Article

Structural, Dynamical, and Entropic Differences between SARS-CoV and SARS-CoV-2 s2m Elements Using Molecular Dynamics Simulations

Adam H. Kensinger,[||] Joseph A. Makowski,[||] Kendy A. Pellegrene,[||] Joshua A. Imperatore, Caylee L. Cunningham, Caleb J. Frye, Patrick E. Lackey, Mihaela Rita Mihailescu, and Jeffrey D. Evanseck*

Cite This: https://doi.org/10.1021/acsphyschemau.2c00032

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The functional role of the highly conserved stem-loop II motif (s2m) in SARS-CoV and SARS-CoV-2 in the viral lifecycle remains enigmatic and an intense area of research. Structure and dynamics of the s2m are key to establishing a structure–function connection, yet a full set of atomistic resolution coordinates is not available for SARS-CoV-2 s2m. Our work constructs three-dimensional coordinates consistent with NMR solution phase data for SARS-CoV-2 s2m and provides a comparative analysis with its counterpart SARS-CoV s2m. We employed initial coordinates based on PDB ID 1XJR for SARS-CoV s2m and two models for SARS-CoV-2 s2m: one based on 1XJR in which we introduced the mutations present in SARS-CoV-2 s2m and the second based on the available SARS-CoV-2 NMR NOE data supplemented with knowledge-based methods. For each of the

three systems, 3.5 μs molecular dynamics simulations were used to sample the structure and dynamics, and principal component analysis (PCA) reduced the ensembles to hierarchal conformational substates for detailed analysis. Dilute solution simulations of SARS-CoV s2m demonstrate that the GNRA-like terminal pentaloop is rigidly defined by base stacking uniquely positioned for possible kissing dimer formation. However, the SARS-CoV-2 s2m simulation did not retain the reported crystallographic SARS-CoV motifs and the terminal loop expands to a highly dynamic "nonaloop." Increased flexibility and structural disorganization are observed for the larger terminal loop, where an entropic penalty is computed to explain the experimentally observed reduction in kissing complex formation. Overall, both SARS-CoV and SARS-CoV-2 s2m elements have a similarly pronounced L-shape due to different motif interactions. Our study establishes the atomistic three-dimensional structure and uncovers dynamic differences that arise from s2m sequence changes, which sets the stage for the interrogation of different mechanistic pathways of suspected biological function.

KEYWORDS: COVID-19, antiviral strategy, structure–function relationships, functional dynamics, RNA hairpin
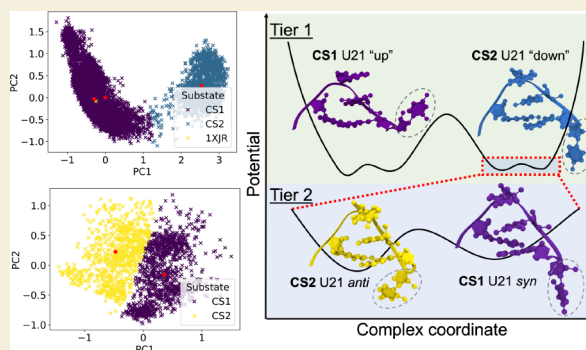
INTRODUCTION

Despite various advances in science, medicine, and logistics,[1−5] severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for COVID-19, remains a source of illness and death worldwide.[6] Although vaccine development is highly successful in combating ancestral SARS-CoV-2, the emergence of new variants has resulted in an escape from immunity, refocusing efforts on antiviral therapies.[7,8] Specifically, conserved elements within coronavirus genomes are enticing targets for antiviral intervention due to an implied biological function and phylogenetic stability unlike other targets that vary due to constant mutation and contribute to immune escape.[9−11] However, complete atomistic structures of conserved sequences from experimental data have yet to be

reported to guide the design and development of SARS-CoV-2 antiviral therapies.

The stem-loop II motif (s2m), a 41-nucleotide (nt) sequence in the 3′-untranslated region (UTR) of many coronaviruses, is one such highly conserved sequence which has been suggested to confer an advantage for replication despite ongoing debate over its function.[12−22] Although there is abundant speculation on the function of s2m in the
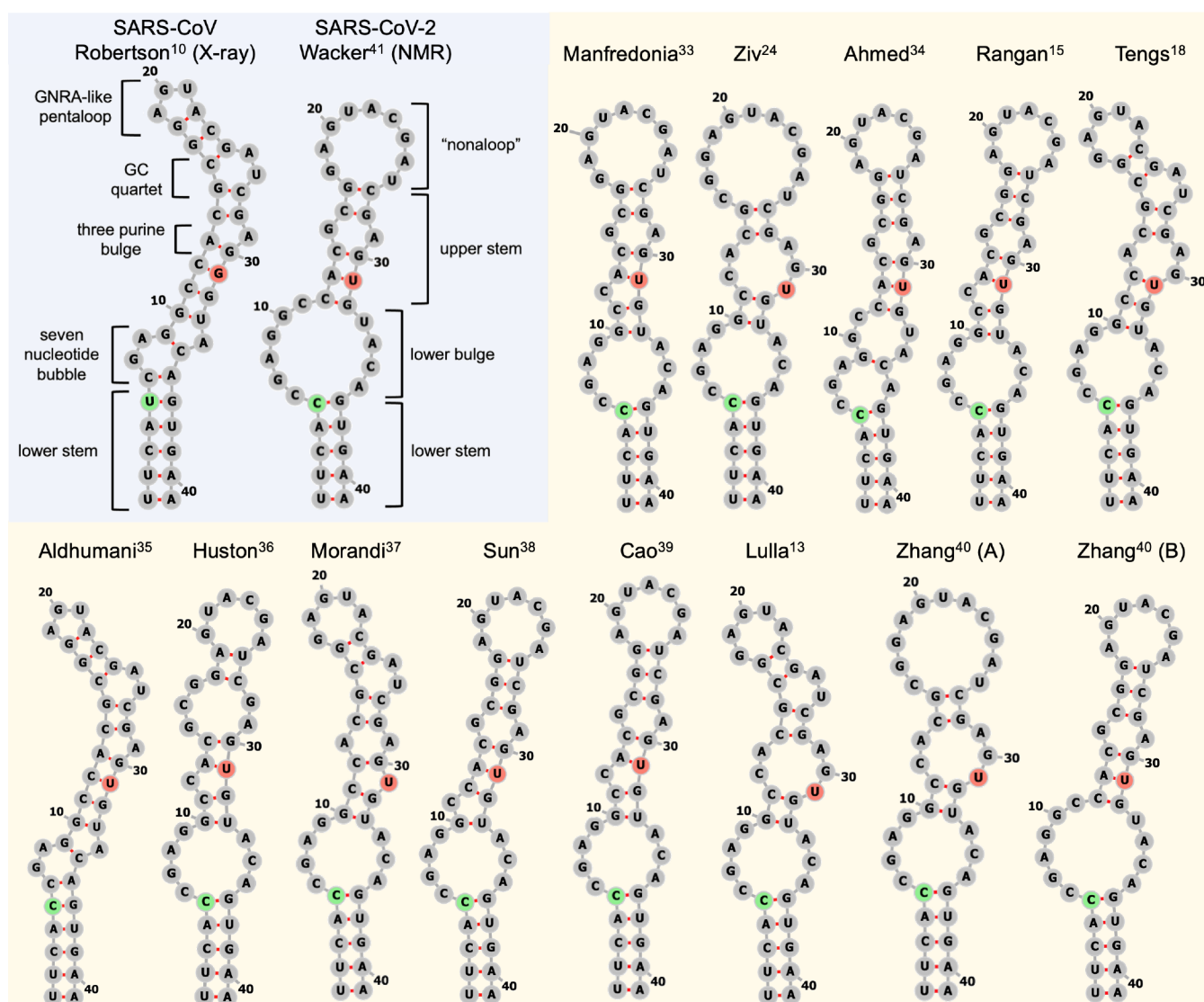
Figure 1. Secondary structures reported for the s2m in SARS-CoV, and SARS-CoV-2 used in this study highlighted in blue with secondary structures of SARS-CoV-2 s2m reported in the literature highlighted in tan. U5C mutation in green and G31U in red.

field,[10,13,17−21,23,24] we have reported that s2m contains a palindromic sequence in its terminal loop (GUAC, nt 20−23, Figure 1), a striking similarity to the conserved dimerization initiation site (DIS) in the human immunodeficiency virus (HIV) and X55 region in the hepatitis C virus (HCV).[11,25,26] During kissing complex formation between monomer RNA hairpin loops, unpaired terminal loop nucleotides form intermolecular base pairs facilitated through palindromic sequences.[27] Functionally, in these viral systems, these hairpins form kissing complexes that are converted to extended duplexes by the viral capsid protein, being involved in the genome dimerization.[11] However, mutations in SARS-CoV-2 s2m reorganize the hairpin secondary structure and impact kissing complex and extended duplex formation compared to SARS-CoV s2m.[25,26] While prior to SARS-CoV-2 the s2m element showed high conservancy, relative to the SARS-CoV s2m (Tor2 isolate), the SARS-CoV-2 s2m (Wuhan-Hu-1 isolate) differs in sequence by two nucleotides: U29732C and G29758U, named U5C and G31U in this study for simplicity (Figure 1).[11,28−30] Our experimental results showed that SARS-CoV s2m readily forms kissing dimers, whereas SARS-

CoV-2 s2m exists in an equilibrium between monomers and kissing dimers. Moreover, the viral N protein can only convert the SARS-CoV-2 s2m kissing dimers into the extended duplex conformation but not the SARS-CoV s2m. These results were surprising considering that both s2m elements have an identical GUAC palindromic sequence in their terminal loop and indicate that the two mutations present in SARS-CoV-2 alter the s2m tertiary structure.[11] Thus, further elucidation of how s2m mutations affect structure and dynamics is necessary for the understanding of mechanism and function. Consequently, there is a critical need to fill this gap in knowledge, where atomistic molecular dynamics (MD) simulation of the s2m in SARS-CoV and SARS-CoV-2 under physiological conditions can establish meaningful structural, dynamic, and thermodynamic information.

Robertson et al. reported the crystallographic structure of the SARS-CoV s2m to a resolution of 2.7 Å.[10] No crystallographic coordinates have been published for SARS-CoV-2 s2m; however, several secondary structures of s2m have been reported using computational techniques and low resolution[31,32] experiments (Figure 1).[13,15,18,24,33−40] In

general, the predicted SARS-CoV-2 s2m secondary structures fail to converge on a consensus base pairing pattern. Wacker et al. determined the solution phase secondary structure of SARS-CoV-2 s2m by 10 high resolution NMR NOE assignments, and in contrast with the findings of Robertson et al., they found that the terminal loop of SARS-CoV s2m expands into a larger loop in SARS-CoV-2, containing nine nucleotides ("nonaloop") due to a register shift produced by the G31U sequence alteration.[10,41] These findings indicate that the secondary structure of s2m in SARS-CoV-2 unambiguously deviate from s2m in SARS-CoV despite the high degree of sequence conservation (Figure 1).

Three-dimensional ensemble prediction employing Rosetta's FARFAR2[15,42] and coarse-grained folding in SimRNA have been reported.[33,43] Specifically, the 10 lowest energy SARS-CoV-2 s2m structures generated have many additional base pairs not identified by NMR NOE data, particularly in the hairpin terminal loop. Additionally, three-dimensional SARS-CoV-2 s2m structures based on cryo-electron microscopy (cryoEM) and SHAPE secondary structures were further developed through algorithmic structure prediction giving 12 base pairs instead of the 10 found by NMR.[13,34] The noted differences generated from each technique with NMR underscores the importance of our work, where prior to performing atomistic simulations of RNA, it is essential to identify experimentally derived, biologically relevant starting structures.[44−47]

In our study, we elucidate a hierarchy of structural and dynamical features and thermodynamic parameters of the s2m in three systems. In the first phase of our study, we simulated a model of the SARS-CoV s2m based on crystallographic coordinates (PDB ID: 1XJR) to establish a baseline of structure and dynamics. In the second phase, we assess the robustness of a traditional 1XJR MD homology model of SARS-CoV-2. In the third phase, we employ knowledge-based RNAComposer software to extrapolate the Wacker et al. secondary structure from NMR NOE assignments to generate an initial 3D structure for simulations.[48] With initial coordinate agreement with Wacker et al. and ability to access conformations and dynamics that occur on microsecond time scales with nonbiased molecular dynamics, we sample and classify a broad hierarchy of structural and dynamical features distinguishing SARS-CoV and SARS-CoV-2 s2m. Finally, to better understand our results showing different s2m homo-dimerization kinetics between SARS-CoV and SARS-CoV-2,[11] we estimate the absolute entropy of each s2m based on MD simulation position covariances. Ultimately, our simulation work explains the different dimerization properties of SARS-CoV and SARS-CoV-2 s2m.[11] However, our results also provide atomistic structure and dynamics of the s2m element that can be broadly used for exploration of different mechanistic pathways of suspected biological functions.

■ METHODS

Dynamic Hierarchy

RNA dynamics occurs across at least 16 orders of time scale magnitude, with complex behavior from static to highly dynamic structures,[49] which has been critically reviewed.[50−52] Systematic decomposition of RNA dynamics into a hierarchy of tiers defined by the time scale and free-energy barriers of transitions associated with dynamical modes helps to simplify the rugged conformational landscape into a framework that is easy to describe and apply to functional understanding.[53,54] The approach applied in this work

classifies RNA dynamics into Tier-0 (distinct secondary structure dynamics, >0.1 s), Tier-1 (base-pair and tertiary dynamics, milliseconds to seconds), and Tier-2 (thermodynamic jittering of bases and base stacking, pico- to microseconds) which was introduced by Mustoe et al. based on the original hierarchical descriptions of functionally important protein dynamics developed by Frauenfelder et al.[53,54]

Molecular Dynamics Simulations

The accurate description of structure and dynamics expected for biological function, in terms of hierarchical RNA dynamics tiers, requires computational techniques involving an experimentally relevant starting structure,[44−47] a force field attuned to the system,[55,56] and a time scale that can access the functionally important conformational changes.[50−52] Overall, three s2m systems were modeled and simulated in this study: SARS-CoV s2m based on 1XJR coordinates, SARS-CoV-2 s2m homology model based on 1XJR coordinates, and SARS-CoV-2 s2m leveraging NMR NOE assignments[41] and knowledge-based RNAComposer[48] coordinates. All three s2m models were solvated and ionized using the tLeap program included in AmberTools20.[57] The experiments used for deriving the starting structures were performed using various concentrations of different ions, but our interest is to align simulated conditions to our dimerization experiments. As discussed and outlined below, we solvated each system with 15 Å of TIP3P[58] waters and added one Mg^{2+} cation, resulting in concentrations of ca. 3.5 mM Mg^{2+}, within range of our previously reported experimental dimerization conditions.[11] The residual negative charge was neutralized with Na^+ atoms. The s2m simulations were run at two temperatures, 283 and 310 K, to represent the experimental temperature from the Wacker et al. 2D NMR study and physiological temperature, respectively.[41] Systems, including ions, were minimized for 1000 steps with the conjugate gradient algorithm and equilibrated under the NPT ensemble with periodic boundary conditions using AMBER $ff99\chi OL3$[59,60] force field parameters with NAMD[61] molecular dynamics engine. Systems were deemed equilibrated when potential energy and volume stabilized. Production run simulations were performed for 3.5 μs using the NPT ensemble.[62]

An extensive framework defining the ionic atmosphere of RNA has been established,[63−70] where condensation effects result in an accumulation of cations around RNA.[63,69] However, the impact of ions upon RNA in experiment are found to be different in simulation.[49,71,72] The controversy over ionic species,[71,73,74] placement,[75−79] and conditions[80,81] between experiment and modern simulations have been exquisitely detailed and reviewed by Šponer and co-workers.[49] Despite extensive studies and advances, fundamental issues, intrinsic differences, and counterintuitive effects have been identified from both experiment[73,81−83] and computation[71,84−87] that prevent an exact alignment of ionic conditions that achieves meaningful structural and dynamical information from MD simulations. However, the procedure of simple charge neutralization by monovalent cations, either Na^+ or K^+ interchangeably,[71,72,85] and experimentally identified Mg^{2+}, as in our simulations, is predicted to result in acceptably small error on the short time scale of MD simulation.[49,71,84] In the crystallographic structure 1XJR, two Mg^{2+} cations were identified, and the experimental $[Mg^{2+}]$ range of 1−10 mM corresponds to one (3.5 mM) or two (7.0 mM) Mg^{2+} cations in the system size simulated.[10,11] Our interest is in the condensed liquid phase with ~3.5 mM $[Mg^{2+}]$ and not the 60 mM Mg^{2+} solid phase 1XJR conditions. As a check of our ionization process, we studied the effects of charge neutralization with $n \cdot Mg^{2+}$ (n = 1, 2, 13) and Na^+ in SARS-CoV s2m, and we observed minimal differences between n = 1, 2 and the expected dynamic dampening effect for n = 13 (Figure S1).[74] With the computed ion condensation effects, it is plausible that our simulations overestimate the dynamics and structural variability. Thus, given the limitations and successes known to the field, we elected to follow the approach of performing MD simulations that include divalent cations when explicitly reported by experiment and monovalent cations to complete the neutralization. While recent advances in polarizable force fields represent headway into the

accurate portrayal of the ionic atmosphere of RNA,[87,88] our simulations using a standard protocol provide a reasonable estimate of the structure and dynamics of SARS-CoV and SARS-CoV-2 s2m.[84]

Simulation Analysis

Visual Molecular Dynamics (VMD) software was used to visualize simulated systems and perform measurements of distances and angles, including H-bond occupancies.[89] The traditional use of root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) were used through our study, as implemented in VMD. Unless otherwise specified, the first frame following equilibration was used as our reference structure for RMSD or RMSF. Base stacking interactions were studied using nucleotide center of mass coordinates for distances, and stacking energies were obtained through motif identifier for nucleic acid trajectories (MINT) using the AMBER force field.[90] Details regarding our configuration of MINT are provided in the Supporting Information.[70] Structural analysis and helical parameter calculation were implemented in Web 3DNA 2.0, and secondary structures were visualized in ViennaRNA Forna Web Services.[91,92]

Multivariate Statistical Analysis

Principal component analysis (PCA) uses the eigenvalues and eigenvectors of the covariance matrix $\mathbf{C}$ of atomic coordinates, whose entries c_{ij} are given by

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

to reduce the dimensionality of data down to a subspace responsible for the variance. PCA facilitates the classification of a hierarchy of conformational substates (CS) sampled through a simulation.[54,93] The covariance matrix of atomic positions was determined for each simulation, and the eigenvectors ("principal components") capturing the greatest proportion of the total variance were used as the basis of a lower dimensional subspace. We performed these computations with different indices of selectivity to capture a range of dynamics from global to local motion. Generally, no less than 50% of the system variance was captured within three or fewer eigenvectors as judged by scree plots, which depict the proportion of total variance captured within each principal component. Each position vector, originally containing $3N$ spatial variables over N atoms, was projected onto the subspace spanned by the principal components, yielding 2- or 3-D data capturing the functional dynamics and conformational substates within each system.[94,95]

Estimation of Absolute Entropy

Based on foundational work by Schlitter,[96] Andricioaei and Karplus reported a quasiharmonic approximation to the absolute entropy of macromolecules.[97] Assuming atomic fluctuations follow a multivariate Gaussian distribution, vibrational frequencies calculated from the eigenvalues of a mass-weighted covariance matrix $\mathbf{C}'$ are used to calculate the absolute entropy. We leveraged the ability to perform these computations with different indices of selectivity to approximate the absolute entropy of all atoms in each s2m oligomer, the nucleotides encompassing the SARS-CoV-2 terminal loop (residues 17 to 27), and the palindromic sequence. Mass matrices were derived from the AMBER parameters employed for simulation.

■ RESULTS AND DISCUSSION

SARS-COV S2M

Structure and Dynamics of SARS-CoV s2m. Our model of the SARS-CoV s2m first defines changes when transitioning from the crystallographic structure to dilute solution. The RMSD (Figure S2) over the course of the 3.5 μs simulation remained low at 2.45 ± 0.68 Å at 283 K and increased to an average of 5.08 ± 0.98 Å at 310 K. The 283 K simulation average RMSD stayed within crystallographic resolution of 2.7 Å, while the 310 K RMSD more than doubled, indicating increased structural variation. After ca. 1.6 μs, the average

RMSD remained nearly constant for both simulations. Specific sources of structural deviation were identified by RMSF (Figure 2).
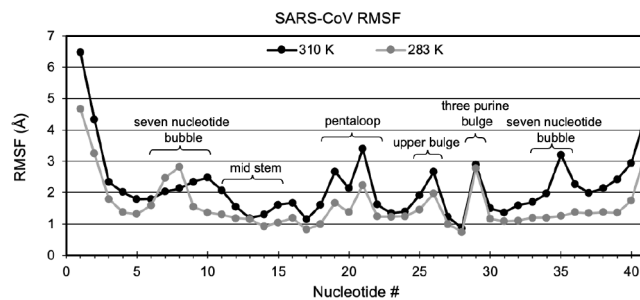

Figure 2. Nucleotide RMSF of SARS-CoV s2m labeled by motif.

From the RMSF analysis, the greatest fluctuation was found at the terminal 3′ and 5′ ends, bulges, and terminal loop secondary structural elements. Not surprisingly, simulations at both temperatures experience base fraying since our physical model included only the s2m element (41 nt) without the two additional base pairs from the crystallographic structure.[10]

The RMSF analyses mirror each other except for two differences in the seven-nucleotide bubble, which were found to be influenced by stem fraying and should not be overinterpreted. However, and most important to our study, the terminal GNRA-like pentaloop of C18-G22, including the palindromic sequence of G20-C23, is computed to have higher RMSF at 310 K, consistent with the reported crystallographic U21 distortions. The average RMSF per motif revealed rigidity in the constitutive sets of nucleotides, each having less than 2.5 Å of fluctuation for both temperatures (Figure S3).

Comparison of Simulation to Crystal Structure Motifs. The exquisite crystallographic analysis reported by Robertson[10] is central to understanding the current simulation work. The first motif of interest is a GNRA-like pentaloop having relatively stable bases, except for U21, which is exposed to solvent and is thermodynamically unstable based on B-values (Figure S4). While the lower stem of our model experienced fraying, absent in 1XJR, which is stabilized by an additional two base pairs compared to our model, nucleotide fluctuation in the upper portion of the hairpin is in qualitative agreement with B-factor dispersion, especially for U21. Expanding upon the reported U21 distortions, the simulations demonstrate oscillatory behavior of U21, alternating between *syn* (37%) and *anti* (45%) dihedral angles.[98] The U21 base is computed to have unique flexibility compared to the other palindromic nucleotides and is of future interest in kissing dimer initiation. The χ angle of the remaining pentaloop nucleotides remained stable and reflect the angle reported in 1XJR, however the χ angle of G20 oscillates about a mean angle that is 78° higher than that reported in the crystal structure (Table S1). The same behavior was observed at 283 K.

Finally, MINT analysis identified base stacks with high occupancy, but weak interaction energies, as a function of the Coulombic and van der Waals energies. The center of mass (COM) distances between the stacked bases in the terminal loop are relatively rigid with low standard deviation, indicating no stack reshuffling occurred during the simulation, having expected entropic and geometric consequences, discussed in more detail in the entropy section (Table S2).

The second motif of interest is the GC quartet composed of 10 hydrogen bonds (Figure S5). The GC quartet was found to be unusually contained within the double-helical axis with the consequence of rendering A25 and U26 to be excluded and solvent accessible, explaining the relatively high RMSF of these nucleotides (Figure 2). The GC quartet persists throughout the entirety of the 283 and 310 K simulations with high hydrogen bond occupancy (65−94%) and heavy atom interaction distance measurements ranging between 2.88 and 3.13 Å with low standard deviation (0.1−0.3 Å) (Figure S5 and Table S3). The inflexible GC quartet leads to stabilization of other key structures such as the GNRA-like pentaloop and L-shape kink formed by deformation of the backbone. Specifically, the occupancies and distances calculated for the quartet indicate that this region is thermodynamically stable, lowering the entropy for this motif and adjacent structures by sampling such a tight conformational space.

Finally, the Mg^{2+} binding tunnel, a motif of considerable pharmaceutical interest, is composed of the three-purine bulge and seven-nucleotide asymmetric bubble. From crystallographic analysis, two conformations involving A29 and the G7/A8 stacked unit were reported. The predominant conformation has a long-range interaction with a distance of 2.7 Å between N1 of A29 and N2 of G7 forming a negatively charged cavity in which the phosphates of the bubble are turned inward and to which Mg^{2+} can bind (Figure S6). In contrast, our simulation showed a significant increase in this specific distance of 14.5 ± 2.2 Å at 310 K (Figure S7). The second long-range interaction was not observed between A29 and A8. The increase in distance and shift in position of the residues may be due to the lack of crystal packing forces as it melts into the condensed liquid state. However, the non-canonical G-A base pair in the purine bulge remained stable based on relatively high base pair occupancies and low standard deviation of interaction distances (Table S4).

We elected to retain the Mg^{2+} within the tunnel to simulate a concentration consistent with physiological conditions and that reported by *in vitro* experiments.[11] To ensure that the removal of 1 Mg^{2+} ion did not significantly alter the structure of our system, we simulated for 3.5 μs the SARS-CoV s2m from the edited 1XJR system but maintained the coordinates of both crystallographic Mg^{2+} ions. Traditional RMSD and RMSF analyses and PCA confirmed that the system with both Mg^{2+} ions behaved analogously in structure and dynamics to the simulation with only one Mg^{2+} ion. Direct comparison of RMSD and RMSF between the two systems did not reveal significant change (Figure S8, parts A and B). A small increase in RMSF was computed for C35, which can be attributed to a fleeting interaction with U1 due to fraying (Figure S8C).

Overall Shape of the SARS-CoV s2m. The shape of the SARS-CoV s2m from crystallography shows an L-shape fold due to a kink stabilized by the interactions found in the four mentioned motifs. Residues A25 and U26 produce a tight bend in the backbone of the region, allowing the L-shape kink to form (Figure 3A). The angle of the L-shape kink in the crystal structure was found to be 65°.

After 3.5 μs of MD simulation, this angle had widened to 80.6° ± 5.0° at 283 K and to 67° ± 7.5° at 310 K (Figure S9), remaining nearly constant through the simulation. A two-sample t test performed at 95% confidence interval indicated these values were not significantly different from one another. A25 and U26 are positioned at the fulcrum of the L-shape hinge, yet with limited motion about the L-shape kink as
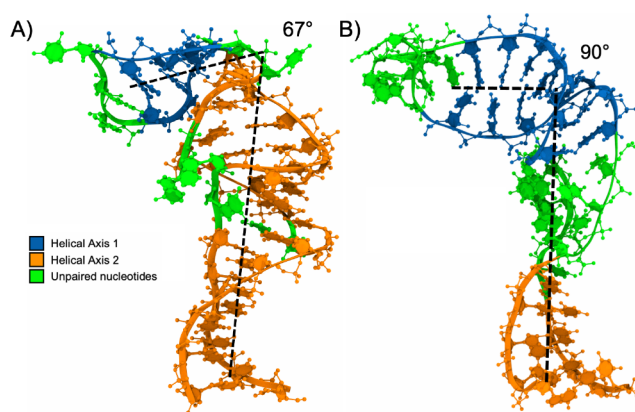
Figure 3. (A) L-shape kink angle defined by C3:N4, U26:P, and G18:N3 in SARS-CoV s2m 310 K model. (B) L-shape kink angle defined by C3:N4, C14:C2, and U26:O4 in SARS-CoV-2 s2m 310 K.

indicated by the low standard of deviation of 7.5° at 310 K. Thus, the overall three-dimensional shape, structural stability of adjacent motifs, and hinge-like dynamics of the s2m are largely influenced by the stiffness provided by the GC quartet. Based on this analysis, disruption of the GC quartet would plausibly modify any structure−function relationship of the SARS-CoV s2m. Ultimately, we find that the RMSD and RMSF is aligned with crystallography, where all the motifs have RMSF values less than the crystallographic resolution of 2.7 Å, setting a reference frame for the structural and dynamics comparison with the SARS-CoV-2 s2m.

Defining Structural Hierarchy with PCA. To gain deeper insight into the dynamics of the system beyond the crystallographic description, principal component analysis (PCA) was carried out using all the nucleotide non-hydrogen atoms. For both the 310 and 283 K simulations, the resulting scree plots captured over 50% of the structural variation in the first three strongest eigenvalues (Figure S10, parts a and b). The structural variance and information extracted by PCA was found to be dominated by lower stem fraying, as expected from the physical model used for the simulations (Figure S11).

As such, we refined our index of selectivity to target the different tertiary motifs. Specifically, this is the GNRA-like pentaloop, since it is the element of concern when extending a structure, dynamics, and function connection on the suspected formation of the s2m kissing dimer complex. PCA was applied to the non-hydrogen atoms of the GNRA-like pentaloop (nucleotides G18-A22) over the entire 3.5 μs. The resulting scree plot captured more than 50% of the structural variation in the first three strongest eigenvalues (Figure S10c).

Analysis of 2D projections of the PCA was used to extract 3D structures for comparison (Figure S11). Figure S10a−c contains other combinations of 2D plots. At 310 K, two substates, CS1 and CS2, were identified, where the difference is defined by a U21 "up" conformation in CS1 versus a U21 "down" conformation in CS2 (Figure 4A).

Enumeration of the simulation frames belonging to each CS (Figure 4B) shows that the pentaloop remains in the U21 "up" substate for approximately 90% of the simulation (Figure 4C), the remaining 10% (~350 ns) where U21 is in the "down" substate (Figure 4C). While "down", we observe that U21 forms new backbone tertiary interactions with A23 and C24. Base stacking analysis with Barnaba and MINT of each CS centroid structure revealed that the nucleotides aside from U21
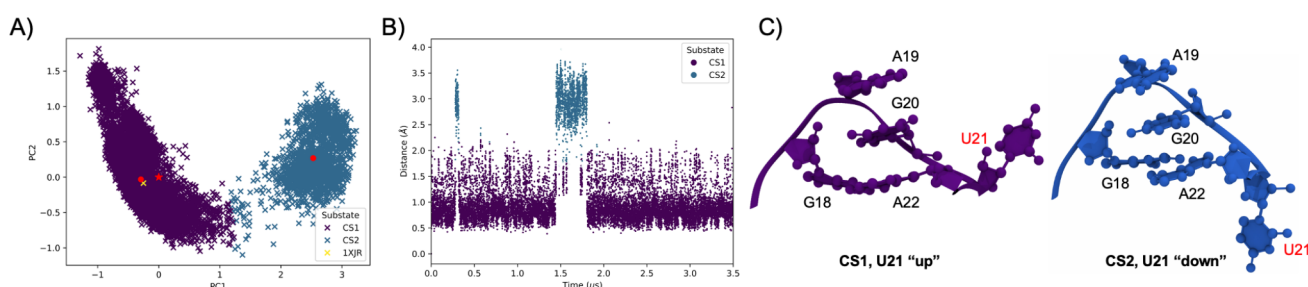
Figure 4. (A) 2D PCA plot for the GNRA-like SARS-CoV s2m pentaloop at 310 K. The *k*-means method was used to identify the centroid of each conformational substate (red dot), while the average structure of the entire simulation is represented by a red star. The 1XJR structure is given by the yellow cross. (B) Pentaloop RMSD colored by CS. (C) Centroid structures for CS1 and CS2 with U21 up or down.

in the pentaloop are rigidly stabilized by base stacking interactions (Figure 5). It was found that an inflexible "stack
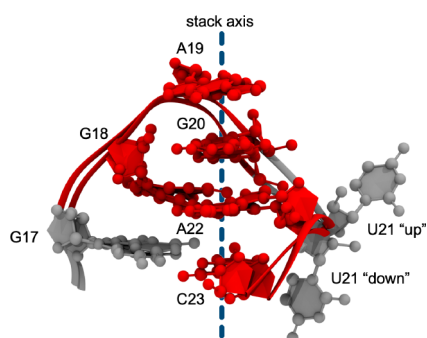
Figure 5. Superimposed CS1 and CS2 centroids of SARS-CoV 310 K pentaloop and adjacent stem base-pair (G17:C23) identified by MINT. Dashed line indicates the axis along which the red nucleotides form a rigidly preorganization stack

axis" (Figure 5) consisting of three of four palindromic nucleotides persists through the entire simulation—regardless of CS—with occupancies greater than 96%. These interactions contribute greatly to the stability and rigidity of the SARS-CoV s2m pentaloop, and the preorganization of palindromic bases for kissing dimmer complexation, with the notable exception of U21, which remains unstacked through CS1 in the "up" conformation and only makes touching interactions with C23 (52%) and G24 (24%) while in the CS2 "down" conformation.

Per the theory of RNA dynamical hierarchy,[53] we expect that multiple Tier-2 substates exist within any one Tier-1 substate. We separated the Tier-1 CS structures and performed PCA again on each partition. In agreement with classical structure analysis above, we found that the dominant mode within the "up" and "down" substates was rotation of U21 from *syn* to *anti*. This further corroborates our characterization of the s2m pentaloop as rigid since no other motions are detected. In this manner, PCA combined with a clustering algorithm has been shown to be a potentially useful tool for elucidating a hierarchy of RNA substates.

In conclusion for SARS-CoV, through detailed description of how each motif evolves through the simulation, a foundation for a potential structure–function relationship is produced. Most importantly, the terminal pentaloop is stabilized by stacking interactions with the thermodynamically disordered U21 excluded from this organization. The GNRA-like pentaloop is additionally stabilized from below by stem content and by backbone rigidity provided by the GC quartet, but U21, swung-out and exposed to solvent, was found to be

characterized by two tiers of dynamical modes: up/down swinging and *syn/anti* torsion. The hinge-like dynamics about the L-shape kink, which was maintained during the shift from condensed solid phase to solution phase, were also described for the first time. Ultimately, our results confirm the crystallographic description of Robertson et al. with SARS-CoV s2m yet expand upon the knowledge by providing a hierarchy of s2m dynamics in solution that underscores the rigidity and geometric preorganization of the palindromic sequence for eventual kissing complexation.

SARS-COV-2 S2M

Structure and Dynamics of SARS-CoV-2 s2m Homology. It is commonly assumed that sequence conservation implies three-dimensional structure conservation.[99] Indeed, two MD studies have simulated SARS-CoV s2m on 10 and 180 ns time scales, respectively, including comparison models of SARS-CoV-2 s2m using 1XJR as starting coordinates.[23,35] Thus, our first model of SARS-CoV-2 s2m used the protocol of homology modeling, using the 1XJR crystallographic structure of SARS-CoV s2m (see Methods). However, after 3.5 μs of homology simulation, the model deviated only slightly from the secondary and tertiary structure of SARS-CoV, according to an RMSD of 2.32 ± 0.5 Å (Figure S12A). Notably, each of the SARS-CoV s2m motifs described by Robertson et al. was maintained throughout the 3.5 μs simulation of the homology model. Furthermore, the computed RMSF was aligned with SARS-CoV s2m model (Figure S12B). Ultimately, the SARS-CoV-2 s2m homology model simulations indicate nearly identical structures with SARS-CoV despite the C5U and G31U mutations, where the maintained secondary structure aligned with SARS-CoV is inconsistent with that reported by NOE NMR results by Wacker, establishing challenges for homology modeling of SARS-CoV-2 s2m (see Table S5).[41]

Structure and Dynamics of SARS-CoV-2 s2m NMR. Wacker et al. determined secondary structure characteristics of the RNA element using high-resolution 2D NMR techniques.[41] The group identified 10 nucleic acid base pairs with eight of the 10 being corroborated by DMS footprinting.[41] Two transient base pairs at the base of the terminal loop (C16−G28, G17−C27) were additionally not detected by ^1H,^15N-HNN-COSY or ^1H,^1H-NOESY experiments at 298 K, but were detected at 283 K.[41] Thus, we used these NOEs to construct an appropriate starting model consistent with the experiment, as described in Methods.

Global Structure Analysis, L-Shape Kink. We find that over the first 1.2 μs of the trajectory, the 310 K SARS-CoV-2 s2m backbone RMSD steadily increased, plateauing to 10.5 ± 0.62 Å (Figure S13). The source of the large RMSD, as

compared to that computed for SARS-CoV s2m, originates from the observed register shift in base pairing, expansion of the lower bulge, and reduction of the lower stem content and ultimately leads to increased fraying. In contrast, at the lower 283 K temperature, the RMSD was lower on average with a greater standard deviation, suggesting significant differences in the dynamics of each system.

RMSF was then used to uncover the source of the large amplitude fluctuations on a per nucleotide basis. As noted in the RMSD analysis, the SARS-CoV-2 s2m nucleotides with the greatest fluctuation are the fraying base pairs at the base of the lower stem, where nucleotides 1−2 and 39−41 result between 6.5 to 15 Å from the initial structure. However, terminal loop nucleotides 18 to 26 resulted in increased RMSF values ranging from 3.0 to 8.0 Å, indicating increased flexibility and motion for this region. In particular, the magnitude of the RMSF from the pentaloop ranges between 1.75 to 3.25 Å for SARS-CoV (Figure 2), whereas the range for the terminal nonaloop more than doubles to 4 to 8 Å in SARS-CoV-2 (Figure 6). Comparing SARS-CoV and SARS-CoV-2 s2m this way, we uncovered that the expansion from a pentaloop to a nonaloop is a source of increased flexibility.
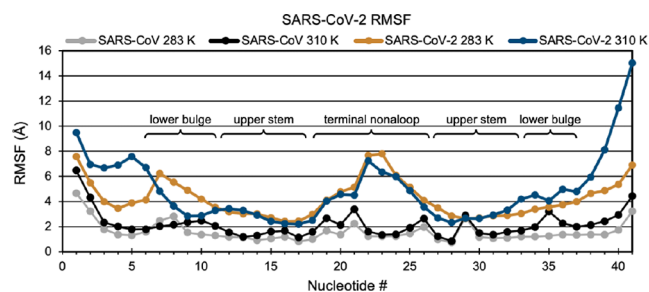
Figure 6. Nucleotide RMSF of SARS-CoV-2 models, structure labels apply to SARS-CoV-2 s2m. Both are compared to SARS-CoV models in gray (283 K) and black (310 K).

Our multivariate statistical analysis of the SARS-CoV-2 s2m revealed in finer structural detail the source of changes to the RMSD and its low standard deviation relative to the 283 K simulation (Figure 7A). Figure S10d−f contains other combinations of 2D plots. PCA analysis of the whole SARS-CoV-2 s2m identified four CSs (Figure 7A) defined dominantly by dynamics of the lower stem (Figure 7C) Following equilibration, the structure begins in CS1 and is found to make a tertiary kink angle involving a long-range interaction between G8 and A29. Beginning at ca. 600 ns, the

simulation enters CS2, characterized by a stretching of the structure. In the lower stem, base pairs are broken, and the structure is stabilized by base stacking and sugar−phosphate interactions up the backbone.

Despite progressive change from the starting structure postequilibration, the long-range interaction stabilizes the structure in the aforementioned kink angle, preventing large-magnitude tertiary fluctuations. The transition to CS3 occurs at 1.2 μs, where the s2m frays completely and causes the lower stem and bulge to fold into artificial tertiary interactions. CS4 remains in a similar folded local minimum and is defined in part by stack swapping and swing-out dynamics between the frayed stem and lower bulge. Overall, the SARS-CoV-2 s2m oligomer dynamics are computed to become more disorganized in solution, suggesting a structural and dynamic basis for extended duplex formation (described in conclusions).

In comparison, the 283 K simulation SARS-CoV-2 s2m never adopted the stabilizing kink angle which persisted through the 310 K simulation, and the PCA-reduced data did not cluster beyond an isotropic point cloud. Our *ansatz* was that this tertiary structural dynamic accounted for the difference in RMSD magnitude and standard deviation. Least squares linear regression between the "head-to-tail" distance of the 283 K s2m and the RMSD yielded an R^2 value of approximately 0.6, suggestive of a weak, yet explanatory correlation. Thus, at physiological temperature, the ensemble of structures sampled is described by four conformational substates (CS) defined in large part by lower stem fraying dynamics. The fraying revealed through simulation may explain the unique ability of the SARS-CoV-2 s2m to form an extended duplex in the absence of N protein.[11]

In contrast, a relatively small proportion of the SARS-CoV s2m experienced significant fraying, also consistent with our expectation that preorganization of the palindromic nucleotides facilitates its tendency to convert to the kissing homodimer. Finally, the tertiary interaction resulting in the 310 K simulation kink angle results in an ensemble of structures qualitatively similar to SARS-CoV s2m, but fine structural analysis below will reveal fundamental differences between the two structures.

The SARS-CoV-2 s2m simulated model deviates from SARS-CoV s2m in relation to the motifs identified by Robertson et al., yet some familiar global architectures are retained that are presumed necessary for biological function such as the L-shape kink. In SARS-CoV-2 s2m at 310 K, the hinged structure had an angle of 90.0° ± 12.1° which is comparable to the kink angle in SARS-CoV s2m, despite drastically different starting coordinates (Figure 3B).
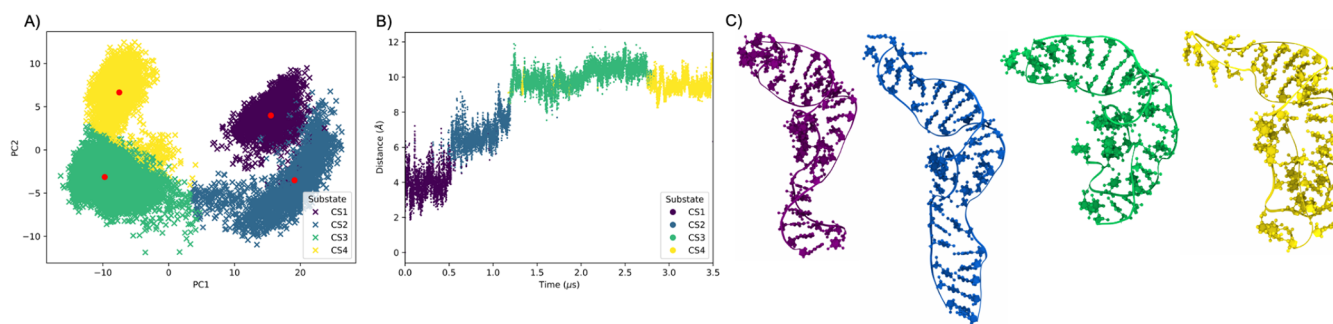
Figure 7. (A) SARS-CoV-2 s2m 310 K heavy atom PCA partitioned into four CSs illustrating the time evolution, CS centroids given as red dots. (B) RMSD colored by CS as the trajectory progresses in time. (C) CS1−CS4 centroid structures.
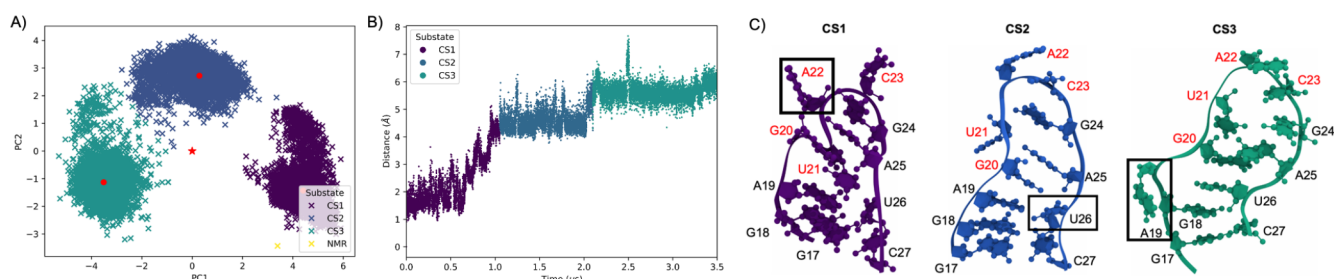
Figure 8. SARS-CoV-2 s2m heavy atom terminal loop PCA. (A) Terminal loop heavy atom PCA, three CS identified. PCA average (red star), knowledge-based NMR starting structure (yellow). (B) Terminal loop RMSD colored according to the conformational substate. (C) CS1−CS3 centroid structures with palindromic nucleotides in red. Major sources of variation tend to be swung-out (boxed), unstacked nucleotides.

Contrasting the 67° backbone twisting anchored at the GC quartet hinge in SARS-CoV, the origin of the SARS-CoV-2 kink is located within the GA base pair present in the stem. Two long-range tertiary interactions stabilize the bend throughout the simulation and were observed to disappear fleetingly at ca. 1.5 μs, which corresponds to an overall opening of the L-shape kink three-dimensional structure. These interactions were characterized to be backbone to base interactions G9O2′-A29OP2 and G9N2-G28O5′ with average distances of 3.52 ± 1.4 and 4.12 ± 2.3 Å, respectively. The deformation of the backbone impacted shape of the upper stem helix and Web 3DNA 2.0 was unable to classify much of the helix as the canonical A form (Table S6).

The transient making and breaking of these interactions cause a global dynamic of opening and closing of the L-shape kink. No long-range interaction was found between G9 and G28 or A29 at 283 K. However, the global kinked shape of the SARS-CoV-2 s2m has been previously reported from 3D reconstructions calculated from cryoEM, and our structure shares clear visual similarities to this data.[13]

Characterization of Motifs. Next, we systematically investigated the SARS-CoV-2 s2m for the presence of each motif characterized in the SARS-CoV s2m crystal structure and monitored those regions for the duration of the 3.5 μs simulations at both temperatures. The three-purine bulge and seven nucleotide asymmetric bubble are not present in the secondary structure of SARS-CoV-2 s2m (Figure S14).

Most importantly, the inflexible GC quartet was not observed in the SARS-CoV-2 s2m simulations, engendering major structural differentiation in the upper stem. At the base of the upper stem, a triple-interaction between G32, C12, and the phosphate of C11 helps to stabilize the expanded lower bulge region and was observed to form at both 283 and 310 K with interaction occupancies of 73−90% (Figure S14). Characterization of the Mg^{2+} ion in the cavity of s2m revealed a binding pocket analogous to SARS-CoV s2m. Phosphate oxygens of G7 and A8 were found to coordinate with the ion for the entire 3.5 μs with an average distance of 1.93−1.94 Å. Similar locations of Mg^{2+} binding in SARS-CoV and SARS-CoV-2 s2m expose the importance of metal binding to the three-dimensional shape.

Nonaloop and Palindrome. The GNRA-like pentaloop found in SARS-CoV s2m was replaced with an expanded nonaloop due to the U5C and G31U mutations, which we found to be more flexible and disordered. Dihedral angles (χ) were measured as a function of time for residues 18−26 to quantitate the level of disorder conferred to the terminal loop from the loop expansion, with particular attention paid to the palindromic GUAC nucleotides 20−23. In general, at 310 K,

we find the standard deviation in χ to be ca. 50% of the average value over the 3.5 μs simulation. Specifically, G20 started the simulation in the *syn* conformation but converted to *high-anti* and *anti* at ca. 1.0 and 2.1 μs, respectively with an average angle of 189.4° ± 81.8°; U21 was stable in the *syn* conformation until it converted to *anti* at 1.1 μs with an average angle of 191.8° ± 87.0°; A22 was dynamically disordered throughout the simulation but stabilized at ca. 1.8 μs with an average angle of 188.2° ± 47.2°; C23 began the simulation in the *high-anti* conformation but drifted down slowly to a stable *anti* conformation with an average angle of 219.5° ± 28.2°. The standard deviations are in sharp contrast to the highly rigid GNRA-like pentaloop in SARS-CoV of 9−20°. Dissecting the terminal loop in terms of individual dihedral angle variation unveiled concerted conformational changes for nucleotides G18, A19, G20, and U21 that lead to stack reshuffling dynamics that require deeper conformational analysis to understand the structural impact.

PCA Terminal Loop Dynamics. It is immediately apparent from the Wacker et al. secondary structure of the SARS-CoV-2 s2m that there is a substantial Tier-0 change in the terminal loop secondary structure. Due to a lack of stabilizing base pairing, we expected that the nonaloop is characterized by a greater range of dynamics occurring on a smaller time scale than the pentaloop. PCA and *k*-means of the nonaloop region reveal three conformational substates (Figure 8A). As anticipated, multiple dynamical modes (Figure 8, parts B and C) distinguish the centroids. While each CS is characterized by an abundance of base stacking interactions, we observed stack-swapping dynamics that change which nucleotides participate in stabilizing the loop. Consequentially, we identify unique unstacked, solvent-exposed nucleotides in each CS, which were measured by intra-CS RMSF to account for the structural disorder (Figure S15).

Terminal Loop Base Stacking. To quantify the differences in CS base stacking interactions, MINT and Barnaba were applied to each centroid structure (Figure S16A). In contrast to SARS-CoV, no two centroids were found to have identical base stacking patterns, due in part to the swung-out nucleotides distinguishing each CS. The A19/G20 stacking interaction identified by MINT in CS2, for instance, is broken in CS3 as A19 is swung out. Although A22 corresponds to a local maximum in RMSF in all CSs, it is evident that different nucleotide stacking combinations yield unique peak magnitudes within each CS. Distortions in the backbone also influence which nucleotides participate in stacking interactions. The palindromic sequence also differs in stacking orientations between CS.

H

In contrast with the SARS-CoV s2m, the palindromic sequence of SARS-CoV-2 is relatively free and exposed to solvent in SARS-CoV-2 (Figure S16B). Transient base stacking interactions fix the palindromic nucleotides in place within individual CS, but no conformation is as rigid as the SARS-CoV pentaloop. Due to the increased flexibility, we expect that there will be a large entropic penalty associated with kissing dimerization, and the computed geometric positioning and base melting should not encourage kissing dimer formation.

In the same manner as SARS-CoV s2m, each Tier-1 nonaloop CS was subject to multivariate analysis again. Within all Tier-1 CSs, a total of seven Tier-2 CSs were extracted. Generally, this extra step of refinement managed to isolate individual modes better, with most of the variation in within a given CS being attributed to motions not dissimilar to the Tier-1 up/down dynamics in SARS-CoV. While the relatively large Tier-2 ensemble is suggestive of greater dynamic flexibility, we recommend a single iteration of PCA as the most economical approach for characterizing important changes in highly disordered systems.

In summary for SARS-CoV-2, homology modeling using the 1XJR crystal structure proved to be inconsistent with reported SARS-CoV-2 s2m NMR data and should not be considered suitable for interpretation of SARS-CoV-2 s2m. Fundamentally, our simulations of SARS-CoV-2 s2m using NMR NOE assignments and knowledge-based methods establish that careful selection of initial coordinates is a requirement for modeling the s2m element. Furthermore, our SARS-CoV-2 s2m simulations not only are in agreement with the NMR secondary structure but also provide valuable dynamical knowledge and atomistic structures that can be used to guide understanding the plausible biological function of s2m. We speculate that the simulated differences in s2m structure and dynamics between SARS-CoV and SARS-CoV-2 described from our analysis explain differences in our experimental dimerization results.[11]

Entropy in the S2M

Internal bulges and loops are an important class of RNA secondary structural features found to impact the ability of a hairpin structure to form a kissing dimer or extended duplex and form protein binding recognition motifs, as reported for HIV-1 DIS.[100,101] The s2m element is a hairpin with varying stem content comparable to DIS. Specifically, for SARS-CoV-2, 3′-UTR hairpin bulges have been investigated for druggability.[16] Implicit to these structural features are the entropic costs associated with reducing helical content in a hairpin or adopting a larger terminal loop.[102,103] Multiple models of RNA verify that different secondary structures and topologies affect entropy.[102,104,105] Having characterized the structure and dynamics of each model s2m, the data suggest a difference in disorder and flexibility between SARS-CoV and SARS-CoV-2 s2m. Entropy is the thermodynamic natural variable associated with these qualities, but direct calculation of entropy for NPT ensembles with correlated, many-body dynamics is costly to compute.[106] Andricioaei and Karplus have presented a straightforward and successful quasiharmonic approximation of entropy for biomolecular systems.[96,97] Briefly, using the eigenvalues of the mass-weighted covariance matrix, vibrational frequencies were obtained to estimate the absolute entropy of each oligomer. Like PCA, we considered different indices of selectivity to obtain a more detailed view of the entropic consequences of the mutations in each s2m.

Terminal Loop Entropy. Since our main interest is in the terminal loop and palindromic sequence, we estimated absolute entropy in residues 17 to 27 for each model (Table 1). Although it was necessary to expand our index of selectivity

Table 1. Estimated Absolute Entropy of the s2m Using the Quasiharmonic Approximation.[a]

oligomer	terminal loop region relative entropy [kcal mol^{-1} K^{-1}]	palindrome relative entropy [kcal mol^{-1} K^{-1}]	entire s2m relative entropy [kcal mol^{-1} K^{-1}]
SARS-CoV	0.000	0.000	0.000
SARS-CoV-2	0.125	0.045	0.314

[a]Based on data from 17500 frames (3.5 μs) of MD simulation, systems simulated at 310 K. Terminal loop: all atoms in residues 17 to 27 were used. Palindrome: all atoms in residues 20 to 23 were used.

in SARS-CoV beyond the pentaloop to enable comparison with the nonaloop, the relative entropy of its sequence is the lowest measured. This finding is consistent with structural and dynamical analyses revealing increased stem content and rigid base stacking. In contrast, the nonaloop is far more entropic than the corresponding sequence in SARS-CoV s2m, translating to a free energy contribution of approximately 38 kcal/mol at 310 K representing a significant energy penalty for dimerization compared to SARS-CoV s2m.

Palindromic Sequence Entropy. To assess directly possible entropic effects on the formation of kissing dimer base pairs, we limited the index of selectivity to only the palindromic sequence, residues 20 through 23 (Table 1). SARS-CoV was again found to be the lowest in entropy, distinguished energetically from the SARS-CoV-2 solvent-exposed palindrome by 14 kcal/mol at 310 K due to entropy. From these results, we would therefore expect SARS-CoV to form Watson—Crick palindrome base pairing more spontaneously, while SARS-CoV-2 faces a relatively steep entropic penalty—hindering homodimerization. This is consistent with our experimental dimerization findings.[11]

Entire s2m Entropy. The last index of selectivity investigated is the entire s2m (Table 1), which includes the L-shape and tertiary motifs. As discussed previously, entropic effects from internal bulges and loops have been shown to affect formation of kissing dimers and extended duplex conformations in other viruses. Since we know from a combination of experimental and computational evidence that the s2m secondary structure varies between viruses, it is worthwhile to investigate those differences in stem content along with the overall L-shape kink, and local tertiary motifs explain experimentally determined duplexation patterns.[100,101] The SARS-CoV s2m again has the lowest estimated entropy (Table 1). This is again consistent with our structural and dynamical analyses above, which broadly characterize the SARS-CoV s2m as relatively rigid and high in stem content. The entire structure of the SARS-CoV-2 s2m is the highest in entropy, corresponding to modes found from PCA involving decreased stem content, L-shape kink tertiary interaction dynamics, and fraying in the lower stem region.

According to the experimental homodimerization results, SARS-CoV-2 s2m forms more extended duplex compared to SARS-CoV in the absence of the N-protein chaperone.[11] The relatively high entropy inherent to the structure of SARS-CoV-2 determined in this work supports our conjecture that decreased stem content, L-shape kink tertiary interaction

dynamics, and fraying in the lower stem region facilitates base melting necessary for spontaneous extended duplex formation.

CONCLUSIONS

Our simulation work shows structural, dynamical, and entropic differences between SARS-CoV and SARS-CoV-2 s2m. Simulation of SARS-CoV-2 s2m from PDB 1XJR, revealed that the homology modeling approach is insufficient for capturing the physiologically relevant structures known by NMR at time scale simulated. As such, we employed a knowledge-based method which retained NOE assignments, placing the SARS-CoV-2 s2m into a more appropriate location in phase-space for experimental comparison. The overall dynamics of the SARS-CoV s2m model retained the crystallographic structure with moderate fraying on the 3′ and 5′ ends. However, simulations of SARS-CoV-2 resulted in the greatest amount of fraying, contributing to the disruption of tertiary and three-dimensional structure in the lower stem and bulge. The overall shape of SARS-CoV and SARS-CoV-2 were each computed to have a distinct L-shape kink through entirely different tertiary interactions. Structural, dynamical, and thermodynamic differences between SARS-CoV and SARS-CoV-2 result from the register shift and terminal loop expansion from the GNRA-pentaloop to the nonaloop. As a baseline, the terminal loop of SARS-CoV was found to be rigid and structurally preorganized where the palindromic G20, A22, and C23 nucleotides are involved in persistent base stacking interactions aligning the bases for possible kissing complex dimerization. U21 is unique in both crystallography and simulations having high flexibility and displaced from base stacking interactions. SARS-CoV-2 was found to sample a greater ensemble of structures within the terminal loop characterized by base stacking reorganization with backbone deformations. Consequently, both the flexibility and structural positions of the palindromic bases are altered from the highly organized placement from that found in SARS-CoV. The simulation results provide a rationalization of the experimental homodimerization results. First, the computed predisposed alignment and rigidity of the palindromic sequence coupled with the greater stem content suggests that SARS-CoV should convert to the kissing dimer complex, as found by experiment, where SARS-CoV forms the kissing complex exclusively. Second, the expansion of the terminal loop results in a palindromic sequence that is computed to be highly flexible and structurally diverse discouraging conversion to the homodimer. The interpretation is consistent with experiment, where the PAGE results show that SARS-CoV-2 s2m remains mostly as a monomer with small amounts of kissing complex and duplex formation. Our work provides the foundation for future studies on the mechanism of homodimerization in SARS-CoV-2 and variants, building a foundation to bridge a connection between structure and function. Ultimately, our study establishes the atomistic three-dimensional structure and uncovers dynamic differences that arise from s2m sequence changes in the SARS-CoV-2 virus.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphyschemau.2c00032.

Additional structural and dynamical data, principal component analysis, simulation details, and methods,

including details about the estimation of absolute entropy (PDF)

AUTHOR INFORMATION

Corresponding Author

Jeffrey D. Evanseck − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States;* ⓞ orcid.org/0000-0003-2616-7354; Phone: (412) 396-6337; Email: evanseck@duq.edu

Authors

Adam H. Kensinger − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States;* ⓞ orcid.org/0000-0002-1931-9703

Joseph A. Makowski − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States;* ⓞ orcid.org/0000-0002-2597-8211

Kendy A. Pellegrene − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States;* Present Address: LifeSci Communications, New York City, NY, 10019

Joshua A. Imperatore − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States;* Present Address: National Institute of Standards and Technology, Gaithersburg, MD, 20899

Caylee L. Cunningham − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States*

Caleb J. Frye − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States*

Patrick E. Lackey − *Department of Biochemistry and Chemistry, Westminster College, New Wilmington, Pennsylvania 16172, United States;* ⓞ orcid.org/0000-0003-2528-1147

Mihaela Rita Mihailescu − *Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsphyschemau.2c00032

following undergraduates for participation in related aspects of this work to be eventually published: Morgan Shine (REU), Izayah Bojanac (REU), Nicole Mackenstein (REU), and Bryan Kelleher.

■ REFERENCES

(1) Bettini, E.; Locci, M. SARS-CoV-2 MRNA Vaccines: Immunological Mechanism and Beyond. *Vaccines (Basel)* **2021**, *9* (2), 147.

(2) Ahmed, T. Immunotherapy for Neuroblastoma Using MRNA Vaccines. *Advances in Cancer Biology - Metastasis* **2022**, *4*, 100033.

(3) Kis, Z.; Tak, K.; Ibrahim, D.; Papathanasiou, M. M.; Chachuat, B.; Shah, N.; Kontoravdi, C. Pandemic-Response Adenoviral Vector and RNA Vaccine Manufacturing. *NPJ. Vaccines* **2022**, *7* (1), 29.

(4) Fang, E.; Liu, X.; Li, M.; Zhang, Z.; Song, L.; Zhu, B.; Wu, X.; Liu, J.; Zhao, D.; Li, Y. Advances in COVID-19 MRNA Vaccine Development. *Signal Transduction and Targeted Therapy* **2022**, *7* (1), 1−31.

(5) Rele, S. COVID-19 Vaccine Development during Pandemic: Gap Analysis, Opportunities, and Impact on Future Emerging Infectious Disease Development Strategies. *Human Vaccines & Immunotherapeutics* **2021**, *17* (4), 1122−1127.

(6) *WHO Coronavirus (COVID-19) Dashboard*. https://covid19.who.int/ (accessed 2021-10-26).

(7) Xu, S.; Ding, D.; Zhang, X.; Sun, L.; Kang, D.; Huang, B.; Liu, X.; Zhan, P. Newly Emerging Strategies in Antiviral Drug Discovery: Dedicated to Prof. Dr. Erik De Clercq on Occasion of His 80th Anniversary. *Molecules* **2022**, *27*, 850.

(8) Fischer, W. A.; Eron, J. J.; Holman, W.; Cohen, M. S.; Fang, L.; Szewczyk, L. J.; Sheahan, T. P.; Baric, R.; Mollan, K. R.; Wolfe, C. R.; Duke, E. R.; Azizad, M. M.; Borroto-Esoda, K.; Wohl, D. A.; Coombs, R. W.; James Loftis, A.; Alabanza, P.; Lipansky, F.; Painter, W. P. A Phase 2a Clinical Trial of Molnupiravir in Patients with COVID-19 Shows Accelerated SARS-CoV-2 RNA Clearance and Elimination of Infectious Virus. *Science Translational Medicine* **2022**, *14* (628), 7430.

(9) Li, M.; Lou, F.; Fan, H. SARS-CoV-2 Variant Omicron: Currently the Most Complete "Escapee" from Neutralization by Antibodies and Vaccines. *Signal Transduction and Targeted Therapy* **2022**, *7* (1), 28.

(10) Robertson, M. P.; Igel, H.; Baertsch, R.; Haussler, D.; Ares, M.; Scott, W. G. The Structure of a Rigorously Conserved RNA Element within the SARS Virus Genome. *PLoS Biology* **2004**, *3* (1), e5.

(11) Imperatore, J. A.; Cunningham, C. L.; Pellegrene, K. A.; Brinson, R. G.; Marino, J. P.; Evanseck, J. D.; Mihailescu, M. R. Highly Conserved S2m Element of SARS-CoV-2 Dimerizes via a Kissing Complex and Interacts with Host MiRNA-1307−3p. *Nucleic Acids Res.* **2022**, *50* (2), 1017−1032.

(12) Gilbert, C.; Tengs, T. No Species-Level Losses of S2m Suggests Critical Role in Replication of SARS-Related Coronaviruses. *Sci. Rep.* **2021**, *11* (1), 1−5.

(13) Lulla, V.; Wandel, M. P.; Bandyra, K. J.; Ulferts, R.; Wu, M.; Dendooven, T.; Yang, X.; Doyle, N.; Oerum, S.; Beale, R.; O'Rourke, S. M.; Randow, F.; Maier, H. J.; Scott, W.; Ding, Y.; Firth, A. E.; Bloznelyte, K.; Luisi, B. F. Targeting the Conserved Stem Loop 2 Motif in the SARS-CoV-2 Genome. *J. Virol.* **2021**, *95* (14), 1−21.

(14) Li, J.; Sczepanski, J. T. Targeting a Conserved Structural Element from the SARS-CoV-2 Genome Using l-DNA Aptamers. *RSC Chemical Biology* **2022**, *3* (1), 79−84.

(15) Rangan, R.; Watkins, A. M.; Chacon, J.; Kretsch, R.; Kladwang, W.; Zheludev, I. N.; Townley, J.; Rynge, M.; Thain, G.; Das, R. De Novo 3D Models of SARS-CoV-2 RNA Elements from Consensus Experimental Secondary Structures. *Nucleic Acids Res.* **2021**, *49* (6), 3092−3108.

(16) Sreeramulu, S.; Richter, C.; Berg, H.; Wirtz Martin, M. A.; Ceylan, B.; Matzel, T.; Adam, J.; Altincekic, N.; Azzaoui, K.; Bains, J. K.; Blommers, M. J. J.; Ferner, J.; Fürtig, B.; Göbel, M.; Grün, J. T.; Hengesbach, M.; Hohmann, K. F.; Hymon, D.; Knezic, B.; Martins, J. N.; Mertinkus, K. R.; Niesteruk, A.; Peter, S. A.; Pyper, D. J.; Qureshi,

N. S.; Scheffer, U.; Schlundt, A.; Schnieders, R.; Stirnal, E.; Sudakov, A.; Tröster, A.; Vögele, J.; Wacker, A.; Weigand, J. E.; Wirmer-Bartoschek, J.; Wöhnert, J.; Schwalbe, H. Exploring the Druggability of Conserved RNA Regulatory Elements in the SARS-CoV-2 Genome. *Angewandte Chemie - International Edition* **2021**, *60* (35), 19191−19200.

(17) Gilbert, C.; Tengs, T. No Species-Level Losses of S2m Suggests Critical Role in Replication of SARS-Related Coronaviruses. *Sci. Rep.* **2021**, *11* (1), 1−5.

(18) Tengs, T.; Delwiche, C. F.; Monceyron Jonassen, C. A Genetic Element in the SARS-CoV-2 Genome Is Shared with Multiple Insect Species. *Journal of General Virology* **2021**, *102* (3), 1−5.

(19) Tengs, T.; Kristoffersen, A. B.; Bachvaroff, T. R.; Jonassen, C. M. A Mobile Genetic Element with Unknown Function Found in Distantly Related Viruses. *Virology Journal* **2013**, *10* (132), 1−9.

(20) Tengs, T.; Jonassen, C. Distribution and Evolutionary History of the Mobile Genetic Element S2m in Coronaviruses. *Diseases* **2016**, *4* (4), 27.

(21) Frye, C. J.; Shine, M.; Makowski, J. A.; Kensinger, A. H.; Cunningham, C. L.; Milback, E. J.; Evanseck, J. D.; Lackey, P. E.; Mihailescu, M. R. Bioinformatics Analysis of the S2m Mutations within the SARS-CoV-2 Omicron Lineages. *Journal of Medical Virology* **2022**, x.

(22) Hegde, S.; Tang, Z.; Zhao, J.; Wang, J. Inhibition of SARS-CoV-2 by Targeting Conserved Viral RNA Structures and Sequences. *Frontiers in Chemistry* **2021**, *9*, 1122.

(23) Ryder, S. P.; Morgan, B. R.; Coskun, P.; Antkowiak, K.; Massi, F. Analysis of Emerging Variants in Structured Regions of the SARS-CoV-2 Genome. *Evolutionary Bioinformatics* **2021**, *17*, 1−18.

(24) Ziv, O.; Price, J.; Shalamova, L.; Kamenova, T.; Goodfellow, I.; Weber, F.; Miska, E. A. The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell* **2020**, *80* (6), 1067−1077.e5.

(25) Shetty, S.; Kim, S.; Shimakami, T.; Lemon, S. M.; Mihailescu, M. R. Hepatitis C Virus Genomic RNA Dimerization Is Mediated via a Kissing Complex Intermediate. *RNA* **2010**, *16* (5), 913−925.

(26) Mihailescu, M.-R.; Marino, J. P. A Proton-Coupled Dynamic Conformational Switch in the HIV-1 Dimerization Initiation Site Kissing Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (5), 1189−1194.

(27) Horiya, S.; Li, X.; Kawai, G.; Saito, R.; Katoh, A.; Kobayashi, K.; Harada, K. RNA LEGO: Magnesium-Dependent Formation of Specific RNA Assemblies through Kissing Interactions. *Chemical Biology* **2003**, *10* (7), 645−654.

(28) Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; Icenogle, J. P.; Peñaranda, S.; Bankamp, B.; Maher, K.; Chen, M.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J. L.; Chen, Q.; Wang, D.; Erdman, D. D.; Peret, T. C. T.; Burns, C.; Ksiazek, T. G.; Rollin, P. E.; Sanchez, A.; Liffick, S.; Holloway, B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Günther, S.; Osterhaus, A. D. M. E.; Drosten, C.; Pallansch, M. A.; Anderson, L. J.; Bellini, W. J. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* **2003**, *300* (5624), 1394−1399.

(29) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579* (7798), 265−269.

(30) Dyer, O. Two Strains of the SARS Virus Sequenced. *BMJ.* **2003**, *326* (7397), 999−999.

(31) Sloma, M. F.; Mathews, D. H. *Improving RNA Secondary Structure Prediction with Structure Mapping Data*, 1st ed.; Elsevier Inc.: 2015; Vol. 553. DOI: 10.1016/bs.mie.2014.10.053.

(32) Kladwang, W.; Vanlang, C. C.; Cordero, P.; Das, R. Understanding the Errors of SHAPE-Directed RNA Structure Modeling. *Biochemistry* **2011**, *50* (37), 8049−8056.

(33) Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T. K.; Marinus, T.; Ogando, N. S.; Snijder, E. J.; van Hemert, M. J.; Bujnicki, J. M.; Incarnato, D. Genome-Wide Mapping of SARS-

CoV-2 RNA Structures Identifies Therapeutically-Relevant Elements. *Nucleic Acids Res.* **2020**, *48* (22), 12436−12452.

(34) Ahmed, F.; Sharma, M.; Al-Ghamdi, A. A.; Al-Yami, S. M.; Al-Salami, A. M.; Refai, M. Y.; Warsi, M. K.; Howladar, S. M.; Baeshen, M. N. A Comprehensive Analysis of Cis-Acting RNA Elements in the SARS-CoV-2 Genome by a Bioinformatics Approach. *Frontiers in Genetics* **2020**, *11*, 1−18.

(35) Aldhumani, A. H.; Hossain, M. I.; Fairchild, E. A.; Boesger, H.; Marino, E. C.; Myers, M.; Hines, J. v. RNA Sequence and Ligand Binding Alter Conformational Profile of SARS-CoV-2 Stem Loop II Motif. *Biochem. Biophys. Res. Commun.* **2021**, *545*, 75−80.

(36) Huston, N. C.; Wan, H.; Strine, M. S.; de Cesaris Araujo Tavares, R.; Wilen, C. B.; Pyle, A. M. Comprehensive in Vivo Secondary Structure of the SARS-CoV-2 Genome Reveals Novel Regulatory Motifs and Mechanisms. *Mol. Cell* **2021**, *81* (3), 584−598.e5.

(37) Morandi, E.; Manfredonia, I.; Simon, L. M.; Anselmi, F.; van Hemert, M. J.; Oliviero, S.; Incarnato, D. Genome-Scale Deconvolution of RNA Structure Ensembles. *Nat. Methods* **2021**, *18* (3), 249−252.

(38) Sun, L.; Li, P.; Ju, X.; Rao, J.; Huang, W.; Ren, L.; Zhang, S.; Xiong, T.; Xu, K.; Zhou, X.; Gong, M.; Miska, E.; Ding, Q.; Wang, J.; Zhang, Q. C. In Vivo Structural Characterization of the SARS-CoV-2 RNA Genome Identifies Host Proteins Vulnerable to Repurposed Drugs. *Cell* **2021**, *184* (7), 1865−1883.e20.

(39) Cao, C.; Cai, Z.; Xiao, X.; Rao, J.; Chen, J.; Hu, N.; Yang, M.; Xing, X.; Wang, Y.; Li, M.; Zhou, B.; Wang, X.; Wang, J.; Xue, Y. The Architecture of the SARS-CoV-2 RNA Genome inside Virion. *Nature Communications* **2021**, *12* (1), 1−14.

(40) Zhang, Y.; Huang, K.; Xie, D.; Lau, J. Y.; Shen, W.; Li, P.; Wang, D.; Zou, Z.; Shi, S.; Ren, H.; Wang, Y.; Mao, Y.; Jin, M.; Kudla, G.; Zhao, Z. In Vivo Structure and Dynamics of the SARS-CoV-2 RNA Genome. *Nat. Commun.* **2021**, *12* (1), 5695.

(41) Wacker, A.; Weigand, J. E.; Akabayov, S. R.; Altincekic, N.; Bains, J. K.; Banijamali, E.; Binas, O.; Castillo-Martinez, J.; Cetiner, E.; Ceylan, B.; Chiu, L.; Davila-Calderon, J.; Dhamotharan, K.; Duchardt-Ferner, E.; Ferner, J.; Frydman, L.; Fürtig, B.; Gallego, J.; Grün, J. T.; Hacker, C.; Haddad, C.; Hähnke, M.; Hengesbach, M.; Hiller, F.; Hohmann, K. F.; Hymon, D.; de Jesus, V.; Jonker, H.; Keller, H.; Knezic, B.; Landgraf, T.; Löhr, F.; Luo, L.; Mertinkus, K. R.; Muhs, C.; Novakovic, M.; Oxenfarth, A.; Palomino-Schätzlein, M.; Petzold, K.; Peter, S. A.; Pyper, D. J.; Qureshi, N. S.; Riad, M.; Richter, C.; Saxena, K.; Schamber, T.; Scherf, T.; Schlagnitweit, J.; Schlundt, A.; Schnieders, R.; Schwalbe, H.; Simba-Lahuasi, A.; Sreeramulu, S.; Stirnal, E.; Sudakov, A.; Tants, J.; Tolbert, B. S.; Vögele, J.; Weiß, L.; Wirmer-Bartoschek, J.; Wirtz Martin, M. A.; Wöhnert, J.; Zetzsche, H. Secondary Structure Determination of Conserved SARS-CoV-2 RNA Elements by NMR Spectroscopy. *Nucleic Acids Res.* **2020**, *48* (22), 12415−12435.

(42) Watkins, A. M.; Rangan, R.; Das, R. FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure* **2020**, *28* (8), 963−976.e6.

(43) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: A Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction. *Nucleic Acids Res.* **2016**, *44* (7), e63.

(44) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. from Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5* (10), 1771−1782.

(45) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to Understand Atomistic Molecular Dynamics Simulations of RNA and Protein−RNA Complexes? *Wiley Interdisciplinary Reviews RNA* **2017**, *8* (3), 1405.

(46) Ditzler, M. A.; Otyepka, M.; Šponer, J.; Walter, N. G. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Acc. Chem. Res.* **2010**, *43* (1), 40−47.

(47) Hashem, Y.; Auffinger, P. A Short Guide for Molecular Dynamics Simulations of RNA Systems. *Methods* **2009**, *47* (3), 187−197.

(48) Biesiada, M.; Pachulska-Wieczorek, K.; Adamiak, R. W.; Purzycka, K. J. RNAComposer and RNA 3D Structure Prediction for Nanotechnology. *Methods* **2016**, *103*, 120−127.

(49) 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.

(50) Al-Hashimi, H. M.; Walter, N. G. RNA Dynamics: It Is about Time. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 321−329.

(51) Rinnenthal, J.; Buck, J.; Ferner, J.; Wacker, A.; Fürtig, B.; Schwalbe, H. Mapping the Landscape of RNA Dynamics with NMR Spectroscopy. *Acc. Chem. Res.* **2011**, *44* (12), 1292−1301.

(52) Watters, K. E.; Strobel, E. J.; Yu, A. M.; Lis, J. T.; Lucks, J. B. Cotranscriptional Folding of a Riboswitch at Nucleotide Resolution. *Nature Structural & Molecular Biology* **2016**, *23* (12), 1124−1131.

(53) Mustoe, A. M.; Brooks, C. L.; Al-Hashimi, H. M. Hierarchy of RNA Functional Dynamics. *Annu. Rev. Biochem.* **2014**, *83* (1), 441−466.

(54) Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. Protein States and Proteinquakes. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5000−5004.

(55) Dror, R. O.; Jensen, M.; Borhani, D. W.; Shaw, D. E. Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations. *J. Gen. Physiol.* **2010**, *135* (6), 555−562.

(56) Wang, L.; O'Mara, M. L. Effect of the Force Field on Molecular Dynamics Simulations of the Multidrug Efflux Protein P-Glycoprotein. *J. Chem. Theory Comput.* **2021**, *17* (10), 6491−6508.

(57) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruziero, 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.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovski, O.; 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, R. C.; Walker, J.; Wilson, W. L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. *AmberTools20*; University of California: San Francisco, CA, 2020.

(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(59) Cornell, W. D.; Cieplak, P.; Bayly, C. L.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(60) 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.

(61) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.

(62) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: 2017; Vol. *1*. DOI: 10.1093/oso/9780198803195.001.0001.

(63) Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. Generalized Manning Condensation Model Captures the RNA Ion Atmosphere. *Phys. Rev. Lett.* **2015**, *114* (25), 1−6.

(64) Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. A Repulsive Field: Advances in the Electrostatics of the Ion Atmosphere. *Current Opinions in Chemical Biology* **2008**, *12* (6), 619−625.

(65) Jacobson, D. R.; Saleh, O. A. Quantifying the Ion Atmosphere of Unfolded, Single-Stranded Nucleic Acids Using Equilibrium Dialysis and Single-Molecule Methods. *Nucleic Acids Res.* **2016**, *44* (8), 3763−3771.

(66) Krepl, M.; Vögele, J.; Kruse, H.; Duchardt-Ferner, E.; Wöhnert, J.; Sponer, J. An Intricate Balance of Hydrogen Bonding, Ion Atmosphere and Dynamics Facilitates a Seamless Uracil to Cytosine Substitution in the U-Turn of the Neomycin-Sensing Riboswitch. *Nucleic Acids Res.* **2018**, *46* (13), 6528−6543.

(67) Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. Generalized Manning Condensation Model Captures the RNA Ion Atmosphere. *Phys. Rev. Lett.* **2015**, *114* (25), 1−6.

(68) Leipply, D.; Lambert, D.; Draper, D. E.*Ion−RNA Interactions*, 1st ed.; Elsevier Inc.: 2009; Vol. *469*. DOI: 10.1016/s0076-6879(09)69021-2.

(69) Manning, G. S. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions. 7. Electrophoretic Mobility and Conductance. *J. Phys. Chem.* **1981**, *85* (11), 1506−1515.

(70) Woodson, S. A. Compact Intermediates in RNA Folding. *Annual Review of Biophysics* **2010**, *39* (1), 61−77.

(71) Beššeová, I.; Banáš, P.; Kührová, P.; Košinová, P.; Otyepka, M.; Sponer, J. Simulations of A-RNA Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J. Phys. Chem. B* **2012**, *116* (33), 9899−9916.

(72) Pinamonti, G.; Bottaro, S.; Micheletti, C.; Bussi, G. Elastic Network Models for RNA: A Comparative Assessment with Molecular Dynamics and SHAPE Experiments. *Nucleic Acids Res.* **2015**, *43* (15), 7260−7269.

(73) Kognole, A. A.; MacKerell, A. D. Contributions and Competition of Mg 2+ and K + in Folding and Stabilization of the Twister Ribozyme. *RNA* **2020**, *26* (11), 1704−1715.

(74) Fischer, N. M.; Polěto, M. D.; Steuer, J.; van der Spoel, D. Influence of Na+ and Mg2+ Ions on RNA Structures Studied with Molecular Dynamics Simulations. *Nucleic Acids Res.* **2018**, *46* (10), 4872−4882.

(75) Ennifar, E.; Dumas, P. Polymorphism of Bulged-out Residues in HIV-1 RNA DIS Kissing Complex and Structure Comparison with Solution Studies. *J. Mol. Biol.* **2006**, *356* (3), 771−782.

(76) Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal Structures of Coaxially Stacked Kissing Complexes of the HIV-1 RNA Dimerization Initiation Site. *Nature Structural & Molecular Biology* **2001**, *8* (12), 1064−1068.

(77) Mundigala, H.; Michaux, J. B.; Feig, A. L.; Ennifar, E.; Rueda, D. HIV-1 DIS Stem Loop Forms an Obligatory Bent Kissing Intermediate in the Dimerization Pathway. *Nucleic Acids Res.* **2014**, *42* (11), 7281−7289.

(78) Auffinger, P.; Bielecki, L.; Westhof, E. Anion Binding to Nucleic Acids. *Structure* **2004**, *12* (3), 379−388.

(79) Fiala, R.; Špačková, N.; Foldynová-Trantírková, S.; Šponer, J.; Sklenář, V.; Trantírek, L. NMR Cross-Correlated Relaxation Rates Reveal Ion Coordination Sites in DNA. *J. Am. Chem. Soc.* **2011**, *133* (35), 13790−13793.

(80) Kührová, P.; Otyepka, M.; Šponer, J.; Banáš, P. Are Waters around RNA More than Just a Solvent? − An Insight from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10* (1), 401−411.

(81) Hurwitz, C.; Rosano, C. L. The Intracellular Concentration of Bound and Unbound Magnesium Ions in Escherichia Coli. *J. Biol. Chem.* **1967**, *242* (16), 3719−3722.

(82) Record, M. T.; Anderson, C. F.; Lohman, T. M. Thermodynamic Analysis of Ion Effects on the Binding and Conformational Equilibria of Proteins and Nucleic Acids: The Roles of Ion Association or Release, Screening, and Ion Effects on Water Activity. *Q. Rev. Biophys.* **1978**, *11* (2), 103−178.

(83) Rook, M. S.; Treiber, D. K.; Williamson, J. R. An Optimal Mg 2+ Concentration for Kinetic Folding of the Tetrahymena Ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (22), 12471−12476.

(84) Réblová, K.; Špačková, N.; Koča, J.; Leontis, N. B.; Šponer, J. Long-Residency Hydration, Cation Binding, and Dynamics of Loop E/Helix IV RRNA-L25 Protein Complex. *Biophys. J.* **2004**, *87* (5), 3397−3412.

(85) Sklenovský, P.; Florová, P.; Banáš, P.; Réblová, K.; Lankaš, F.; Otyepka, M.; Šponer, J. Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *J. Chem. Theory Comput.* **2011**, *7* (9), 2963−2980.

(86) Florová, P.; Sklenovský, P.; Banáš, P.; Otyepka, M. Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact. *J. Chem. Theory Comput.* **2010**, *6* (11), 3569−3579.

(87) Sengul, M. Y.; MacKerell, A. D. Accurate Modeling of RNA Hairpins Through the Explicit Treatment of Electronic Polarizability with the Classical Drude Oscillator Force Field. *Journal of Computational Biophysics and Chemistry* **2022**, *21* (04), 461−471.

(88) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836−3849.

(89) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics and Modelling* **1996**, *14* (1), 33−38.

(90) Górska, A.; Jasiński, M.; Trylska, J. MINT: Software to Identify Motifs and Short-Range Interactions in Trajectories of Nucleic Acids. *Nucleic Acids Res.* **2015**, *43* (17), e114.

(91) Li, S.; Olson, W. K.; Lu, X. J. Web 3DNA 2.0 for the Analysis, Visualization, and Modeling of 3D Nucleic Acid Structures. *Nucleic Acids Res.* **2019**, *47* (W1), W26−W34.

(92) Kerpedjiev, P.; Hammer, S.; Hofacker, I. L. Forna (Force-Directed RNA): Simple and Effective Online RNA Secondary Structure Diagrams. *Bioinformatics* **2015**, *31* (20), 3377−3379.

(93) Stein, S. A. M.; Loccisano, A. E.; Firestine, S. M.; Evanseck, J. D.Chapter 13 Principal Components Analysis: A Review of Its Application on Molecular Dynamics Data. In *Annual Reports in Computational Chemistry*; Pittsburgh, PA, 2006; Vol. *2*, pp 233−261. DOI: 10.1016/S1574-1400(06)02013-5.

(94) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412−425.

(95) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193−226.

(96) Schlitter, J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance Matrix. *Chemical Physical Letters* **1993**, *215* (6), 617−621.

(97) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115* (14), 6289−6292.

(98) Klyne, W.; Prelog, V. Description of Steric Relationships across Single Bonds. *Experientia* **1960**, *16* (12), 521−523.

(99) Rivas, E. Evolutionary Conservation of RNA Sequence and Structure. *Wiley Interdisciplinary Reviews RNA* **2021**, *12* (5), 1649.

(100) Mujeeb, A.; Ulyanov, N. B.; Georgantis, S.; Smirnov, I.; Chung, J.; Parslow, T. G.; James, T. L. Nucleocapsid Protein-Mediated Maturation of Dimer Initiation Complex of Full-Length SL1 Stemloop of HIV-1: Sequence Effects and Mechanism of RNA Refolding. *Nucleic Acids Res.* **2007**, *35* (6), 2026−2034.

(101) Hermann, T.; Patel, D. J. RNA Bulges as Architectural and Recognition Motifs. *Structure* **2000**, *8* (3), R47−R54.

(102) Cao, S.; Chen, S.-J. Predicting RNA Folding Thermodynamics with a Reduced Chain Representation Model. *RNA* **2005**, *11* (12), 1884−1897.

(103) Cao, S.; Chen, S.-J. Structure and Stability of RNA/RNA Kissing Complex: With Application to HIV Dimerization Initiation Signal. *RNA* **2011**, *17* (12), 2130−2143.

(104) Zoli, M. Entropic Penalties in Circular DNA Assembly. *J. Chem. Phys.* **2014**, *141* (17), 174112.

(105) Mak, C. H.; Phan, E. N. H. Topological Constraints and Their Conformational Entropic Penalties on RNA Folds. *Biophys. J.* **2018**, *114* (9), 2059−2071.

(106) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. Estimating Entropies from Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *120* (6), 2652−2661.