, the PAZ domain.[11] The target RNA is mainly forming interactions with the guide RNA strand, with a few flanking nucleotides interacting with the Ago2 surface.

We tested the MD simulations of an Ago2-RNA complex using three traditional non-polarizable force fields: Amber RNA force field OL3[21] with protein force fields ff14SB[22] and ff19SB,[23] the all-atom force field OPLS4,[24] and the polarizable AMOEBA[17,25] force field. The results show that the non-polarizable force fields tend to over-estimate the electrostatic interactions between the polarized RNA backbone and the protein. This leads to less freedom of movement for both molecules and might hide some biologically relevant conformational changes. The polarizable force field shows more flexibility especially in the RNA backbone and smaller number of H bonds between the protein and the RNA. However, there is a significant loss of simulation speed when using a polarizable force field.

Methods

System preparation

The complex used for the simulations is the crystal structure of Ago2 bound to guide and target RNA 4W5O in PDB (Protein Data Bank).[19] The missing nucleotides in the PDB model were built manually in PyMOL (Version 2.5.1).[26] The missing loops in the protein structure were modelled in BioLuminate (Schrödinger 2021.3), and afterwards the built loops were refined using Prime (Schrödinger 2021.3).[27] The N-terminal was not present in the crystal structure and is unstructured in all argonaute crystal structures. It was thus omitted and the residue numbering starts from 22. N-termini was capped by adding ACE (N-acetyl group), while C-termini was left as a charged carboxyl group. Both 5´-phosphate and 3´-phosphate were eliminated from the ends of RNA molecules to prevent possible errors due to the lack of parametrization for them in the Amber force fields. The model was prepared in Protein Preparation Wizard (Schrodinger 2021.3)[27] - the water molecules were deleted beyond 5.0 Å from het groups, the states of het groups were generated using Epik in pH 7.4 ± 2.0; the optimization of H-bonds was performed with PROPKA[28,29] in pH 7.4 followed by restrained minimization of them in OPLS4.[24] The system was then converted to Amber atom names using the pdb4amber script.

Amber simulations – ff14SB and ff19SB

The Amber simulation systems were prepared with the tleap tool of AmberTools21.[30] The RNA-protein complex was solvated with truncated octahedron water box with 15Å buffer using 0.15M NaCl solution with SPCE or OPC water model for ff14SB and ff19SB protein force fields, respectively. The ions used parameters specified with the corresponding water model, and the RNA was parameterized with the RNA

force field OL3.[21] To use the larger timestep (4 fs), the hydrogen masses were repartitioned to the carbon atoms using ParmEd software.[31]

For the ff19SB[23] simulations, the minimization and equilibration steps were following: 1) all non-water atoms constrained, 2) heavy atoms constrained, 3) protein back bone constrained and 4) no constraints. The constraint force was 50 kcal/mol in the minimizations and 10 kcal/mol in the equilibration simulations. The minimizations 1-4 used the steepest descent algorithm with a maximum of 10,000 steps. The equilibration steps 1-3 consisted of 400 ps simulations, and step 4 was a 4,000 ps simulation. In the first equilibration step, the system was heated to 310 K. The temperature and pressure were maintained with Langevin thermostat and Berendsen barostat in the equilibration simulations. For the ff14SB, steps 2-4 were used in the minimization procedure and steps 2 and 4 for the equilibration runs of 20 ps and 2000 ps, respectively. The simulations followed similar parameters as the ones described for ff19SB.

During the production runs for each system the NPT ensemble was used: the 1.0 bar pressure was maintained with the help of Monte Carlo barostat and 310 K temperature was controlled by Langevin thermostat. Frames were recorded each 0.1 ns. The files were made ready for analysis by aligning and centering the complex, stripping away water molecules and writing the output in the format of xtc (compressed Gromacs trajectory) using cpptraj tool. The Amber production simulations were 4*500 ns in length for both force fields, resulting in 2 μs of total simulation time.

Desmond simulations – OPLS4

The RNA-protein complex was solvated with truncated octahedron water box with 15Å buffer using 0.15M NaCl solution with SPCE water model. The simulations consisted of the default minimization and relaxation protocol, and parameters for the production run were like those of the Amber simulations: recording interval each 100 ps, ensemble class NPT with temperature 300 K and pressure 1.01325 bar. The Desmond simulations were 8*250 ns in length, resulting in 2 μs of total simulation time.

OpenMM simulations – AMOEBA

The RNA-protein complex solvated with the Amber protocol was used to start the AMOEBA simulations in OpenMM.[32] To avoid clashes and high energy conformations in this slower force field, we used the last frame of an Amber trajectory as the starting conformation. The system was then minimized using Verlet Integrator with 1 ps timestep for a maximum of 100 iterations. The production simulations were run with the default parameters using Langevin integrator, 1/ps collision frequency and 2 fs timestep. The total simulation time was 10*10ns for the initial short simulations and the longer simulations were 2*100ns (exactly 106.2 ns and 105.6 ns). The simulation frames were saved every 0.02 ns.

Analysis of simulations – cpptraj

The resulting simulations were stripped from water atoms, wrapped into a single periodic box, centered around the protein and converted to .xtc-format to save disk space. These stripped simulations were then analyzed using the in-house cpptraj scripts to calculate the RMSD, RMSF, radius of gyration and hydrogen bond count. The principal component analysis (PCA) was carried out using the coordinates of the heavy atoms of the guide RNA on each snapshot, and the average coordinates of each respective simulation set was used as the reference.

Results

System flexibility and fluctuations – RMSD and RMSF

The flexibility of the protein and RNA strands was calculated using the root-mean-square deviation (RMSD) to the crystal structure. The results were calculated separately for the protein backbone (Cα, C, N, O atoms), guide RNA backbone (sugar + phosphate moieties) and target RNA backbone (sugar + phosphate moieties), and they are presented in Table 1. The distribution of RMSD values in simulations is depicted at the frequency polygon plots for protein, guide and target RNA backbones in Figure 2. The root-mean-square fluctuations of Cα atoms of the protein and the C3’ atoms of the guide RNA are shown in Figure 2.

The tight packing of the RNA and the protein are most clearly seen in the ff14SB protein force field simulations. Previously, ff14SB protein force field has been criticized to overestimate the helical content due

to TIP3P water model which leads to too tight packing of the protein core and less freedom of movement.[23] By changing the protein force field to ff19SB and the water model to OPC, the movements of the RNA are also affected. Due to the more freely moving protein, the RNA can also display more flexibility.

The OPLS4 force field also displays a relatively rigid protein, like in ff14SB. The protein is still fluctuating more than in the ff14SB, and both RNA strands display a larger deviation from the starting structure (Figure 2). The protein and the RNA strands are also fluctuating more than in the Amber force fields (Figure 3).

The AMOEBA 10*10ns simulations display the lowest RMSD values among all the simulations, which is most likely an artifact arising from the very short simulation time. Because it is possible that the shorter simulations are not displaying equilibrium conditions, we ran two longer AMOEBA simulations (~100ns) to get a better overview of the RNA and protein movements. Indeed, the AMOEBA 2*100ns simulations display large deviations from the starting structure (Table 1) and more fluctuations than the short AMOEBA simulations (Figure 3).

Radius of gyration

A visual inspection of the trajectories reveals that the differences observed in the RMSD values arise from different behavior of the system in the traditional and polarizable force fields. This compacting of the RNA-protein complex is clearly visible in the maximum radius of gyration calculated for the simulation snapshots (Figure 4). In this analysis, ff14SB and OPLS4 seem to form the most compact complexes. Even in the AMOEBA 10*10ns simulations, a shift to larger radius of gyration can be observed. With the longer AMOEBA simulations, the relaxation of the system is even more obvious.

The simulations with Amber ff19SB protein force field do not show as much compacting as ff14SB or OPLS4 which can be seen in the overlaid snapshots in Figure 5A. The unexpected compacting of OPLS4 simulations is due to a formation of a strong hydrogen bond between the terminal 5' OH-group of the guide RNA U1 and either of the phosphoryl oxygens of the target RNA A2 (Figure 5B). This behavior is common, as it was observed in 7/8 of the OPLS4 simulations. If the hydrogen bond forms, it will remain until the end of the simulation. In the other force fields, and in the single OPLS4 simulation, the terminal 5´ OH-group of the guide RNA U1 interacts with the protein residues Tyr509 or Gln525.

Principal component analysis - PCA

To study this opening-closing dynamic more closely, we conducted principal component analysis (PCA) of the guide RNA coordinates. This clearly shows that the 5´ end of the guide RNA is moving outwards only in the polarizable AMOEBA force field (Figure 6) (only the AMOEBA 10*10ns simulation data was used for this analysis). In the other force fields the main movements of the guide RNA are observed towards the 3´ end which is either not closely bound to the protein or is bound to the very flexible PAZ domain.

The number of hydrogen bonds

All the force fields show similar numbers for the hydrogen bonds between the RNA strands (Table 2). Theoretically, there should be 17 hydrogen bonds between the seed region of the guide RNA (nt 2-9) and the target RNA. As expected, during the simulations some of those bonds stretch or twist beyond the criteria specified here and thus there is some fluctuation in the numbers, but overall, the numbers remain close to the theoretical number.

The number of hydrogen bonds between the protein and the guide RNA is shown in Table 2 and Figure 6 shows frequency polygons of hydrogen bond distribution. The amount of hydrogen bonds in the AMOEBA 10*10ns simulations is similar to those observed in the non-polarizable force fields. In the longer AMOEBA simulations, where the system has more time to relax and open, the amount of hydrogen bonds between the guide RNA and the protein goes down by ~10 bonds. This indicates that the non-polarizable force fields over-stabilize the hydrogen bonding at the protein-RNA interface which has been reported before.[3,16,33]

Discussion

All the studied force fields can be used with RNA-protein complexes. All force fields produced reasonable simulations without major artefacts that could be traced back to the parameterization of the force fields. The RNA-protein complexes are chemically and biologically special systems, for which there are no specifically

tailored parameters in any currently available force field. Thus, at least for now the simulations need to be run as a combination of two different force fields or an all-atom force field which is a compromise of a many different compounds.

Checking the simulation trajectories visually is a necessary step for the force field selection. Based on the numerical values of RMSD, RMSF, radius of gyration and hydrogen bond count, the all-atom force field OPLS4 displays similar performance to the Amber force fields. However, the visual checking of the trajectories revealed, that the usage of the OPLS4 force field lead to the most compact RNA-protein complex due to a hydrogen bond forming between the 5´ OH group and a phosphoryl oxygen on the target RNA (Figure 4B). It is possible that the formation of this hydrogen bond could be prevented with more careful water/ion placement. Generally, the 5´ terminus of small RNA molecules that bind to Ago2 is phosphorylated, as in the crystal structure that we used in this study.[19] This phosphate group was removed in our simulations due to missing parameters in the Amber OL3 force field. We agree that for more precise RNA-Ago2 interaction calculations, the 5´ phosphate parameters need to be developed also for these force fields.

The protein force field affects RNA-protein complex dynamics. Even though both Amber simulation sets used the same RNA force field, the complexes behaved slightly differently. It is known that the Amber ff14SB force field underestimates helicity which in connection to 3-point water models leads to overly compact protein structures.[23] This behavior is enhanced when RNA is bound to the protein, as the electrostatic interactions with the RNA backbone make the complex even more compact. Changing the protein force field to ff19SB and the water model to the 4-point OPC alleviate the problem, as the results show more flexible protein backbone movement and more open conformation based on the radius of gyration. Using a more modern protein force field and a more advanced water model leads, at least in the case of this RNA-protein complex, to a more realistic behavior of the complex.

The non-polarizable force fields lead to too tight packing of RNA-protein complexes. The all-atom force field OPLS4 displayed similar maximum radius of gyration as the older ff14SB Amber protein force field. Both simulation sets used the older 3-point water model SPC/E. In Amber ff19SB simulations we used OPC, a more advanced 4-point water model that has been reported to improve the results of biomolecular simulations.[34] It seems that using this water model in combination with the newer Amber ff19SB force field will lead to results that are closest to those obtained with the polarizable force field. Even in this case, the electrostatic interactions of RNA with the protein and the other RNA strand are overestimated which leads to less freedom of movement to all biomolecules. The too strong charge-charge interactions are a known issue in non-polarizable force fields[33,35–37], for which there are no easy solutions.

Generally, only adjusting Lennard-Jones parameters, vdW parameters or changing the water model are proposed to help with the issue of too strong electrostatic interactions.[4,23,37–39] Electronic continuum correction approach can be used to adjust the charges of ions to enhance the ion representation in non-polarizable force fields.[37] This methodology could be partially also employed on proteins, where the charged side chains are relatively distant from the backbone. However, in the case of RNA, the ribose-phosphate backbone is heavily charged and any simple modifications to the point charges would compromise the description of the backbone dihedrals which are directly related to the point charges.

Another commonly used non-polarizable force field CHARMM36 was not included in this study, as results of RNA-protein complex simulations with it in comparison with the Amber force fields are presented in another very recent study.[33] In that study, CHARMM36 simulations showed lower hydrogen bond count than the Amber simulations and less interaction between the RNA and the protein. The studied system had the RNA bound on the protein surface which allowed freer movement for it than the Ago2 binding site which is in-between of the protein domains. Based on the CHARMM36 results and our results with OPLS4, it seems that the Amber force fields result in higher hydrogen bond count and less fluctuating complex. This could be caused by the Amber force fields mixing and matching the partial charges and other parameters when calculating the interaction of molecules from different force fields, i.e. ff19SB for protein and OL3 for RNA.

The polarizable AMOEBA force field overcomes too tight packing but is computationally expensive. The AMOEBA force field was the only studied force field here where the RNA-protein complex started opening outwards instead of compacting. This was obvious already in the initial very short 10 ns simulations, and the trend continued also in the longer AMOEBA simulations (2*100ns). However, using the polarizable force field comes with a significant computational cost. While the ff19SB + OL3 simulations achieved a simulation

speed of ~160 ns/day, the AMOEBA simulations only reached a simulation speed of ~4 ns/day. Both simulations were run in a similar environment, on a single NVIDIA Volta V100 GPU. Due to this significant increase in computational cost, in some cases it might be more beneficial to use non-polarizable force fields with an enhanced sampling method to get reasonably accurate results.

Conclusions

Even though extensive effort has been put to parameterize the force fields for RNA and proteins, the RNA-protein complex simulations remain problematic because they need to provide reasonable interactions at the interface of these two chemically different molecules. This study further confirms that the force field and other simulation parameters selection is always dependent on the studied system. Based on our results all the non-polarizable force fields can be used to simulate RNA-protein complexes. In this case, the combination of ff19SB and OL3 had fewer issues with additional artificial hydrogen bonds and less compact packing. However, all the non-polarizable force fields tended to make the complexes very compact which might prevent the formation of some biologically relevant conformations. To avoid this, and when the computational cost is not an issue, the polarizable force field AMOEBA should be preferred. We perceive that the polarizable force fields are the future of biomolecular simulations also beyond the RNA-protein complexes after sufficient development of software and hardware makes them a computationally reasonable alternative.

Data availability

The simulation trajectories in .xtc format and their corresponding .pdb files have been uploaded into the Zenodo database under the following DOI: 10.5281/zenodo.6605469. The trajectories are wrapped into a single periodic boundary box, centered around the protein Cα atoms and the water molecules have been stripped out to conserve disk space.

Funding

This work was supported by Erasmus+ program which funded the research visit of H.B. at the University of Eastern Finland, during which most of the simulations were ran and analyzed. In addition, P.B. is funded by the competitive funding to strengthen university research profiles, 5th call, funding to University of Eastern Finland, funded by the Academy of Finland [grant number 325022].

Acknowledgements

The authors thank CSC IT Center for Science for the computational resources and Professor Antti Poso for critical comments on the manuscript.

References

(1) Berro, J. “Essentially, All Models Are Wrong, but Some Are Useful”—a Cross-Disciplinary Agenda for Building Useful Models in Cell Biology and Biophysics. *Biophys Rev* **2018**, *10* (6), 1637–1647. https://doi.org/10.1007/s12551-018-0478-4.

(2) Bissaro, M.; Sturlese, M.; Moro, S. Exploring the RNA-Recognition Mechanism Using Supervised Molecular Dynamics (SuMD) Simulations: Toward a Rational Design for Ribonucleic-Targeting Molecules? *Front. Chem.* **2020**, *8*. https://doi.org/10.3389/fchem.2020.00107.

(3) Estarellas, C.; Otyepka, M.; Koča, J.; Banáš, P.; Krepl, M.; Šponer, J. Molecular Dynamic Simulations of Protein/RNA Complexes: CRISPR/Csy4 Endoribonuclease. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850* (5), 1072–1090. https://doi.org/10.1016/j.bbagen.2014.10.021.

(4) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. *Proceedings of the National Academy of Sciences* **2018**, *115* (7), E1346–E1355. https://doi.org/10.1073/pnas.1713027115.

(5) Krepl, M.; Dendooven, T.; Luisi, B. F.; Sponer, J. MD Simulations Reveal the Basis for Dynamic Assembly of Hfq–RNA Complexes. *Journal of Biological Chemistry* **2021**, *296*. https://doi.org/10.1016/j.jbc.2021.100656.

(6) Bochicchio, A.; Krepl, M.; Yang, F.; Varani, G.; Sponer, J.; Carloni, P. Molecular Basis for the Increased Affinity of an RNA Recognition Motif with Re-Engineered Specificity: A Molecular Dynamics and Enhanced Sampling Simulations Study. *PLOS Computational Biology* **2018**, *14* (12), e1006642. https://doi.org/10.1371/journal.pcbi.1006642.

(7) Chang, S.; Zhang, D.-W.; Xu, L.; Wan, H.; Hou, T.-J.; Kong, R. Exploring the Molecular Basis of RNA Recognition by the Dimeric RNA-Binding Protein via Molecular Simulation Methods. *RNA Biology* **2016**, *13* (11), 1133–1143. https://doi.org/10.1080/15476286.2016.1223007.
(8) Kong, R.; Xu, L.; Piao, L.; Zhang, D.; Hou, T.-J.; Chang, S. Exploring the RNA-Bound and RNA-Free Human Argonaute-2 by Molecular Dynamics Simulation Method. *Chem Biol Drug Des* **2017**, *90* (5), 753–763. https://doi.org/10.1111/cbdd.12997.
(9) Kalia, M.; Willkomm, S.; Claussen, J. C.; Restle, T.; Bonvin, A. M. J. J. Novel Insights into Guide RNA 5′-Nucleoside/Tide Binding by Human Argonaute 2. *Int J Mol Sci* **2015**, *17* (1). https://doi.org/10.3390/ijms17010022.
(10) Habibian, M.; Harikrishna, S.; Fakhoury, J.; Barton, M.; Ageely, E. A.; Cencic, R.; Fakih, H. H.; Katolik, A.; Takahashi, M.; Rossi, J.; Pelletier, J.; Gagnon, K. T.; Pradeepkumar, P. I.; Damha, M. J. Effect of 2′-5′/3′-5′ Phosphodiester Linkage Heterogeneity on RNA Interference. *Nucleic Acids Res* **2020**, *48* (9), 4643–4657. https://doi.org/10.1093/nar/gkaa222.
(11) Jiang, H.; Sheong, F. K.; Zhu, L.; Gao, X.; Bernauer, J.; Huang, X. Markov State Models Reveal a Two-Step Mechanism of MiRNA Loading into the Human Argonaute Protein: Selective Binding Followed by Structural Re-Arrangement. *PLOS Computational Biology* **2015**, *11* (7), e1004404. https://doi.org/10.1371/journal.pcbi.1004404.
(12) Bhandare, V.; Ramaswamy, A. Structural Dynamics of Human Argonaute2 and Its Interaction with SiRNAs Designed to Target Mutant Tdp43. *Adv Bioinformatics* **2016**, *2016*. https://doi.org/10.1155/2016/8792814.
(13) Kandeel, M.; Kitade, Y. Molecular Dynamics and Binding Selectivity of Nucleotides and Polynucleotide Substrates with EIF2C2/Ago2 PAZ Domain. *Int J Biol Macromol* **2018**, *107* (Pt B), 2566–2573. https://doi.org/10.1016/j.ijbiomac.2017.10.145.
(14) Harikrishna, S.; Pradeepkumar, P. I. Probing the Binding Interactions between Chemically Modified SiRNAs and Human Argonaute 2 Using Microsecond Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017**, *57* (4), 883–896. https://doi.org/10.1021/acs.jcim.6b00773.
(15) Liu, Y.; Yu, Z.; Zhu, J.; Wang, S.; Xu, D.; Han, W. Why Is a High Temperature Needed by Thermus Thermophilus Argonaute During MRNA Silencing: A Theoretical Study. *Front. Chem.* **2018**, *6*. https://doi.org/10.3389/fchem.2018.00223.
(16) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **2018**, *118* (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.
(17) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14* (4), 2084–2108. https://doi.org/10.1021/acs.jctc.7b01169.
(18) Wang, H.-W.; Noland, C.; Siridechadilok, B.; Taylor, D. W.; Ma, E.; Felderer, K.; Doudna, J. A.; Nogales, E. Structural Insights into RNA Processing by the Human RISC-Loading Complex. *Nat Struct Mol Biol* **2009**, *16* (11), 1148–1153. https://doi.org/10.1038/nsmb.1673.
(19) Schirle, N. T.; Sheu-Gruttadauria, J.; MacRae, I. J. Structural Basis for MicroRNA Targeting. *Science* **2014**, *346* (6209), 608–613. https://doi.org/10.1126/science.1258040.
(20) Kehl, T.; Backes, C.; Kern, F.; Fehlmann, T.; Ludwig, N.; Meese, E.; Lenhof, H.-P.; Keller, A. About MiRNAs, MiRNA Seeds, Target Genes and Target Pathways. *Oncotarget* **2017**, *8* (63), 107167–107175. https://doi.org/10.18632/oncotarget.22363.
(21) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.
(22) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(23) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.
(24) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17* (7), 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302.

(25) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and Relative Alchemical Free Energies Using AMOEBA on GPUs. *Journal of Computational Chemistry* **2017**, *38* (23), 2047–2055. https://doi.org/10.1002/jcc.24853.
(26) The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.5.1 Schrödinger, LLC.
(27) Schrödinger Release 2020-4: Glide, Schrödinger, LLC, New York, NY, 2020. Schrödinger Release 2021-3, Schrödinger, LLC, New York, NY, 2021.
(28) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537. https://doi.org/10.1021/ct100578z.
(29) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295. https://doi.org/10.1021/ct200133y.
(30) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Lucho, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, G.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. AMBER 2020. *University of California, San Francisco.* **2020**.
(31) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J Comput Aided Mol Des* **2017**, *31* (1), 147–161. https://doi.org/10.1007/s10822-016-9977-1.
(32) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659.
(33) Gallardo, A.; Bogart, B. M.; Dutagaci, B. Protein–Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**. https://doi.org/10.1021/acs.jcim.2c00121.
(34) Onufriev, A. V.; Izadi, S. Water Models for Biomolecular Simulations. *WIREs Computational Molecular Science* **2018**, *8* (2), e1347. https://doi.org/10.1002/wcms.1347.
(35) Yoo, J.; Aksimentiev, A. New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys Chem Chem Phys* **2018**, *20* (13), 8432–8449. https://doi.org/10.1039/C7CP08185E.
(36) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **2016**, *12* (1), 430–443. https://doi.org/10.1021/acs.jctc.5b00967.
(37) Duboué-Dijon, E.; Javanainen, M.; Delcroix, P.; Jungwirth, P.; Martinez-Seara, H. A Practical Guide to Biologically Relevant Molecular Simulations with Charge Scaling for Electronic Polarization. *J. Chem. Phys.* **2020**, *153* (5), 050901. https://doi.org/10.1063/5.0017775.
(38) Chen, A. A.; García, A. E. High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc Natl Acad Sci U S A* **2013**, *110* (42), 16820–16825. https://doi.org/10.1073/pnas.1309392110.
(39) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing Solute–Water van Der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116* (15), 4524–4534. https://doi.org/10.1021/jp2118373.

Figures and captions

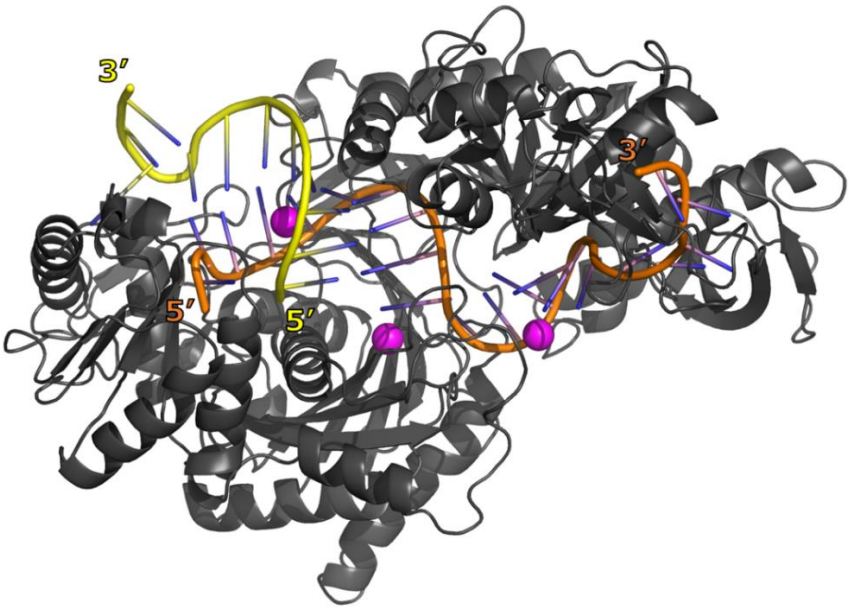

Figure 1. The model of Ago2 bound to guide (orange) and target (yellow) RNA, and three magnesium ions (magenta). The model was constructed from PDB ID 4W5O by predicting the position of missing protein residues and manually inserting the missing nucleotides.

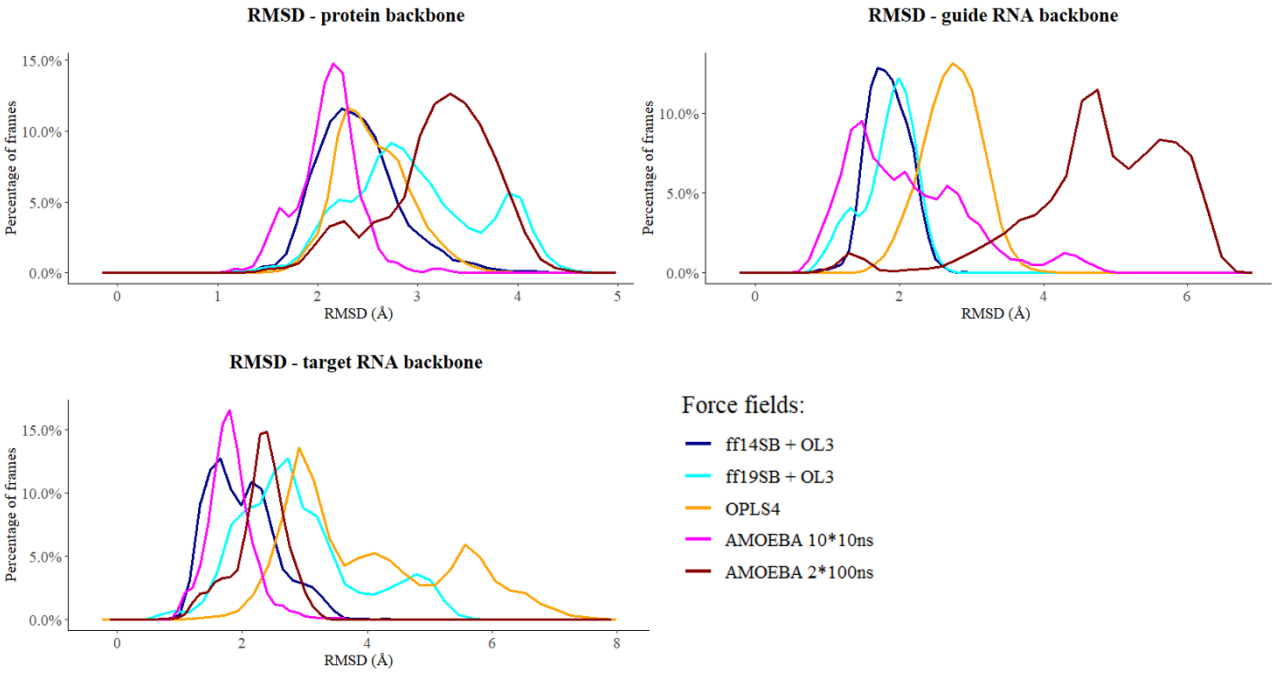

Figure 2. RMSD frequency polygons calculated for the backbones of the protein, the guide and the target RNA.

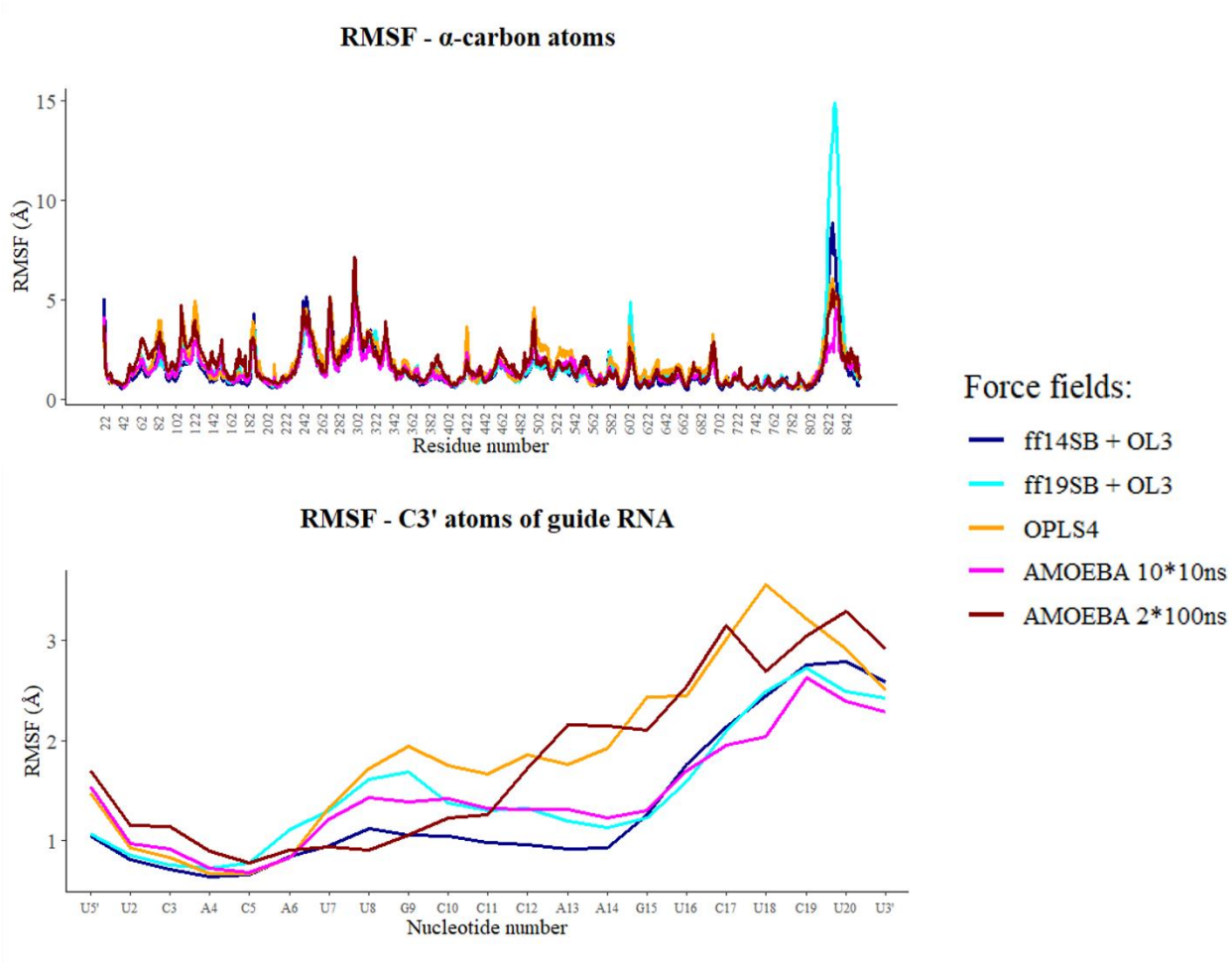

Figure 3. The relative fluctuations of the protein Cα and the guide RNA C3' atoms in the different force fields.

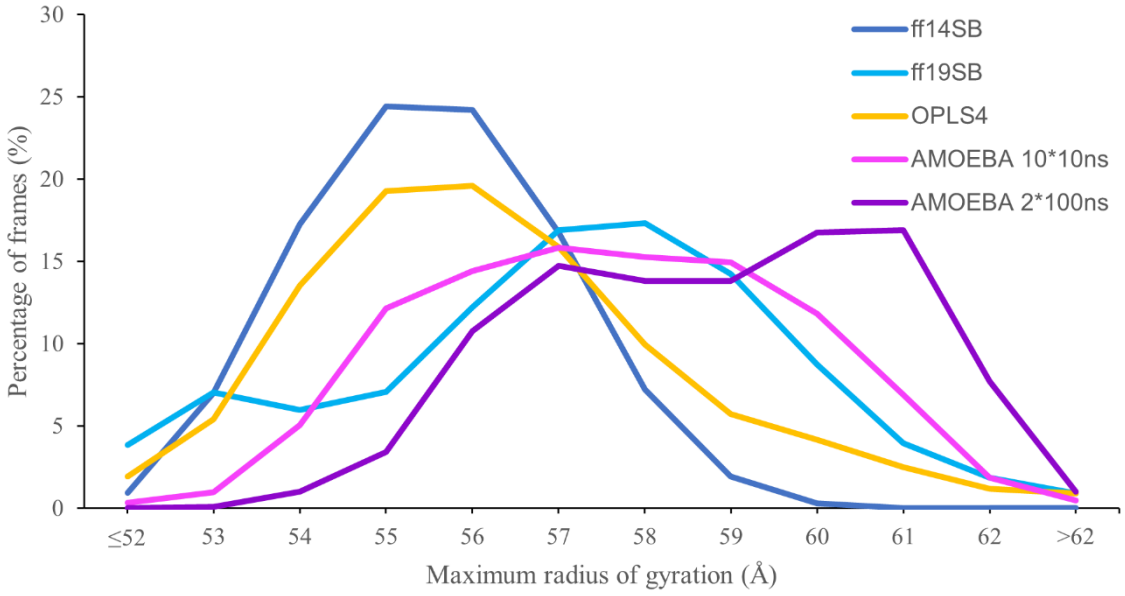

Figure 4. Frequency of the maximum radius of gyration of Cα atoms in the snapshots.

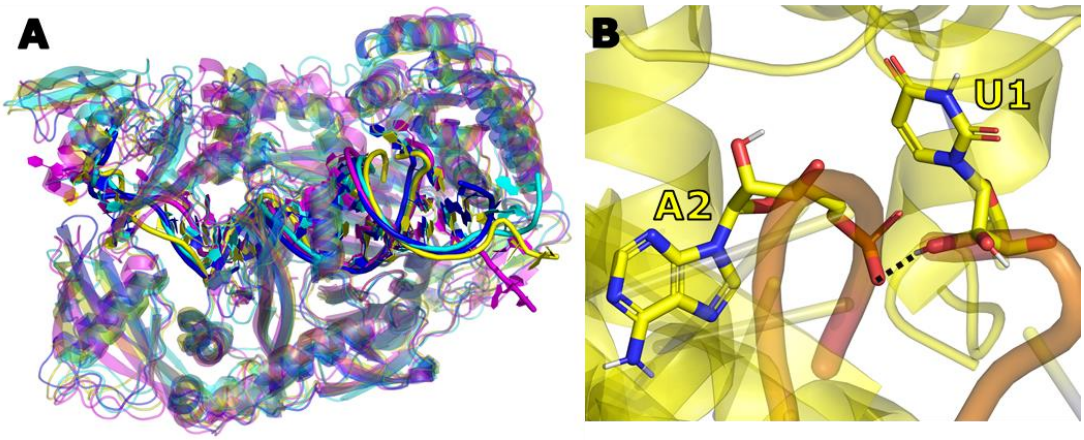

Figure 5. A) Snapshots of simulations from the ff14SB (blue), ff19SB (cyan), OPLS4 (yellow) and AMOEBA (magenta) force fields. The AMOEBA force field, despite its shorter simulation times, keeps the two RNA strands further away from each other than the nonpolarizable force fields. B) The hydrogen bond between the guide RNA U1 and the target RNA A2 which explains the tighter packing in most of the OPLS4 simulations.

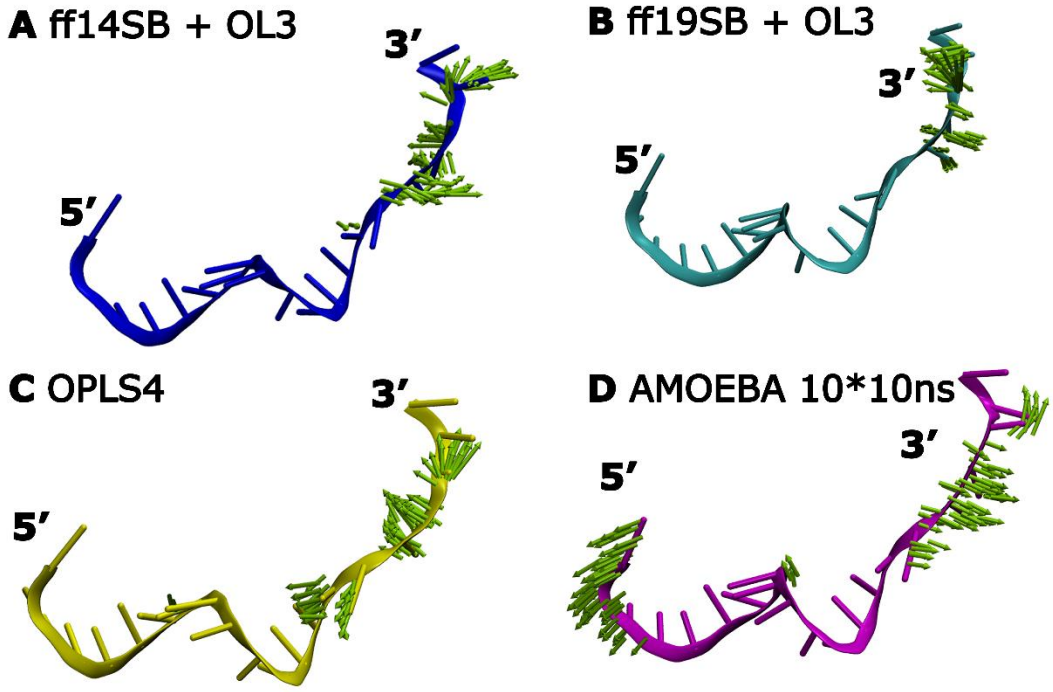

Figure 6. The first principal component from the PCA analysis: A) ff14SB + OL3, B) ff19SB + OL3, C) OPLS4 and D) AMOEBA 10*10ns. The arrows depict the direction and intensity of the movement. For clarity, only larger movements (>2.5 Å) are shown.

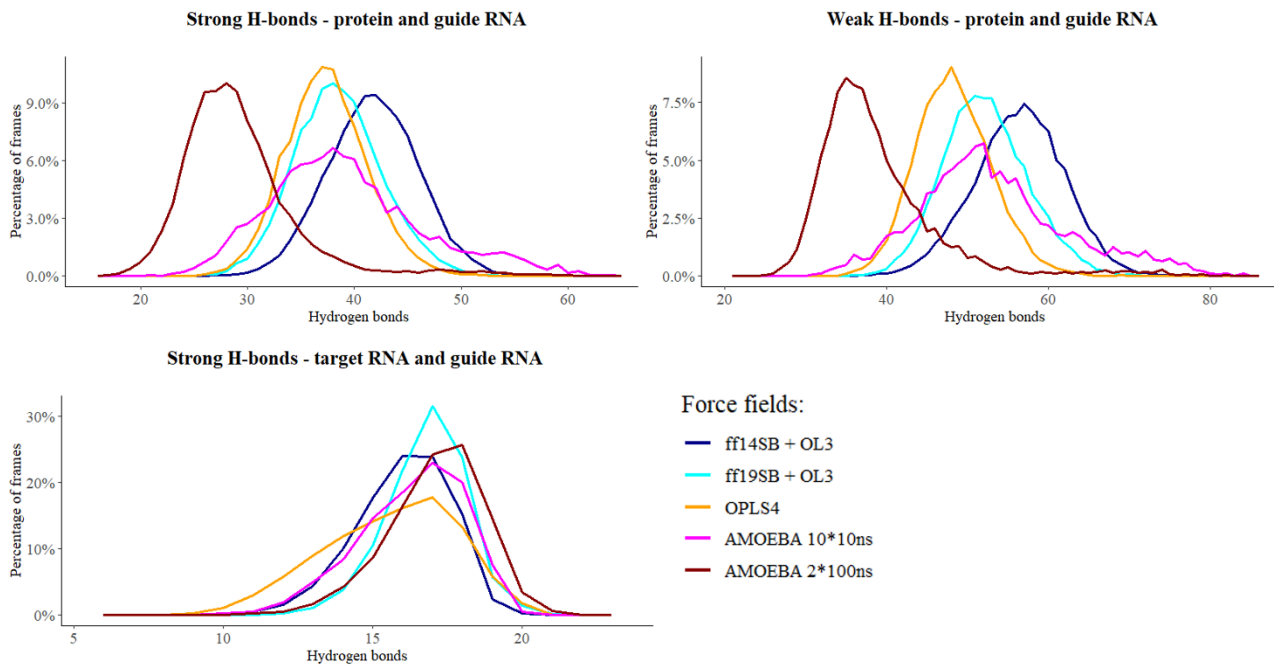

Figure 7. Frequency polygons of hydrogen bonds between the guide RNA and the target RNA and between the protein and the guide RNA.

Tables and captions

Table 1. Average RMSD values in different simulations with corresponding standard deviations (SD).

RMSD	ff14SB + OL3	ff19SB + OL3	OLPS4	AMOEBA 10*10ns	AMOEBA 2*100ns
Protein backbone	2.4 ± 0.4 Å	3.0 ± 0.7 Å	2.5 ± 0.4 Å	2.1 ± 0.3 Å	3.2 ± 0.6 Å
Guide RNA backbone	2.2 ± 0.4 Å	2.5 ± 0.5 Å	3.6 ± 0.8 Å	2.1 ± 0.9 Å	4.8 ± 1.0 Å
Target RNA backbone	2.0 ± 0.6 Å	2.9 ± 1.0 Å	4.0 ± 1.3 Å	1.8 ± 0.4 Å	2.3 ± 0.4 Å

Table 2. The average number of hydrogen bonds between the target RNA and the guide RNA and its standard deviation (SD) during the simulations. Strong bond = Donor-acceptor distance <3.2Å and angle >135°. Weak bond = Donor-acceptor distance <3.5Å and angle >120°.

Intra-RNA	ff14SB + OL3	ff19SB + OL3	OLPS4	AMOEBA 10*10ns	AMOEBA 2*100ns
Strong	16.0 ± 1.6	16.8 ± 1.4	15.6 ± 2.2	16.3 ± 1.8	17.1 ± 1.6
Guide RNA-Protein	**ff14SB + OL3**	**ff19SB + OL3**	**OLPS4**	**AMOEBA 10*10ns**	**AMOEBA 2*100ns**
Strong	41.7 ± 4.2	38.5 ± 4.1	37.4 ± 3.8	39.2 ± 7.2	29.2 ± 5.5
Weak	56.4 ± 5.6	52.2 ± 5.2	48.3 ± 4.7	53.3 ± 9.3	38.8 ± 7.5