Supporting Information

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

Adam H. Kensinger‡1, Joseph A. Makowski‡1, Kendy A. Pellegrine‡1, Joshua A. Imperatore1, Caylee L. Cunningham1, Caleb J. Frye1, Patrick E. Lackey2, Mihaela Rita Mihaiescu1, and Jeffrey D. Evanseck1*

1Department of Chemistry and Biochemistry and Center for Computational Sciences, Duquesne University, Pittsburgh, PA, 15282, USA
2Department of Biochemistry and Chemistry, Westminster College, New Wilmington, PA, 16172, USA

*Email: evanseck@duq.edu
Figure S1. Consideration of ion condensation onto the RNA hairpin for SARS-CoV and SARS-CoV-2 s2m simulations at different concentrations of Mg$^{2+}$ and Na$^+$. (A) RMSD with respect to the first frame of the SARS-CoV s2m systems with 1, 2, and 13 Mg$^{2+}$ indicating that condensation of more Mg$^{2+}$ reduces the overall structural deviation compared to condensation of mostly Na$^+$ ions. (B) RMSF of the SARS-CoV s2m systems showing common flexible residues, but overall dampened dynamics in the 13 Mg$^{2+}$ system. (C) SARS-CoV s2m isosurface maps of the condensed counterion density calculated with an 95% isovalue averaged over 17500 trajectory frames for 3.5 µs simulations and 7500 frames for 1.5µ simulations. (D) SARS-CoV-2 s2m simulation with 1 Mg$^{2+}$ isosurface map calculated similarly to those in (C).

Figure S2. Backbone heavy atom SARS-CoV s2m RMSD at 310 K (black) and 283 K (gray) for 3.5 µs calculated with reference to the first frame of each system.
Figure S3. Average RMSF for SARS-CoV s2m simulations at 283 K and 310K for each motif reported by Robertson et al.\textsuperscript{13}

Figure S4: Heavy atom representation of SARS-CoV s2m GNRA-like pentaloop. Dashed lines (yellow) indicate stacking interactions from the center of mass of the base.

| Nucleotide | Mean $\chi$ angle [$^\circ$] | 1XJR $\chi$ angle [$^\circ$] |
|------------|-----------------------------|-----------------------------|
| G18        | 212 ± 8.9                   | 208                         |
| A19        | 207 ± 20                    | 197                         |
| G20        | 226 ± 13                    | 148                         |
| U21        | 180 ± 84                    | 266                         |
| A22        | 187 ± 9.0                   | 185                         |

Values calculated for 3.5 μs (17500 frames)

Table S2. Stacking pentaloop nucleotide COM distances interaction energies at 310 K.\textsuperscript{b}
| Stack  | Stacking interaction | Mean COM distance [Å] | Total mean stacking energy [kcal/mol] | Stack Occupancy [%] |
|--------|----------------------|-----------------------|---------------------------------------|--------------------|
| a      | G:18|G:20               | 5.77 ± 0.2               | -0.24 ± 0.9         | 99                 |
| b      | A:19|G:20               | 4.31 ± 0.2               | -0.05 ± 3.0         | 99                 |
| c      | G:20|A:22               | 4.29 ± 0.3               | 1.18 ± 1.9          | 99                 |

Values calculated for 3.5 µs (17500 frames)

![Figure S5. GC quartet showing ten hydrogen bonds in SARS-CoV s2m at 310 K. Watson-Crick interactions in blue dashes and sugar edge interactions in red dashes.](image)

| Hydrogen bond | Watson/Crick or sugar edge | Donor     | Acceptor  | Mean Distance [Å] | Occupancy [%] | 1XJR Distance [Å] |
|---------------|---------------------------|-----------|-----------|-------------------|---------------|--------------------|
| a             | W/C                       | C27:N4    | G15:O6    | 2.88 ± 0.1        | 75.6          | 2.88               |
| b             | W/C                       | G15:N1    | C27:N3    | 2.95 ± 0.1        | 94.3          | 2.99               |
| c             | W/C                       | G15:N2    | C27:O2    | 2.92 ± 0.1        | 91.7          | 2.98               |
| d             | W/C                       | G24:N2    | C16:O2    | 2.96 ± 0.1        | 84.4          | 2.94               |
| e             | W/C                       | G24:N1    | C16:N3    | 2.96 ± 0.1        | 86.3          | 2.96               |
| f             | W/C                       | C16:N4    | G24:O6    | 2.85 ± 0.1        | 64.9          | 2.89               |
| g             | sugar edge                | C16:O2'   | G15:N3    | 2.98 ± 0.2        | 54.2          | 2.96               |
| h             | sugar edge                | G15:N2    | C16:O2    | 2.89 ± 0.1        | 79.1          | 2.72               |
| i             | sugar edge                | G24:N2    | C27:O2    | 2.88 ± 0.1        | 85.9          | 2.71               |
| j             | sugar edge                | C27:O2'   | G24:N3    | 3.13 ± 0.3        | 40.6          | 2.87               |

Values calculated for 3.5 µs (17500 frames)
Figure S6. Three-dimensional representation of SARS-CoV s2m backbone (310 K) highlighting three-purine bulge (red), Mg$^{2+}$ tunnel (orange), and seven-nucleotide bubble (brown) with the Mg$^{2+}$ cation (cyan) bound in the tunnel. (Top-right inset) Noncanonical GA base pair of the three-purine bulge. (Bottom-right inset) Long range stacking interaction between A29 of the purine bulge and G9 of the seven-nucleotide bubble.

SARS-CoV (310 K) Purine Bulge: A29N1-G7N2 Interaction

| Donor   | Acceptor | Mean distance [Å] | Occupancy [%] | 1XJR distance [Å] |
|---------|----------|-------------------|---------------|-------------------|
| A13:N6  | G30:O6   | 2.98 ± 0.2        | 79.4          | 2.93              |
| A13:N1  | G30:N1   | 3.01 ± 0.2        | 46.7          | 2.64              |

Values calculated for 3.5 ms (17500 frames)
Figure S8. (A) Backbone heavy atom RMSD of SARS-CoV s2m model with both crystallographic Mg$^{2+}$ ions maintained (yellow) compared to the SARS-CoV s2m simulations with one Mg$^{2+}$ with reference to the first frame of each system. (B) Two-Mg$^{2+}$ ion SARS-CoV s2m model RMSF (yellow) compared to the one Mg$^{2+}$ systems 310 K (black) and 283 K (gray). (C) Aligned centroid structures from two Mg$^{2+}$ system PCA with CS1 in purple and CS2 in yellow. Red ring highlights the transient interaction that forms from the fraying stem.

Figure S9. L-shaped kink angle for SARS-CoV s2m model calculated over the 3.5μs simulations at both 310 K (black) and 283 K (gray).
Figure S10a. Nucleic heavy atom SARS-CoV s2m 310 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and PDB ID: 1XJR as a yellow cross.
Figure S10b. Nucleic heavy atom SARS-CoV s2m 283 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and 1XJR as a yellow cross.
Figure S10c. Nucleic heavy atom SARS-CoV s2m terminal loop 310 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and 1XJR as a yellow cross.
Figure S10d. Nucleic heavy atom SARS-CoV-2 s2m 310 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and NMR-based starting coordinates as a yellow cross.
Figure S10e: Nucleic heavy atom SARS-CoV-2 s2m terminal loop 310 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and NMR-based starting coordinates as a yellow cross.
Figure S10: Nucleic heavy atom SARS-CoV-2 s2m 283 K PCA data. CS centroids are given as red dots, the simulation average structure as a red star, and 1XJR as a yellow cross.
Figure S1: SARS-CoV s2m heavy atom PCA. CS centroids are given as red dots, the simulation average structure as a red star, and 1XJR as a yellow cross. (a) PCA of the 310 K simulation. (b) PCA of the 283 K simulation. (c) Centroid structures of CS1 and CS2 from the 310 K simulation, revealing the main source of structural variation is in moderate lower stem fraying.
Figure S12: (A) Backbone heavy atom RMSD calculated for SARS-CoV-2 s2m homology (black) compared with SARS-CoV s2m (310 K, dark gray; 283 K, light blue) and SARS-CoV-2 s2m (310 K, red; 283 K, light gray). Each RMSD is calculated with respect to the first frame of the individual system. (B) RMSF calculated per nucleotide for SARS-CoV-2 s2m homology (black) compared with SARS-CoV s2m (310 K, dark gray; 283 K, light blue) and SARS-CoV-2 s2m (310 K, red; 283 K, light gray).
### Table S5. Secondary Structure Comparison.

| Base pair | Wacker et al. NMR NOE\(^{14}\) | RNAComposer\(^{15,16}\) | Robertson et al. PDB: 1XJR\(^{13}\) | SARS-CoV-2 Homology |
|-----------|---------------------------------|-------------------------|---------------------------------|---------------------|
| 1         | U1:A41                           |                         | U1:A41                          | –                   |
| 2         | U2:A40                           | U2:A40                  | U2:A40                          | –                   |
| 3         | C3:G39                           | C3:G39                  | C3:G39                          | C3:G39              |
| 4         | A4:U38                           | A4:U38                  | A4:U38                          | A4:U38              |
| 5         | C5:G37                           | C5:G37                  | U5:G37                          | U5:G37              |
| 6         | C12:G32                          | C12:G32                 | C6:A36                          | C6:A36              |
| 7         | A13:U31                          | A13:U31                 | G9:C35                          | G9:C35              |
| 8         | C14:G30                          | C14:G30                 | G10:U33                         | G10:U33             |
| 9         | C16:G28                          | C16:G28                 | C11:G32                         | C11:G32             |
| 10        | G17:C27                          | G17:C27                 | C12:G31                         | C12:G31             |
| 11        |                                 |                         | A13:G30                         | A13:G30             |
| 12        | C14:G28                          | C14:G28                 |                                 |                     |
| 13        |                                 |                         | G15:C27                         | G15:C27             |
| 14        | C16:G24                          | C16:G24                 |                                 |                     |
| 15        |                                 |                         | G17:C23                         | G17:C23             |
| 16        |                                 |                         | G18:A22                         | G18:A22             |

**Figure S13:** Backbone heavy atom SARS-CoV-2 s2m RMSD at 310 K (black) and 283 K (gray) with respect to the first frame of the simulation.
Table S6. web3DNA SARS-CoV-2 s2m CS1 Helical parameters.a

| step | Xp  | Yp  | Zp  | XpH | YpH | ZpH | Form |
|------|-----|-----|-----|-----|-----|-----|------|
| 1    | -1.91 | 7.52 | -1.98 | -6.37 | 5.26 | -5.22 |
| 2    | -1.9  | 8.56 | -1.31 | 6.18  | 7.43  | 4.49  |
| 3    | ---   | ---  | ---  | ---   | ---   | ---   |
| 4    | ---   | ---  | ---  | ---   | ---   | ---   |
| 5    | -1.99 | 8.42 | 0.43  | -2.54 | 8.35  | -1.34 |
| 6    | -2.44 | 9.09 | 1.44  | -3.44 | 9.25  | 0.86  |
| 7    | -2.21 | 7.24 | 2.55  | -16.86| 0.36  | 7.75  |
| 8    | -3.07 | 7.97 | 1.74  | -3.62 | 7.83  | 2.58  |
| 9    | -2.63 | 8.1  | 2.58  | -7.12 | 7.33  | 4.26  | A    |
| 10   | -2.11 | 8.07 | 2.78  | -5.85 | 4.95  | 6.88  | A    |
| 11   | -1.33 | 7.77 | 3.51  | -4.83 | 7.49  | 3.97  |
| Idealb | -1.18 | 8.35 | 2.65  | -5.07 | 7.39  | 4.71  | A    |

aClassification of each dinucleotide step in a right-handed nucleic acid structure: no standard helical form identified. bIdeal A-form helix RNA (polyA-polyU) was generated and analyzed using web3DNA for comparison to the CS1 centroid values. Note that the deformation may be related to the treatment of the RNA ionic environment.

Figure S14. Three-dimensional representation of SARS-CoV-2 s2m backbone (310 K) highlighting long-range interactions (red), Mg$^{2+}$ binding cavity (orange), and base-triple (brown) with the Mg$^{2+}$ ion (cyan) bound in cavity. (Top inset) Triple-interaction at base of upper stem. (Bottom inset) Long range interaction between A29 of the purine bulge and G9 of the seven-nucleotide bubble.
Figure S15. RMSF of SARS-CoV-2 s2m 310 K terminal loop CS. Each CS centroid was used as the reference structure.

Figure S16. Base stacking interactions in the SARS-CoV-2 s2m terminal loop. (A) Stacking interactions as calculated by MINT. Gray nucleotides are either unstacked or in an unstable, positive potential stack; colored nucleotides are in a stabilizing stack. Different colors indicate collections of nucleotides in separate composite stacks. (B) Secondary structure and base stacking interactions determined by Barnaba, base stacks depicted by oriented arrows$^{17}$ and base pairs in Leontis-Westhoff notation.$^{18}$
Supplemental Methods

Multivariate Analysis
The $k$-means clustering algorithm was employed to determine the boundaries of distinct CS within the PCA reduced position data. A distance-based clustering method, $k$-means finds the optimal grouping of data into $k$ disjoint clusters defined by a pre-determined number of $k$ cluster centroids. In the context of MD position data, $k$-means partitions simulation frames into groups defined by structural similarity, where the centroids are a set of representative, average structures which together depict the greatest conformational changes in the simulation. PCA and $k$-means computations were performed using an in-house script, which depends on the Python libraries MDTraj, Scikit-learn, and Matplotlib.

Through iterative minimization of the within-cluster sum of squares (WCSS) criterion, the $k$-means method generally converges to partitioning that minimizes the distance between points assigned to the same cluster. The number of distinct clusters appropriate for each simulated system was determined by visual inspection of the PCA data. In pathological cases where separation into optimal clusters did not match the topology of the point cloud, we elected to increase $k$ to coerce proper partitions.

With respect to the specific structural characteristics that define the greatest variance in biomolecular structure, simulation frames within a conformational substate defined in this manner will generally be more similar to each other than structures from another conformational substate. As such, within any given conformational substate, we expect relatively small variance in the characteristics which, across the whole simulation, do result in great variance. Therefore, by performing PCA again on each conformational substate, a new set of eigenvectors representing different dynamical motions will yield the greatest variance, enabling deeper investigation of the dynamical hierarchy. We performed repeated PCA in this manner to define a dynamical hierarchy within our simulations. This analysis was also facilitated by our in-house script, which discriminates between frames of trajectory files based on conformational substate.

As a distance-based clustering method, some distributions of data may be clustered in unintuitive ways by $k$-means. We observe that there exist scenarios where the centroids obtained through $k$-means do not depict the full range of a dynamical mode, including our 283 K simulation (Figure S17). In pathological cases such as these, we recommend either increasing $k$ until isolated groups of structures are treated as separate substates or employing a different clustering algorithm. For our systems, spectral clustering, which uses a graph Laplacian and accounts for some notion of topology, occasionally results in better clusters than $k$-means alone (Figure S18).

Figure S17: Unexpected $k$-means clustering. We observe that there exist scenarios where the centroids obtained through $k$-means do not depict the full range of a dynamical mode, including our 283 K simulation (see supplementary information)
Figure S18: Comparison of clustering methods over different data. CS centroids are given as red dots, the simulation average structure as a red star, and 1XJR as a yellow cross. (a) \( k \)-means clustering of SARS-CoV s2m 310 K terminal loop PCA data. (b) \( k \)-means clustering of SARS-CoV s2m 283 K terminal loop PCA data. (c) Spectral clustering of SARS-CoV s2m 283 K terminal loop PCA data.

MINT. To aid in measuring structural changes in our trajectories and conformational substate centroids, hydrogen bonding and base stacking interactions were measured using the Motif Identifier for Nucleic Acids (MINT) software. The default configuration was used for both single frame and trajectory analyses, where the hydrogen bond cutoff criterion is 3.5 Å donor-to-donor and 150° displacement and the stacking cutoff is 10.5 Å. MINT uses a forcefield, by default AMBER, to determine quantitatively the potential associated with each stacking interaction. Base stacking analysis with MINT was supplemented with secondary structure graphics from Barnaba, which uses a distance-based method to identify interactions in the framework of Leontis and Westhof.

Estimation of Absolute Entropy. By finding the eigenvalues of the mass-weighted covariance matrix \( C' \),

\[ M^{1/2} C M^{1/2} = C' \]

for which \( M \) is the block matrix of atomic masses on the diagonal and elsewhere 0, one may determine a set of quasiharmonic frequencies \( \omega \) from the eigenvalues \( \lambda \) of \( C' \) by the relation
\[ \omega_i = \sqrt{\frac{kT}{\lambda_i}} \]

which, by the quantum harmonic oscillator partition function, yields the absolute entropy

\[ S_{ho} = k \sum_{i}^{3n-6} \frac{\hbar \omega_i/kT}{e^{\hbar \omega_i/kT} - 1} - \ln(1 - e^{-\hbar \omega_i/kT}) \]

over the \(3n-6\) vibrational modes with nonzero eigenvalues, \(n\) the number of atoms.

References

(1) 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 Research* **2020**, *48* (22), 12436–12452. https://doi.org/10.1093/nar/gkaa1053.

(2) 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. *Molecular Cell* **2020**, *80* (6), 1067-1077.e5. https://doi.org/10.1016/j.molcel.2020.11.004.

(3) 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* (December), 1–18. https://doi.org/10.3389/fgene.2020.572702.

(4) Rangan, R.; Watkins, A.; Kladwang, W.; Das, R. De Novo 3D Models of SARS-CoV-2 RNA Elements and Small-Molecule-Binding RNAs to Guide Drug Discovery. *Nucleic Acids Research* **2021**, *49* (6), 3092–3108. https://doi.org/10.1093/nar/gkab119.

(5) Tengs, T.; Delwiche, C. F.; Jonassen, C. M. A Genetic Element in the SARS-CoV-2 Genome Is Shared with Multiple Insect Species. *Journal of General Virology* **2021**, *102* (3), 1–5. https://doi.org/10.1099/JGV.0.001551.

(6) 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. *Biochemical and Biophysical Research Communications* **2021**, *545*, 75–80. https://doi.org/10.1016/j.bbrc.2021.01.013.

(7) 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. *Molecular Cell* **2021**, *81* (3), 584-598.e5. https://doi.org/10.1016/j.molcel.2020.12.041.

(8) Morandi, E.; Manfredonia, I.; Simon, L. M.; Anselmi, F.; van Hemert, M. J.; Oliviero, S.; Incarnato, D. Genome-Scale Deconvolution of RNA Structure Ensembles. *Nature Methods* **2021**, *18* (3), 249–252. https://doi.org/10.1038/s41592-021-01075-w.

(9) 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. https://doi.org/10.1016/j.cell.2021.02.008.

(10) 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. https://doi.org/10.1038/s41467-021-22785-x.

(11) 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. *Journal of Virology* 2021, 95 (14), 1–21. https://doi.org/10.1128/jvi.00663-21.

(12) 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. *Nature Communications* 2021, 12 (1), 1–12. https://doi.org/10.1038/s41467-021-25999-1.

(13) 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). https://doi.org/10.1371/journal.pbio.0030005.

(14) 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.; Schwabbe, 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.; Zetzschke, H. Secondary Structure Determination of Conserved SARS-CoV-2 RNA Elements by NMR Spectroscopy. *Nucleic Acids Research* 2020, 48 (22), 12415–12435. https://doi.org/10.1093/nar/gkaa1013.

(15) Biesiada, M.; Pachulskia-Wieczonek, K.; Adamiak, R. W.; Purzycka, K. J. RNAComposer and RNA 3D Structure Prediction for Nanotechnology. *Methods* 2016, 103, 120–127. https://doi.org/10.1016/J.YMETH.2016.03.010.

(16) Antczak, M.; Popenda, M.; Zok, T.; Sarzynska, J.; Ratajczak, T.; Tomczyk, K.; Adamiak, R. W.; Szachniuk, M. New Functionality of RNAComposer: An Application to Shape the Axis of MiR160 Precursor Structure. *Acta Biochimica Polonica* 2016, 63 (4), 737–744. https://doi.org/10.18388/abp.2016_1329.

(17) Jhunjhunwala, A.; Ali, Z.; Bhattacharya, S.; Halder, A.; Mitra, A.; Sharma, P. On the Nature of Nucleobase Stacking in RNA: A Comprehensive Survey of Its Structural Variability and a Systematic Classification of Associated Interactions. *Journal of Chemical Information and Modeling* 2021, 61 (3), 1470–1480. https://doi.org/10.1021/acs.jcim.0c01225.

(18) Leontis, N. B.; Stombaugh, J.; Westhof, E. The Non-Watson-Crick Base Pairs and Their Associated Isostericity Matrices. *Nucleic Acids Research* 2002, 30 (16), 3497–3531. https://doi.org/10.1093/nar/gkf481.
(19) Lloyd, S. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28* (2), 129–137. https://doi.org/10.1109/TIT.1982.1056489.

(20) Macqueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; 1967; Vol. 1, pp 281–297.

(21) *scikit-learn* 2.3.2 *K*-means. https://scikit-learn.org/stable/modules/clustering.html#k-means (accessed 2021-11-23).

(22) van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.

(23) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.

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

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

(26) von Luxburg, U. A Tutorial on Spectral Clustering. *Statistics and Computing* **2007**, *17* (4), 395–416. https://doi.org/10.1007/s11222-007-9033-z.

(27) Górńska, A.; Jasiński, M.; Trylska, J. MINT: Software to Identify Motifs and Short-Range Interactions in Trajectories of Nucleic Acids. *Nucleic Acids Research* **2015**, *43* (17). https://doi.org/10.1093/nar/gkv559.

(28) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reiber, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: Software for Analysis of Nucleic Acid Structures and Trajectories. *RNA* **2019**, *25* (2), 219–231. https://doi.org/10.1261/rna.067678.118.