Conformational Ensembles of Non-Coding Elements in the SARS-CoV-2 Genome from Molecular Dynamics Simulations

Sandro Bottaro[1,*], **Giovanni Bussi**[2], **Kresten Lindorff-Larsen**[1,*]

***For correspondence:**
sandro.bottaro@bio.ku.dk (SB);
lindorff@bio.ku.dk (KLL)

[1]Structural Biology and NMR Laboratory & Linderstrøm-Lang Centre for Protein Science Department of Biology, University of Copenhagen. Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark; [2]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136, Trieste, Italy

Abstract The 5′ untranslated region (UTR) of SARS-CoV-2 genome is a conserved, functional and structured genomic region consisting of several RNA stem-loop elements. While the secondary structure of such elements has been determined experimentally, their three-dimensional structure is not known yet. Here, we predict structure and dynamics of five RNA stem-loops in the 5′-UTR of SARS-CoV-2 by extensive atomistic molecular dynamics simulations, more than 0.5ms of aggregate simulation time, in combination with enhanced sampling techniques. We compare simulations with available experimental data, describe the resulting conformational ensembles, and identify the presence of specific structural rearrengements in apical and internal loops that may be functionally relevant. Our atomic-detailed structural predictions reveal a rich dynamics in these RNA molecules, could help the experimental characterisation of these systems, and provide putative three-dimensional models for structure-based drug design studies.

Introduction

The genome of the SARS-CoV-2 virus is a single, positive stranded RNA consisting of approximately 30 thousand nucleotides. The 5′ and 3′ untranslated regions (UTR) contain cis-acting sequences required for viral transcription and replication (***Lal, 2010***; ***Yang and Leibowitz, 2015***). Recent studies suggest the 5′-UTR to enhance viral transcription (***Schubert et al., 2020***), and undergoes liquid-liquid phase separation with the SARS-CoV-2 nucleocapsid protein (***Iserman et al., 2020***).

Chemical probing experiments, NMR chemical shift measurements, and computational predictions have shown that the 5′-UTR is highly structured (***Wacker et al., 2020***; ***Andrews et al., 2020***; ***Manfredonia et al., 2020***; ***Rangan et al., 2020b***), and consists of several stem-loop elements (Fig 1). The secondary structures of the isolated elements are in close agreement with the full-length construct (***Wacker et al., 2020***), suggesting that they can fold independently. Except for stem loop 2 (SL2) from SARS-CoV-1 (***Lee et al., 2011***), the three-dimensional structures have not been determined experimentally for any of the elements in the 5′-UTR. In a recent computational study, ***Rangan et al.*** (***2020a***) predicted the three-dimensional fold of all stem-loop elements in both 5′ and 3-UTR using Rosetta's FARFAR2 algorithm (***Watkins et al., 2020***). The predicted SL2 model in the 5′-UTR was later refined using MD simulations, (***Bergonzo and Szakal, 2020***) leading to an improved agreement with available NMR data (***Lee et al., 2011***).

In this work we have used atomistic molecular dynamics simulations to predict the structure and dynamics of five 5′-UTR stem-loops. We have generated three-dimensional configurations compatible with the secondary structure recently determined by chemical probing and NMR experiments (***Wacker et al., 2020***). While several fragment-based tools exist to perform such task (***Parisien and Major, 2008***; ***Cheng et al., 2015***), we here adapt an MD-based procedure (***Poblete et al., 2018***). These initial structures are subject to extensive MD sampling, amounting to more than 0.5ms of aggregate simulation time. More precisely, we enhance the sampling of apical and internal loop regions (***Wang et al., 2011***; ***Bussi, 2014***), where we currently lack detailed structural information and where significant conformational changes are more likely to occur. Our simulations are thus a Boltzmann-weighted ensemble of three-dimensional conformations constituting our prediction of structure and conformational heterogeneity of these elements.

First, we apply the computational protocol to stem loop 2 (SL2), and validate the accuracy of our prediction against available experimental NMR measurements on SL2 from SARS-CoV-1 (***Lee et al., 2011***). We then proceed by describing our predictions on systems of increasing complexity, stem loop 5a (SL5a), SL1, SL4, and SL2+SL3 (Fig.1), for which no detailed structural information is available.

Our predictions could help the atomic-detailed experimental characterisation of such systems, as simulations can be easily integrated a-posteriori with NMR or SAXS data (***Cesari et al., 2018***; ***Bottaro and Lindorff-Larsen, 2018***; ***Bottaro et al., 2020***). The predicted models provide a starting point for structure-based drug design studies (***Stelzer et al., 2011***; ***Costales et al., 2020***), and could serve as additional benchmark systems to evaluate the accuracy of atomistic RNA force-field (***Dans et al., 2019***).

Methods

MD simulations were conducted using the Amber ff99SB force field with parmbsc0 (***Pérez et al., 2007***), OL3 (***Zgarbová et al., 2011***), and corrections to the Van der Waals parameters for oxygen (***Steinbrecher et al., 2012***) in conjunction with OPC water model (***Izadi et al., 2014***). This combination has been shown to improve the description of RNA structures compared to other force-fields (***Bergonzo and Cheatham III, 2015***), and its accuracy has been assessed by independent groups (***Bottaro et al., 2018***; ***Kuhrova et al., 2019***). The initial conformations were minimised in vacuum and then solvated in a dodecahedral box with minimum distance from box edge of 1.2nm. Potassium ions were added to neutralise the box (***Joung and Cheatham III, 2008***). The systems were equilibrated in the NVT and NPT ensemble at 1 bar for 1 ns. Production runs were performed in the canonical ensemble using a stochastic velocity rescaling thermostat (***Bussi et al., 2007***). All bonds were constrained using the LINCS (***Hess et al., 1997***) algorithm and equations of motion were integrated with a time step of 2 fs. Simulations were performed using GROMACS 2019.4 (***Pronk et al., 2013***) patched with PLUMED 2.5.3 (***Tribello et al., 2014***). Force-field parameters are available at http://github.com/srnas/ff/tree/opc-water-model/amber_na.ff.

Pulling simulations

The aim of the pulling simulations is quickly to generate three-dimensional RNA conformations with a specific input secondary structure. We adapt a procedure originally introduced for reconstructing atomistic structures from coarse-grained models (***Poblete et al., 2018***). Our protocol consists of three stages, as illustrated in Fig. 2a-b for SL2:

1. Plain MD simulation initialised from extended, single stranded, A-form RNA structures.
2. Pulling. We construct a reference structure with an ideal A-form double stranded RNA template for regions with known secondary structure (shown in blue in Fig. 2a) and perform pulling simulations to minimise the eRMSD (***Bottaro et al., 2014***) distance between the RNA molecule and the reference. We here use adiabatic bias MD simulations (***Marchi and Ballone, 1999***), that introduce a biasing potential damping the fluctuations only when the system

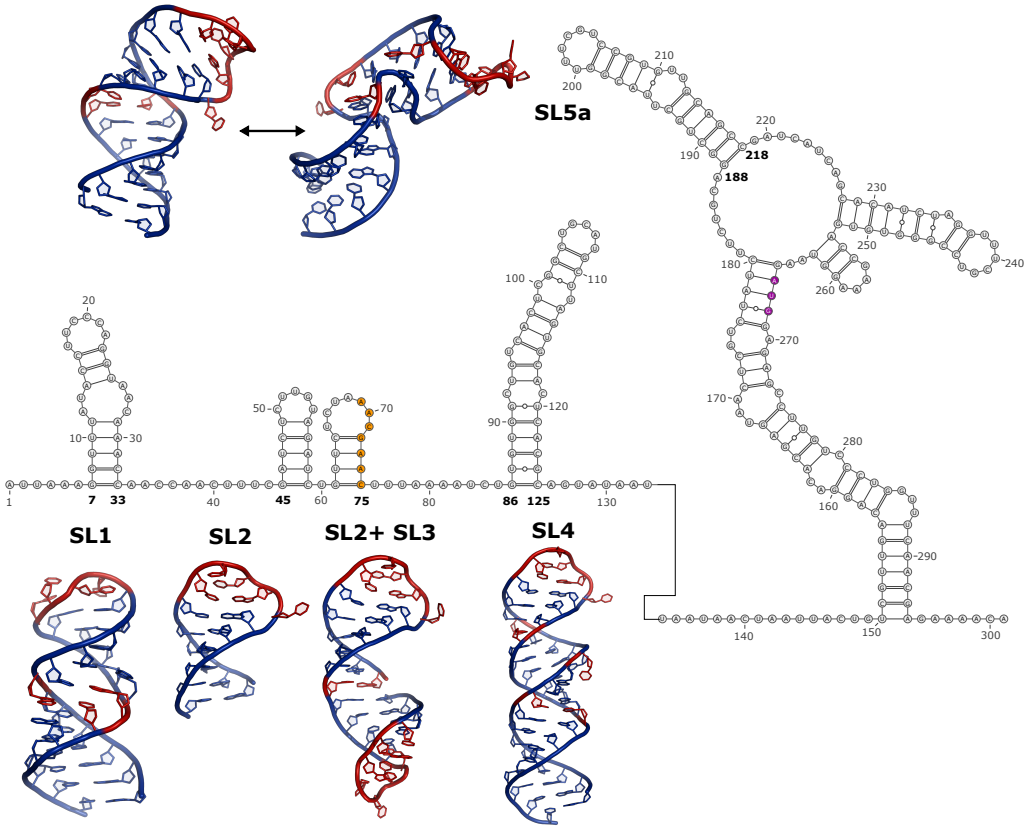

Figure 1. Reference sequence and secondary structure of the 5′-UTR in SARS-CoV-2 genome. The numbering follows the convention in ***Wacker et al.*** (***2020***). The five structured elements considered in this work are labeled: SL1, SL2, SL2+SL3, SL4, and SL5a. The conserved sequence of the transcription regulatory sequence and the AUG start codon are highlighted in orange and purple, respectively. The centroids of the most populated clusters from MD simulations corresponding to each element are shown.

moves further away from the template. In this way the pulling procedure follows the thermal fluctuations of the system, and does not require a pre-defined scheduling for the moving restraint (***Poblete et al., 2018***). When more than one stem is present, the bias acts on multiple templates simultaneously. The constant of the harmonic damping bias is set to 2 kJ/mol.

3. Plain MD. After 40ns of pulling simulation, the external bias is removed and the system freely fluctuates without external forces.

At the end of the procedure we extract samples from the final plain MD stage satisfying the secondary structure constraints, i.e. with eRMSD from template < 0.75. This criterion guarantees that all base-pairs corresponding to those in the target structure are correctly formed (Fig. 2b, bottom panel). Note that individual pulling simulations may fail: stems are sometimes not formed correctly and unfold during the final plain MD stage (Supporting Information 1). The success rate of this procedure ranges from $\approx 80\%$ for SL2 to $\approx 10\%$ for SL4, which consists of 4 stem structures, consistently with previously reported benchmarks (***Poblete et al., 2018***). During preliminary tests we have observed that the formation of the correct secondary structure is hampered by early-forming non-native contacts. For this reason we conduct pulling simulations at 340K.

Partial tempering

We resolvate and minimise the RNA structures obtained from pulling as described above. We use partial tempering to enhance sampling in regions with unknown secondary structure (***Bussi, 2014***).

System	Lenght (nt)	Simulation time [μs][(a)]	Acceptance rate [%] [(b)]	Round trips [(c)]
SL1	29	3.6	15-41	44-191
SL2	17	2.6	26-51	220-435
SL2+SL3	32	1.2	4-23	0-7
SL4	44	1.6	7-30	0-20
SL5a	33	3.6	12-34	6-33

Table 1. Simulated systems and related statistics. (a) Per-replica simulation time. Each simulation consists of 20 replicas and two independent runs for each system are carried out. The first 100ns of each run are discarded during analysis. (b) Minimum and maximum acceptance rate across all replicas and runs (c) Minimum and maximum number of round trips in temperature space across all replicas and runs.

One reference replica exchanges via Metropolis-Hastings algorithm with parallel simulations that are conducted with a modified Hamiltonian. The modification is such that interactions in the cold region (blue and grey in Fig. 2a, together with ions and water molecules) are kept at the reference temperature T=300K. Interactions in the hot region (shown in red in Fig. 2a) are rescaled by a factor $1/\lambda$, while the interactions between cold and hot regions are at an intermediate effective temperature of $T/\sqrt{\lambda}$, with $1 \leq \lambda < 0$. In this study we use 20 geometrically distributed replicas in the effective temperature range of 300 to 1000 K, and attempting exchanges every 240 steps. Hot regions are shown in red in Figs. 2-6, and include the phosphate group at the 3'-end of the selected region. We have found this small addition to be beneficial in preliminary test runs. System length, simulation time per replica, average acceptance rate, and number of round trips in temperature space are reported in Table 1, and additional statistics are available in SI2. To help the identification of slow degrees of freedom and assess convergence we conduct two independent runs with different initial conformations obtained from two separate pulling simulations (SI3).

Analysis and visualisation

Ideal A-form structures are constructed using the nucleic acids builder (NAB) software in Ambertools (***Macke and Case, 1998***). Trajectories are analysed using PLUMED (***Tribello et al., 2014***) and MDtraj (***McGibbon et al., 2015***). Base-pair annotation and puckering angles are calculated with Barnaba (***Bottaro et al., 2019***).

Results

In this section we describe the conformational ensembles obtained from MD simulations. We first show the results obtained on SL2, the simplest system considered in this study. Crucially, in this case it is possible to validate the procedure by comparing with available structural data and NMR measurements.

Stem Loop 2

We run a plain MD simulation of a single stranded, extended RNA with sequence `GGAUCUCUUGUAGAUCU`. After 10ns, we apply a biasing force to promote the formation of an ideal A-form RNA structure in the region where the secondary structure is known (blue nucleotides in Fig. 2a). By construction, the eRMSD distance (***Bottaro et al., 2014***) from the template decreases during pulling (Fig. 2b, top panel). After 40 ns the external bias is removed, and the system fluctuates around a stable structure in which all five Watson-Crick (WC) base-pairs are formed. (Fig. 2b, bottom panel).

While pulling simulations make it possible to quickly generate conformations with a desired stem structure, these runs are not sufficiently long to allow rearrangements in the loop region. Previous simulation studies have shown these motions to occur on long timescales (μs to ms) that require extensive simulations (***Šponer et al., 2018***). For SL2, we thus enhance the sampling in the loop region (red in Fig. 2a) by running partial tempering simulations (20 replicas, each 2.6 μs-long) initialised from a successful pulling run (see Methods).

Analyses of the enhanced simulations show that all five base-pairs in the stem are stable, and a comparison of two independent simulations show that our simulations are well converged (Fig. 2c). In addition, we observe the formation of a stable terminal GU wobble base-pair, as well as of a C50-G53 Watson-Crick interaction in the loop. A three-dimensional cluster analysis (***Bottaro et al., 2019***) reveals that the largest cluster ($\approx 90\%$) is remarkably similar to the available NMR loop structure from SARS-CoV-1 (***Lee et al., 2011***), whose sequence differs from our construct in the two terminal base-pairs. For example, the centroid of the most populated cluster from MD simulations superimposes extremely well to the first NMR model in PDB 2L6I (Fig. 2d). C50 forms a canonical base-pair with G53, U52 stacks on top of C50, and both U51/U54 are solvent exposed. The median distance between the experimental structure and the MD simulation is 1.6 Å heavy-atom root mean squared deviation and 0.5 eRMSD (SI4).

The sugar puckers of U51, U52, and U54 are preferably in C2'-endo conformation, in full agreement with NMR measurements (***Lee et al., 2011***) (Fig. 2e). The great majority (169 out of 179) of upper-bound NOE-derived distances are satisfied in MD simulations (Fig. 2f). The most significant discrepancies consist of short proton-proton distances between G53-U54 that are not sampled in our MD runs (SI5).

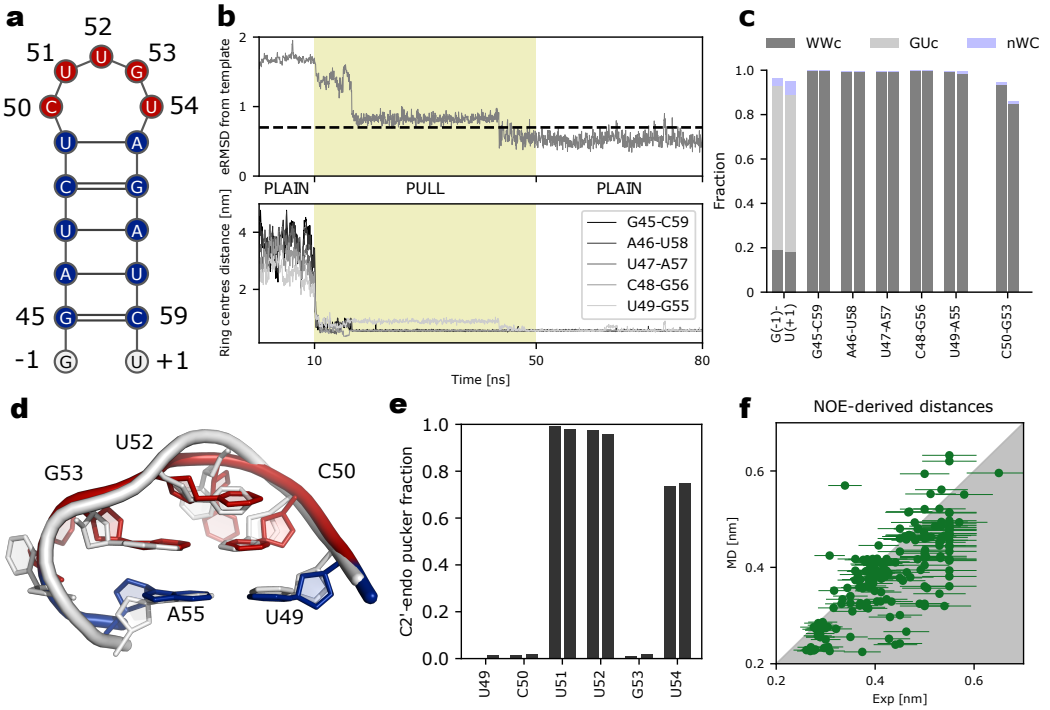

Figure 2. Description of the computational procedure and validation on SL2. **a** Sequence and secondary structure of SL2. The genomic numbering follows the convention in ***Wacker et al.*** (***2020***). The region used for constructing the template A-form three-dimensional structure is shown in blue. The sampling in the region shown in red is enhanced by partial tempering. **b** Behaviour of a successful pulling simulation. eRMSD (top panel) and distance between ring centres of the five WC base-pairs (bottom panel) during the three stages of the procedure: initial plain MD, pulling stage, and final plain MD stage, as labelled. The eRMSD threshold of 0.75 is shown as a dashed horizontal line. **c** Population of base-pairs from 20 replicas $\times$ 2.6 μs partial tempering simulations. Following the Leontis-Westhof classification (***Leontis and Westhof, 2001***), base-pairs are annotated as cis Watson-Crick/Watson-Crick (WWc), GU wobble cis (GUc) or non-Watson-Crick (nWC). For each interaction, the two bars show the statistics from independent simulations. **d** Centroid of the most-populated cluster from MD simulations (blue/red) superposed onto the SL2 loop from SARS-CoV-1 structure (white) (***Lee et al., 2011***). **e** Sugar pucker population in the loop region. **f** Calculated and experimental NOE-derived upper bound distances. The calculated 10 of the 179 NOEs that are larger than the experimental values are located in the white upper-triangular region.

Stem loop 5a

Stem loop 5a (SL5a) consists of two helices (blue nucleotides in Fig. 3a) connected by an internal loop and capped by an hexaloop with sequence UUUCGU. We use the same procedure described for SL2 to generate initial configurations and to sample the ensemble of SL5a by enhancing regions where secondary structure is not known (red regions in Fig. 3a). Helix 1 is highly stable throughout the 20 replicas $\times$ 3.6 μs partial tempering simulations (Fig. 3b). In the internal loop we observe the formation of a U194-U212 base-pair with a population of $\approx 60\%$ (Fig. 3b). The wobble GU base-pair between U195 and G210 in helix 2 is in equilibrium with an alternative WC/WC base-pair between U195 and U211. The most stable conformation for the apical loop is characterised by a wobble interaction between U200 and G204, with U205 completely solvent exposed. In one of the two simulations we observe the transient formation of a U200-U205 base-pair capped by a canonical UUCG tetraloop fold (***Cheong et al., 1990***), as suggested by NMR measurements (***Wacker et al., 2020***). The low population of this state is not surprising, as this fold has been shown not to be stabilised to a sufficient degree within this force-field (***Mráziková et al., 2020***).

In addition to the analysis of individual base-pairs, we identify in simulations frequent transitions from straight to kinked conformations (Figs. 3c-d). The latter structure is reminiscent of a kink-turn motif, but lacks its signature interactions and sequence propensities (***Klein et al., 2001***). This large-scale motion does not affect the stem structures, but only their relative orientation. We also note that kinked conformations are observed only when U212 is solvent exposed, and therefore not engaging base-pairs with U194 nor U195 (Fig. 3e), and suggest that the relative populations could be modulated by binding of proteins or small molecules.

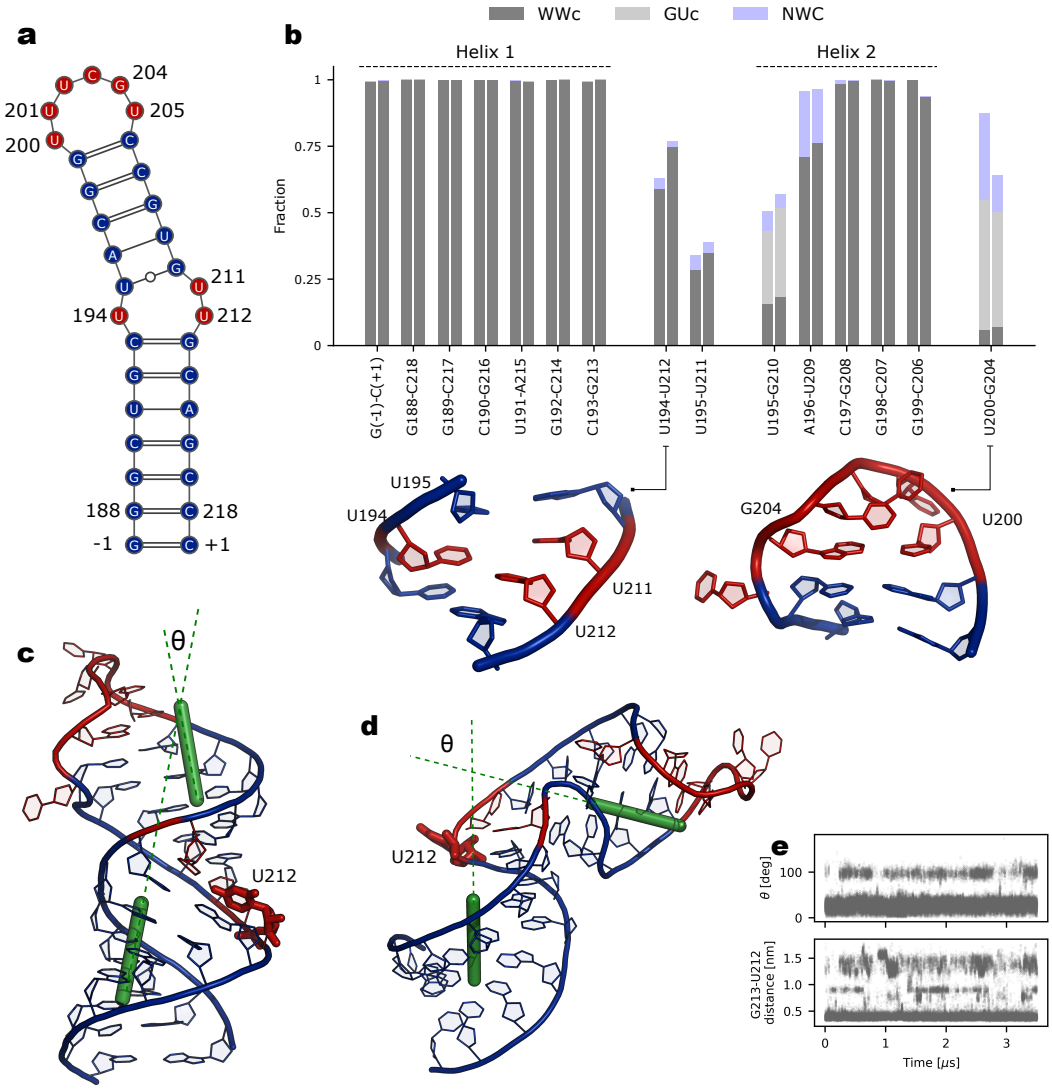

Figure 3. SL5a results. **a** Sequence and secondary structure of SL5a. The genomic numbering follows the convention in ***Wacker et al.*** (***2020***). The region used for constructing the A-form templates is shown in blue. The sampling in the region shown in red is enhanced by partial tempering. **b** Population of base-pairs from 20 replicas × 3.6 μs partial tempering simulations. Centroids of the most populated internal and apical loop structures discussed in the text are shown. **c** Representative straight conformation. The axes defined by helix 1 and 2, together with the angle θ between them are shown in green. The helical axes are calculated as the vector normal to the base-pair plane, averaged over all base-pairs within the helix. **d** Representative kinked conformation. **e** Time series of the angle between helices, θ, (top panel) and of the distance between U212 and G213 (bottom panel) in the neutral replica. We note that due to the replica-exchange setup used in our simulations, these simulations do not reflect the timescales needed to switch between the two conformations.

Stem loop 1

Similarly to SL5a, stem loop 1 (SL1) consists of two helices connected by an asymmetrical internal loop. We use the pulling procedure to initialise our runs from the previously determined secondary structure (***Wacker et al., 2020***) where U13-A26 are base-paired and thus represent the first WC interaction of helix 2. During partial tempering simulations we observe a rich dynamics involving the bulge region, together with its neighbourhood, as well as the apical loop.

While helix 1 is stable, the terminal base-pair U11-A29 either interacts through the Watson-Crick/Watson-Crick or Hoogsteen-sugar edges (Fig. 3b). The presence of this non-canonical interaction allows the formation of a A12-C28 sugar-Hoogsteen interaction as well as of a U13-A27 base-pair. A26 is not completely solvent-exposed, but forms a base triple together with A14 and U25. In our simulations we additionally observe two alternative, lowly-populated secondary structure arrangements (Fig. 4c-d) where A12-A27/U13-A26 are base-paired, and either C28 or A29 are partially excluded from the stack. Qualitatively, the nucleobase arrangements shown in Fig. 3b-d all preserve the coaxial stacking of the two helices. Taken together, our predictions are compatible with the NMR-derived secondary structure that does not show the presence of the U13-A26 WC base-pair. Our models suggest the central role of A27 in mediating alternative bulge conformations that might be required for viral viability (***Li et al., 2008***).

The apical loop is highly dynamic and lacks a major, stable conformation with specific base-base interactions. U17-A22 base-pair interconverts from Watson-Crick/Watson-Crick to sugar/Hoogsteen, and U18 can engage a sugar-Hogsteen interaction with C21, albeit with a small population (SI6).

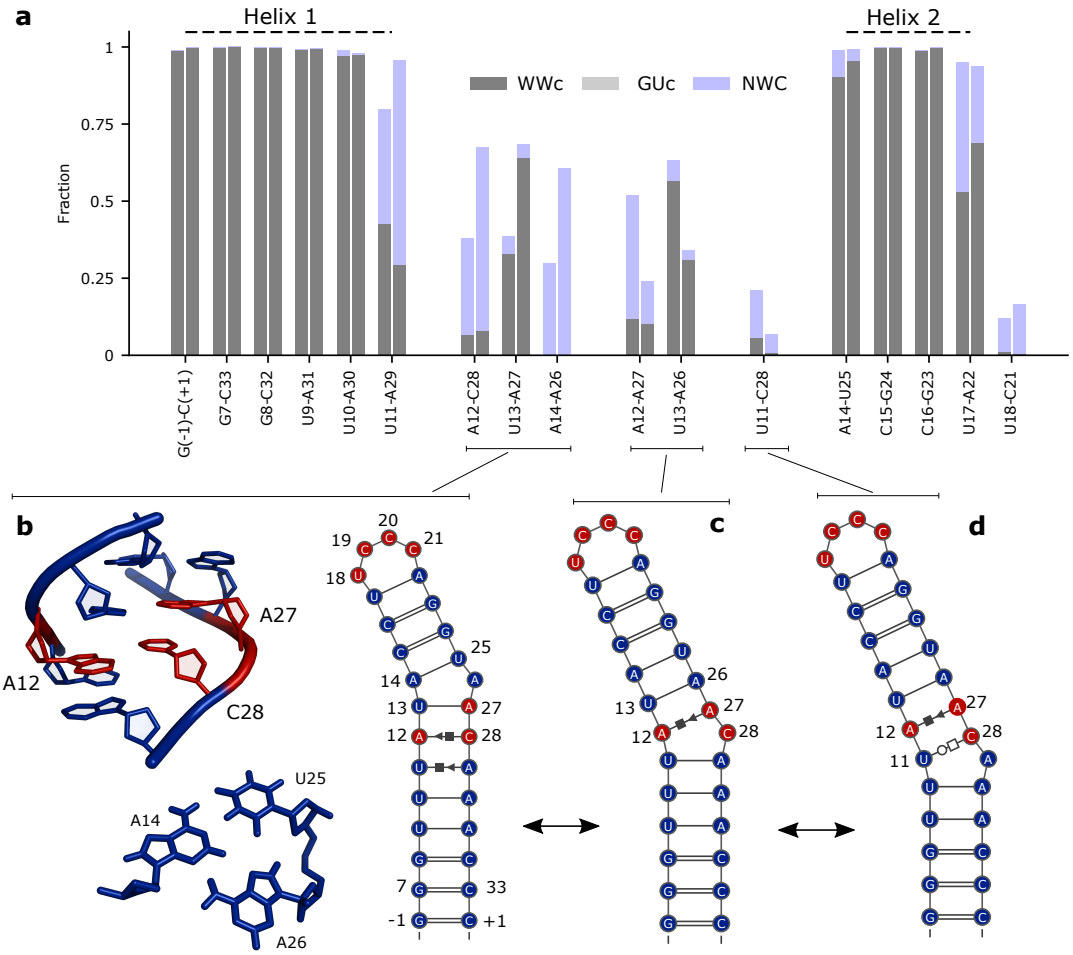

Figure 4. SL1 results. **a** Population of base-pairs from partial tempering simulations. **b** Three-dimensional structure of the most populated internal loop configuration showing the U11-A29 Hogsteen-Sugar, A12-C28 Sugar-Hoogsteen base-pairs, as well as the base A14-U25-U26 base triple. The corresponding secondary structure is shown on the right. The genomic numbering follows the convention in ***Wacker et al.*** (***2020***). The base-paired regions selected for creating initial configurations is shown in blue, while partial tempering acts in regions shown in red. **c-d** Secondary structures of two alternative internal loop arrangements.

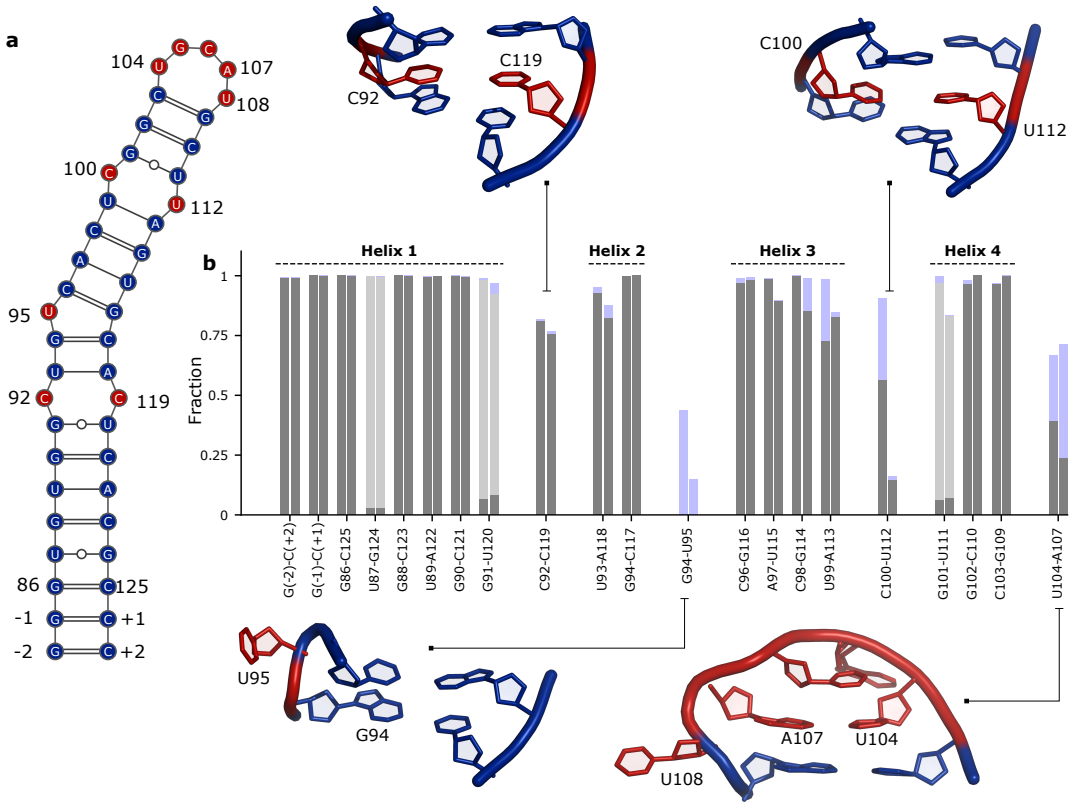

Figure 5. SL4 results. **a** Sequence and secondary structure of SL4. The genomic numbering follows the convention in ***Wacker et al.*** (***2020***). The region used for constructing the A-form templates is shown in blue. The sampling in the region shown in red is enhanced by partial tempering. **b** Population of base-pairs from 20 replicas × 1.6 μs partial tempering simulations. Centroid of the most populated internal and apical loops discussed in the text are shown.

Stem loop 4

Stem loop 4 (SL4) consists of four A-form helices of different lengths linked by short pyrimidine internal loops and capped by a `UGCAU` pentaloop (Fig. 5a). The experimentally-derived secondary structure is to a great extent preserved in parallel tempering simulations (Fig. 5b). The A-form geometry between helix 1 and helix 2 is stabilised by a WC/WC C92-C119 base-pair (Fig. 5b). U95 is solvent-exposed and transiently engages in non-canonical interactions with G94 (Fig. 5b).

In the third internal loop the two simulations behave differently suggesting slow motions that are not fully sampled in our simulations. In the first run C100 preferentially base-pair with U112, while in the second run additional configurations are sampled, where either C100 or U112 are not in the stack. The partial overlap between the two runs, that we ascribe to insufficient sampling, is further confirmed by a principal component analysis of this internal loop region (SI7). Experimental NMR measurements on SL4 agree with A-form helix compatible arrangement of the opposing residues C100 and U112 (***Wacker et al., 2020***), thus supporting the presence of pairing between the two bases. The apical loop structure resembles a tetraloop with U108 solvent-exposed and U104:A107 forming either a Watson-Crick/Watson-Crick or a sugar/Hogsteen non-canonical interaction, as shown in Fig. 5b.

SL2+SL3

We here report the results on the SL2+SL3 construct, consisting of two consecutive stem-loop structures. In our simulations we sample structures where the two helices are stacked coaxially tail-to-tail, thereby forming a unique A-form-like stem (Fig. 6).

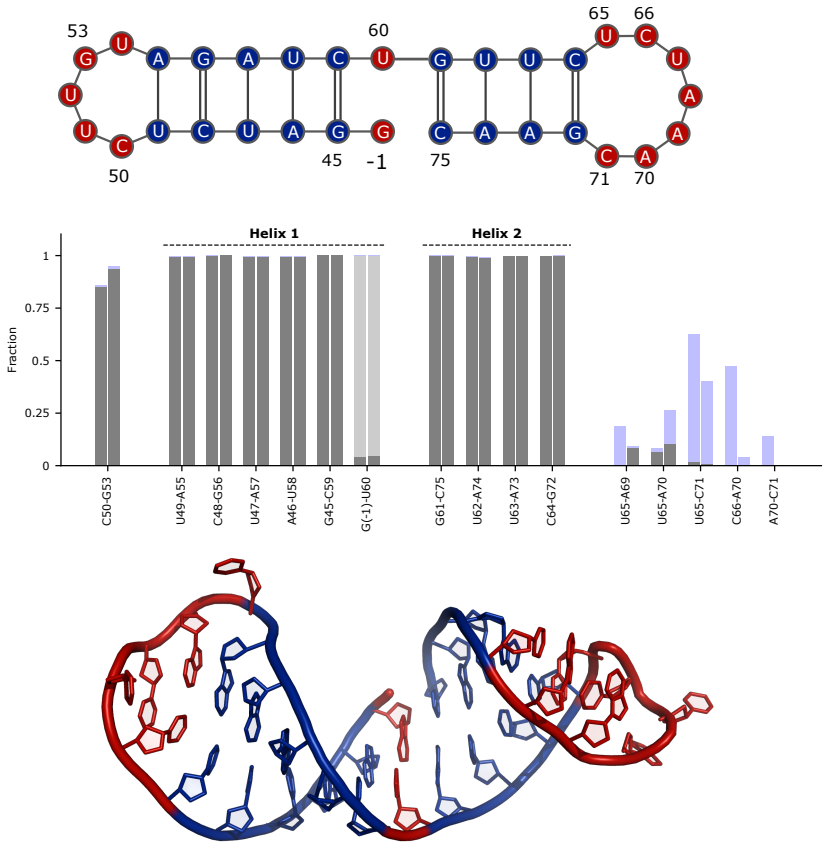

Figure 6. SL2+SL3 results. Top: Sequence and secondary structure of SL2+SL3. The genomic numbering follows the convention in ***Wacker et al.*** (***2020***). The region used for constructing the A-form templates is shown in blue. The sampling in the region shown in red is enhanced by partial tempering. Middle panel: Population of base-pairs from partial tempering simulations. Bottom panel: Centroid of the most populated cluster.

The SL2 loop conformations are similar to those observed for SL2 alone, with a stable C50-G53 base-pair (see above). The `UCUAAAC` heptaloop in SL3 is highly flexible, and adopts several conformations that are partially structured in terms of base-pairing (Fig. 6). Helix 2, containing the core sequence `ACGAAC` is highly stable in simulations, at variance with NMR measurements (***Wacker et al., 2020***). Since SL3 is expected to be only partially stable (***Yang and Leibowitz, 2015***), it is possible that the force-field overstabilise A-form structures, making it difficult to sample alternative conformations.

Discussion

We report the prediction of the structure and conformational heterogeneity of five stem-loop elements, ranging in size between 17 and 44 nucleotides, located at the 5'-UTR in SARS-CoV-2 genome. As a starting point for our study we use the secondary structure determined by NMR experiments (***Wacker et al., 2020***).

Structure of several of these elements have also been predicted using FARFAR2 (***Rangan et al., 2020a***). A principal component analysis on SL1 and SL2 reveals little overlap between the FARFAR2 and MD ensembles, suggesting the two methods to be complementary (SI8). In the case of SL2, the MD predictions overall reproduce the available NMR data, albeit small discrepancies remain. While in this case it appears that the ensemble generated by MD provides a more accurate description of compared to FARFAR2 (SI4), we highlight that our simulations required significant computational resources, amounting to more than 0.5ms of aggregate simulation time.

As in most MD studies, the results presented in this work depend on the amount of sampling. We took care in identifying and discussing states that are visited multiple times during one simulation and in both replicates, indicating these conformations to correspond to *bona fide*, highly populated metastable states. Also, the overall high agreement between the two independent simulations suggest that the conformational ensembles are relatively well converged.

Despite recent improvement to RNA force fields (***Šponer et al., 2018***; ***Dans et al., 2019***) simulation results also depend on the accuracy of the physical models used. While we have validated our simulations against available data, we note also that it should be straightforward to integrate our extensive conformational sampling with data from future solution measurements (e.g. NMR or SAXS) using reweighting approaches (***Cesari et al., 2018***; ***Bottaro and Lindorff-Larsen, 2018***; ***Bottaro et al., 2020***) to refine the atomic-level description of SARS-CoV-2 RNA structure and dynamics.

Acknowledgements

We acknowledge Drs. Harald Schwalbe, Anna Wacker, Julia E. Weigand and other members of the Covid19-NMR consortium for discussions and insights. SB and KLL acknowledge funding from the Lundbeck Foundation BRAINSTRUC structural biology initiative (R155-2015-2666). We acknowledge access to computational resources from PRACE for the COVID-RNA project (COVID19-72).

Data Availability

Analysis scripts, cluster centroids, and input files are available at https://github.com/KULL-Centre/papers/tree/master/2020/COVID_5UTR_MD PLUMED input files are available in the PLUMED NEST (***Bonomi et al., 2019***) under the accession code 20.034 https://www.plumed-nest.org/eggs/20/034/. MD trajectories can be found on Zenodo http://doi.org/10.5281/zenodo.4317075.

References

Andrews RJ, Peterson JM, Haniff HS, Chen J, Williams C, Grefe M, Disney MD, Moss WN. An in silico map of the SARS-CoV-2 RNA Structurome. BioRxiv. 2020; .

Bergonzo C, Cheatham III TE. Improved force field parameters lead to a better description of RNA structure. J Chem Theor Comput. 2015; 11(9):3969–3972.

Bergonzo C, Szakal AL. Using All-Atom Potentials to Refine RNA Structure Predictions of SARS-CoV-2 Stem Loops. International journal of molecular sciences. 2020; 21(17):6188.

Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, Bernetti M, Bolhuis PG, Bottaro S, Branduardi D, et al. Promoting transparency and reproducibility in enhanced molecular simulations. Nature methods. 2019; 16(8):670–673.

Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. In: *Structural Bioinformatics* Springer; 2020.p. 219–240.

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):eaar8521.

Bottaro S, Bussi G, Pinamonti G, Reißer S, Boomsma W, Lindorff-Larsen K. Barnaba: software for analysis of nucleic acid structures and trajectories. Rna. 2019; 25(2):219–231.

Bottaro S, Di Palma F, Bussi G. The role of nucleobase interactions in RNA structure and dynamics. Nucleic Acids Res. 2014; 42(21):13306–13314.

Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: A perfect match? Science. 2018; 361(6400):355–360.

Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. Molecular Physics. 2014; 112(3-4):379–384.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007; 126(1):014101.

Cesari A, Reißer S, Bussi G. Using the maximum entropy principle to combine simulations and solution experiments. Computation. 2018; 6(1):15.

Cheng CY, Chou FC, Das R. Modeling complex RNA tertiary folds with Rosetta. In: *Methods in enzymology*, vol. 553 Elsevier; 2015.p. 35–64.

Cheong C, Varani G, Tinoco I. Solution structure of an unusually stable RNA hairpin, 5GGAC (UUCG) GUCC. Nature. 1990; 346(6285):680–682.

Costales MG, Childs-Disney JL, Haniff HS, Disney MD. How we think about targeting RNA with small molecules. Journal of Medicinal Chemistry. 2020; .

Dans PD, Gallego D, Balaceanu A, Darré L, Gómez H, Orozco M. Modeling, simulations, and bioinformatics at the service of RNA structure. Chem. 2019; 5(1):51–73.

Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. J Comput Chem. 1997; 18(12):1463–1472.

Iserman C, Roden CA, Boerneke MA, Sealfon RS, McLaughlin GA, Jungreis I, Fritch EJ, Hou YJ, Ekena J, Weidmann CA, et al. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. Molecular Cell. 2020; .

Izadi S, Anandakrishnan R, Onufriev AV. Building water models: a different approach. The journal of physical chemistry letters. 2014; 5(21):3863–3871.

Joung IS, Cheatham III TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B. 2008; 112(30):9020–9041.

Klein D, Schmeing T, Moore P, Steitz T. The kink-turn: a new RNA secondary structure motif. The EMBO journal. 2001; 20(15):4214–4221.

Kuhrova P, Mlynsky V, Zgarbová M, Krepl M, Bussi G, Best RB, Otyepka M, Sponer J, Banas P. Improving the performance of the amber RNA force field by tuning the hydrogen-bonding interactions. Journal of chemical theory and computation. 2019; 15(5):3288–3305.

Lal SK. Molecular Biology of the SARS-Coronavirus. Springer Science & Business Media; 2010.

Lee CW, Li L, Giedroc DP. The solution structure of coronaviral stem-loop 2 (SL2) reveals a canonical CUYG tetraloop fold. FEBS letters. 2011; 585(7):1049–1053.

Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. Rna. 2001; 7(4):499–512.

Li L, Kang H, Liu P, Makkinje N, Williamson ST, Leibowitz JL, Giedroc DP. Structural lability in stem–loop 1 drives a 5 UTR–3 UTR interaction in coronavirus replication. Journal of molecular biology. 2008; 377(3):790–803.

Macke TJ, Case DA. Modeling unusual nucleic acid structures. In: *Molecular Modeling of Nucleic Acids* ACS Publications; 1998.p. 379–393.

Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, Wirecki TK, Marinus T, Ogando NS, Snijder EJ, van Hemert MJ, Bujnicki JM, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. Nucleic acids research. 2020; .

Marchi M, Ballone P. Adiabatic bias molecular dynamics: a method to navigate the conformational space of complex molecular systems. The Journal of chemical physics. 1999; 110(8):3697–3702.

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophysical journal. 2015; 109(8):1528–1532.

Mráziková K, Mlynsky V, Kuhrova P, Pokorná P, Kruse H, Krepl M, Otyepka M, Banas P, Sponer J. UNCG RNA tetraloop as a formidable force-field challenge for MD simulations. bioRxiv. 2020; .

Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature. 2008; 452(7183):51–55.

Pérez A, Marchán I, Svozil D, Sponer J, Cheatham III TE, Laughton CA, Orozco M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha gamma Conformers. Biophys J. 2007; 92(11):3817–3829.

Poblete S, Bottaro S, Bussi G. Effects and limitations of a nucleobase-driven backmapping procedure for nucleic acids using steered molecular dynamics. Biochemical and biophysical research communications. 2018; 498(2):352–358.

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindhal E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013; 29(7):845–854.

Rangan R, Watkins AM, Kladwang W, Das R. De novo 3D models of SARS-CoV-2 RNA elements and small-molecule-binding RNAs to guide drug discovery. bioRxiv. 2020; .

Rangan R, Zheludev IN, Hagey RJ, Pham EA, Wayment-Steele HK, Glenn JS, Das R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. Rna. 2020; 26(8):937–959.

Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nature structural & molecular biology. 2020; 27(10):959–966.

Šponer J, Bussi G, Krepl M, Banáš P, Bottaro S, Cunha RA, Gil-Ley A, Pinamonti G, Poblete S, Jurečka P, Walter NG, Otyepka M. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. Chem Rev. 2018; 118(8):4177–4338.

Steinbrecher T, Latzer J, Case D. Revised AMBER Parameters for Bioorganic Phosphates. J Chem Theor Comput. 2012; 8(11):4405–4412.

Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, Al-Hashimi HM. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nature chemical biology. 2011; 7(8):553.

Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Computer Physics Communications. 2014; 185(2):604–613.

Wacker A, Weigand JE, Akabayov SR, Altincekic N, Bains JK, Banijamali E, Binas O, Castillo-Martinez J, Cetiner E, Ceylan B, et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. Nucleic Acids Research. 2020; .

Wang L, Friesner RA, Berne B. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). The Journal of Physical Chemistry B. 2011; 115(30):9431–9438.

Watkins AM, Rangan R, Das R. FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. Structure. 2020; .

Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3 and 5 ends. Virus research. 2015; 206:120–133.

Zgarbová M, Otyepka M, Sponer J, Mlaadek A, Banas P, Cheatham III TE, Jurecka P. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J Chem Theor Comput. 2011; 7(9):2886–2902.