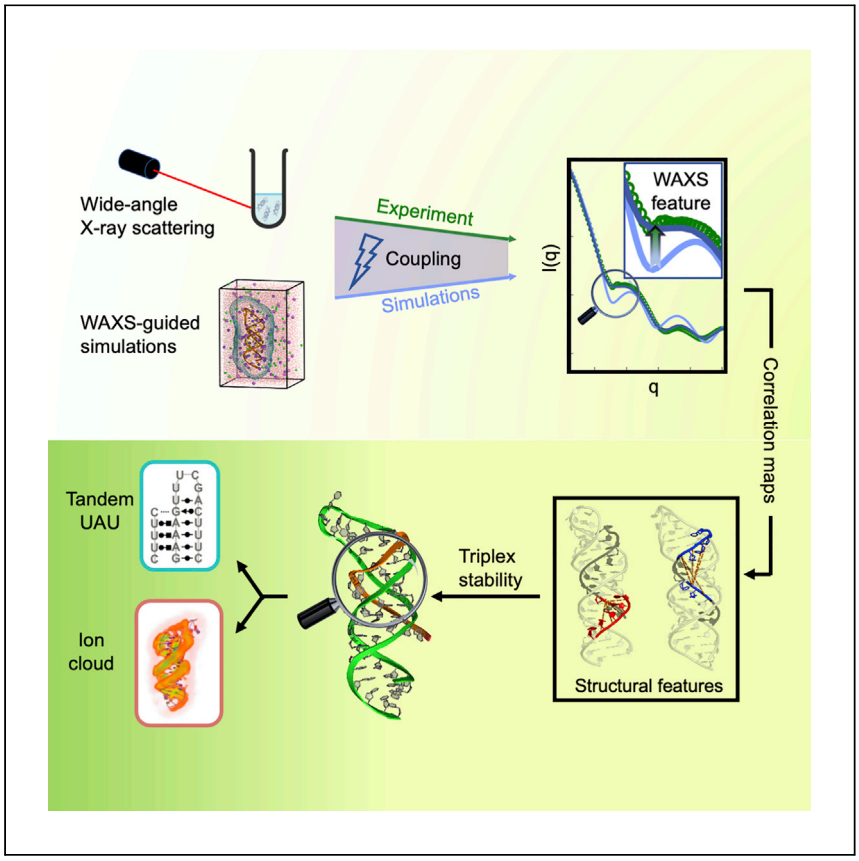
Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Insights into the structural stability of major groove RNA triplexes by WAXS-guided MD simulations



Single RNA strands are often hydrogen bonded to form duplexes, but while the combination of three strands into a triplex is also widespread in nature, it is less well understood. Chen et al. use tight coupling of solution X-ray scattering and atomically detailed simulations to provide new insights into the solution structures of these partially flexible RNA motifs.

Yen-Lin Chen, Weiwei He, Serdal Kirmizialtin, Lois Pollack

serdal@nyu.edu (S.K.)
lp26@cornell.edu (L.P.)

Highlights

Structures of RNA triplexes are derived from computational and experimental data

Data-driven methods allow finding relationship between profiles and structure

RNA major groove expands and counterions move to accommodate triplex-forming RNA strands

Tandem UAU base triples and cation localization provide stabilization for RNA triplexes

Chen et al., Cell Reports Physical Science 3, 100971
July 20, 2022
https://doi.org/10.1016/j.xcrp.2022.100971

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Insights into the structural stability of major groove RNA triplexes by WAXS-guided MD simulations

Yen-Lin Chen,[1,4] Weiwei He,[2,3,4] Serdal Kirmizialtin,[3,*] and Lois Pollack[1,5,*]

SUMMARY

RNA triple helices are commonly observed tertiary motifs that are associated with critical biological functions, including signal transduction. Because the recognition of their biological importance is relatively recent, their full range of structural properties has not yet been elucidated. The integration of solution wide-angle X-ray scattering (WAXS) with molecular dynamics (MD) simulations, described here, provides a new way to capture the structures of major-groove RNA triplexes that evade crystallographic characterization. This method yields excellent agreement between measured and computed WAXS profiles and allows for an atomically detailed visualization of these motifs. Using correlation maps, the relationship between well-defined features in the scattering profiles and real space characteristics of RNA molecules is defined, including the subtle conformational variations in the double-stranded RNA upon the incorporation of a third strand by base triples. This readily applicable approach has the potential to provide insight into interactions that stabilize RNA tertiary structure that enables function.

INTRODUCTION

Recognition of RNA's importance as a powerful tool in supporting life and fighting disease continues to grow. An improved description of its structures is essential to further understand RNA's regulation of biological function and to advance its application as a therapeutic[1] and/or vaccine.[2] Many RNAs contain base paired, duplex regions, connected by a variety of different motifs that often position unpaired nucleotide bases for additional stabilizing interactions. These latter interactions vary in type, and they can involve long-range base pairing or other more intricate tertiary interactions that secure distal parts of the molecule. A survey of these interactions can be found in Butcher et al.[3]

An intriguing class of RNA tertiary motifs contain base triples. Formed by stacked Hoogsteen base triples U-A•U and C-G•C^+, RNA triplexes are now recognized as tertiary motifs[3–5] in naturally occurring functional RNA architectures.[6] As examples, in human telomerase,[7–10] RNA triplexes with two to five base triples are necessary structural elements required for catalytic activity.[11] Spliceosomal RNAs[12,13] use RNA base triples as conserved catalytic sites that coordinate metal ions in the splicing process. RNA triplexes also exist in long noncoding RNAs (lncRNA)[14–16] and viral RNAs,[17,18] where they can stabilize functional domains by forming long, consecutive U-A•U triples with adjacent nucleotides. Such stabilization, notably the sequestration of degradation signals (such as the polyA tail), results in accumulation of RNAs and, in turn, upregulation of gene expressions in both cancer[19] and viral infection[20,21] contexts. In this context, RNA triplexes are valuable therapeutic targets.[22] Finally, some gene-regulatory RNA riboswitches[23–25] leverage triplexes

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA

[2]Department of Chemistry, New York University, New York, NY 10003, USA

[3]Chemistry Program, Science Division, New York University Abu Dhabi, Abu Dhabi 129188, UAE

[4]These authors contributed equally

[5]Lead contact

*Correspondence: serdal@nyu.edu (S.K.), lp26@cornell.edu (L.P.)

https://doi.org/10.1016/j.xcrp.2022.100971

as binding pockets of ligands regardless of nucleobase moiety. Despite their biological or technological importance,[26] the structures of only a few longer triplex structures have been solved (e.g., PDB: 6SVS[27]). Although these structures possess well-defined contacts, the degree of flexibility displayed by the third, triplex-forming oligomer, is unknown, yet is likely important in understanding how these molecules interact with partners. The goal of this work is to explore the conformation(s) of some RNA triplexes that incorporate both stable and flexible regions, and to study the counterion distribution around such highly charged structures.

There has been much recent interest in developing accurate, atomically detailed models of RNAs. In contrast to well-ordered proteins, all-atom MD simulations present challenges for RNA,[28] including concerns about force field accuracy or sampling issues resulting from RNA's rough energy landscape.[29] Despite recent efforts,[30,31] it is not yet clear if current approaches provide a general framework to provide atomically detailed descriptions consistent with experiments. As an alternative approach, all-atom models of RNAs are being successfully refined and tuned to come to agreement with experimental data acquired by NMR,[32–36] cryo-EM,[37] or solution small-[38] or wide-angle scattering (SWAXS).[39,40]

A recent study that motivated this one combines the Sample and Select (SaS) approach with wide-angle X-ray scattering (WAXS) data on RNA duplexes.[39] Results provide insight both into variations from A form structure and counterion distributions around double-stranded RNA. With the goal of modeling structures that are more complex than the duplexes that were the subject of our previous work, we attempted a similar comparison between SWAXS measurements of constructs containing multiple, stacked RNA base triples, and computed profiles of an "ideal" triplex, subject to extended sampling. This SaS approach failed to capture the details reported in the experiment; the "correct" structures do not appear to be present in the pool. This disagreement motivated the application of the data-driven, all-atom MD simulations to enhance our previous SaS approach described herein. This method relies on tight coupling of simulation with solution X-ray scattering data. Notably, the incorporation of experimental data generated by wide-angle X-ray scattering offers the opportunity to measure (and model) finer details of the backbone placement and flexibility, which reflects base pair geometries as well as ion distributions. The detailed features at wide angle are critical in interpreting the features that are recapitulated in the MD simulations. Such approaches have been previously applied to proteins[41] but not to nucleic acids. Because of the rising appreciation of the biological importance of RNAs, it is now critical to gain a more complete understanding of all possible RNA contacts, geometries, and interactions. The ability to navigate simulations with minimal biases in the starting conformation will allow us to rapidly explore novel contacts, structures, and interactions. As mentioned above, similar data-guided approaches are being exploited in conjunction with other experimental measurements to target RNA structures.[32–38]

We apply this integrated approach to reveal the solution structures of hairpin RNAs alone and complexed with triplex-forming oligomers (TFOs). Of great interest are the changes to the duplex structure that accommodate and stabilize base triple interactions. The atomically detailed structural ensembles afforded by MD simulations reveal changes to the cation atmosphere upon triplex formation, and resolve subtle differences in the measured WAXS signals that depend on the presence of particular (physiological) salt ions. Strikingly, our findings elucidate the unique role of U• A-U units in triple helical structures, offering biophysical insight into the interactions that stabilize RNA triplexes.

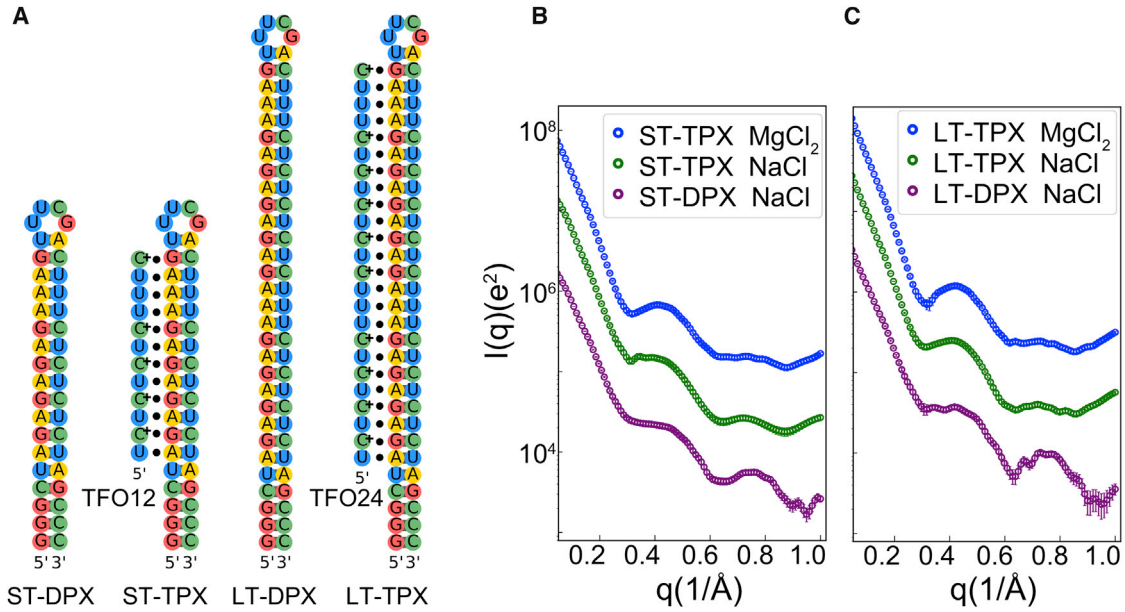


Figure 1. RNA systems under study and their WAXS profiles

(A) Schematics showing the RNA duplex and triplex motifs designed for this work. The Short-Tetraloop Duplex (ST-DPX) has a UUCG tetraloop and a 17-Watson-Crick base-paired stem with one TFO12 binding domain. A custom designed 12 nucleotide long TFO is added to yield multiple base triples, stabilized by Hoogsteen base pairing (dots) resulting in an RNA triplex: ST-TPX. We doubled the TFO12 binding domain to create a triplex-forming partner for the Long-Tetraloop Duplex (LT-DPX). When bound, the duplex-TFO24 pair comprises the longer RNA triplex: LT-TPX. See main text for a more detailed description.

(B and C) Solution X-ray scattering profiles of the different motifs and salts studied: ST-TPX/LT-TPX in 5 mM $MgCl_2$ (blue), ST-TPX/LT-TPX in 200 mM NaCl (green), and ST-DPX/LT-DPX in 200 mM NaCl (purple).

RESULTS

WAXS reveals structural features of RNA triplex

Solution X-ray scattering provides both global and local information about the structure(s) of RNA. While small-angle X-ray scattering reveals the overall size and shape of the molecule, WAXS is now a proven tool for discerning variations in the solution structure(s) of RNA molecules,[39,42,43] importantly reporting higher-resolution information about the distribution of distances present in the molecule *in vitro*. We apply WAXS to understand the structure(s) of motifs with tertiary structures: triple-stranded RNAs.

To obtain triplex structures for measurement, we designed two duplexes as well as a complementary single-stranded TFO that binds to each duplex (see RNA triple helix motif design and sample preparation). We subsequently refer to the duplex motifs as ST-DPX and LT-DPX for short and long constructs, respectively. Their respective triplex counterparts are referred to as ST-TPX and LT-TPX.

The secondary structures of the molecules studied here are shown in Figure 1A. Figures 1B and 1C shows their WAXS profiles, plotted as the log of the intensity of the scattered X-rays, as a function of scattering angle, expressed as $q = (4\pi/\lambda)\sin\theta$, where λ is the incident X-ray wavelength in Å and θ is half of the scattering angle. Profiles of the shorter duplex and triplex (ST-DPX and ST-TPX) are shown in the left panel, in solutions containing different salt ions. Profiles of the longer constructs are shown in the right panel. The difference between duplex and triplex scattering is evident from the curves. The peak in the triplex scattering profile (near $q = 0.4\ \text{Å}^{-1}$) flattens for the duplex and the peak in the duplex scattering

profile (near q = 0.75 $Å^{-1}$) flattens for the triplex. The WAXS profile of the triplex sample is sensitive to the surrounding ions; features in divalent Mg are better defined than those in monovalent Na (note: a similar effect was observed for duplexes, as reported in He et al.[39]). Finally, comparison of Figures 1B and 1C shows that the WAXS features are better defined in the longer constructs. This is not surprising given the increased reinforcement of correlations with length in a regular structure.

Computational method of sampling RNA conformations driven by WAXS

The sensitivity of WAXS to the variations with length, sequence, and salt is promising; however, the information hidden in WAXS curves demands computational modeling to provide precise insights. To reveal the structural differences implied by WAXS, we employed a computational strategy that is summarized in Figure 2A. Simulations begin with an RNA model, built from only secondary structure information (Figure 1). We use this model structure to generate an ensemble of conformations using MD simulations that account for the RNA flexibility, explicit water, and cations (Figure 2B). Details of our MD methodology can be found in methods and supplemental Appendix (see WAXS-driven MD simulations). We compute the WAXS profiles from the atomic coordinates. Later, we assess the agreement of the profile derived from the computational pool by comparing it with the measured scattering profiles from experiment, using a customized χ^2 metric (see The customized statistical metric). If agreement is not achieved, a high-temperature (HT) MD simulation is used to expand the structural pool. The HT trajectory is clustered to identify important conformational sub-states. Each cluster is then evaluated against experimental data. If the χ^2 of the best cluster is satisfactory we select the conformation and further sample conformations around it using unbiased MD simulations. As shown in our past work,[39] this SaS strategy can be effective if the issue is sampling, meaning that the correct structures are present in the pool, but may be difficult to access given the starting point of the simulation (A-form duplex). For more complex structures, where the correct models may not be present within the computational pool, a new approach is required to refine the model. The WAXS-driven MD simulations, described here, allow us to steer the best conformation of the SaS simulation in new directions to achieve the requisite agreement.

To test the validity of our approach in resolving triplex RNA structures, we start from a known structure. The triplex part of the PDB: 6SVS was isolated and its WAXS curve is synthetically generated using the WAXSiS program[44] by adding explicit ions and water mimicking experimental conditions. Details of our approach can be found in the supplemental information (see WAXS computing from MD trajectory). Later, we generated an ideal triplex model using 3D-NuS[45] and sampled conformations committed to fit the WAXS profile. We tested the performance of the WAXS-driven MD in two conditions: (1) when the RNA is protonated at the C•G-C triplet region and (2) when it is left unprotonated. In Figure 3 we summarize our main results and in Figure S4 we show the convergence of the structural sampling during WAXS-driven MD simulations.

As is seen in Figure 3, the WAXS profile of the initial model (black) deviates significantly from the crystal structure (green), evidenced by the large $\chi^2 > 300$. After WAXS-driven MD, both the protonated and unprotonated RNA setups started from the same initial model structure give rise to a perfect fit with the WAXS profile derived from the crystal structure (Figure 3A) with $\chi^2 \approx 6.99$ and 8.24. The best fit models overlaid to the crystal structure show high similarity with the crystal structure. Interestingly, WAXS-driven MD gave rise to a similar performance whether the RNA

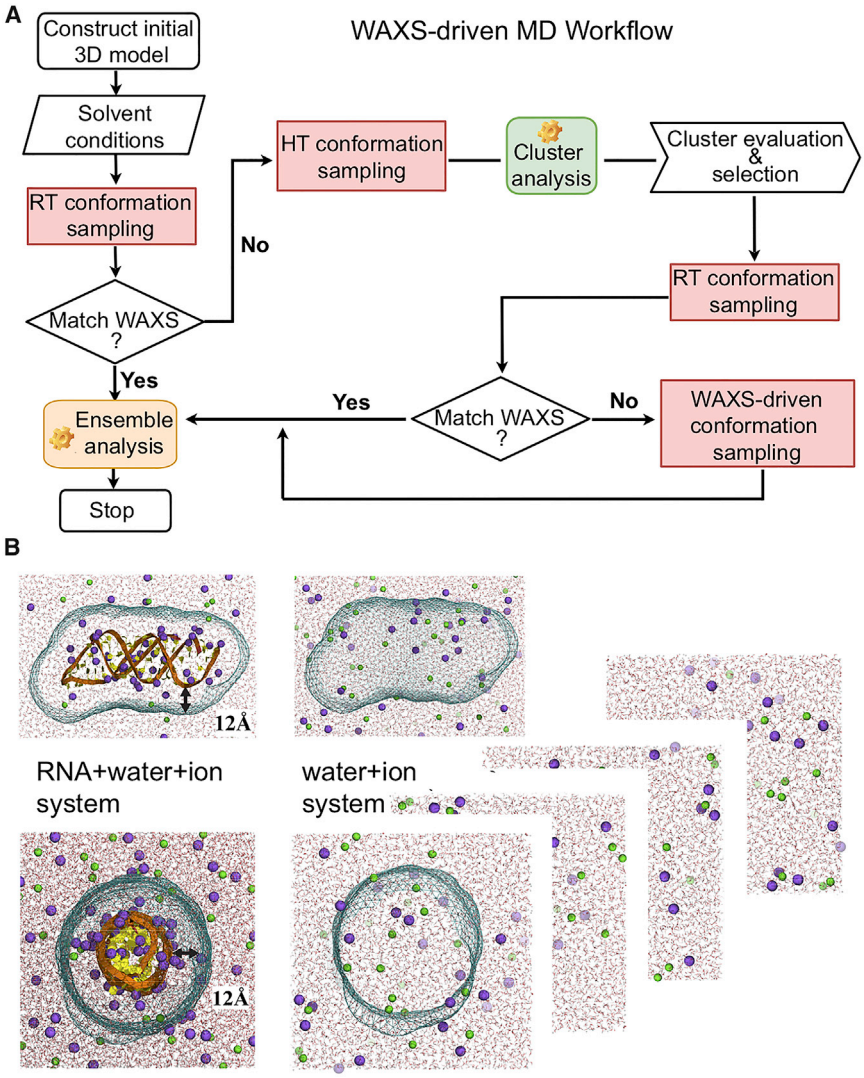


Figure 2. The flowchart of WAXS data-driven MD approach and the method of computing WAXS profiles from simulations

(A) Two feedback loops are considered, and the path through the search is determined by the agreement (or lack of agreement) with WAXS profiles. If agreement is not achieved in the RT pool, the Sample and Select (SaS) approach[39] is incorporated to enhance the sampling. The second feedback loop is triggered when the customized metric in Equation 1 of all the sampled conformations exceeds the threshold. After multiple iterations, the data-driven strategy generates a conformation pool satisfying the experimental WAXS data, assessed by the value of our customized metric.

(B) Illustration of the RNA + water + ion system and water + ion system used in the WAXS data-driven MD simulations. The molecular envelope constructed from a 3D probability density isosurface 12 Å from the helix surface to encompass the solvent/ion shell is depicted by mesh. The scattering intensity of water + ion system was computed by applying the same envelope construct.

is protonated or not, suggesting that the provided experimental data not only improve force field-related issues but also eliminate other uncertainties that may arise due to the solution conditions. This exercise allowed us to critically assess the ability of WAXS-driven MD to refine triplex structures and made it possible to study RNA triplexes where no crystal structure is available.

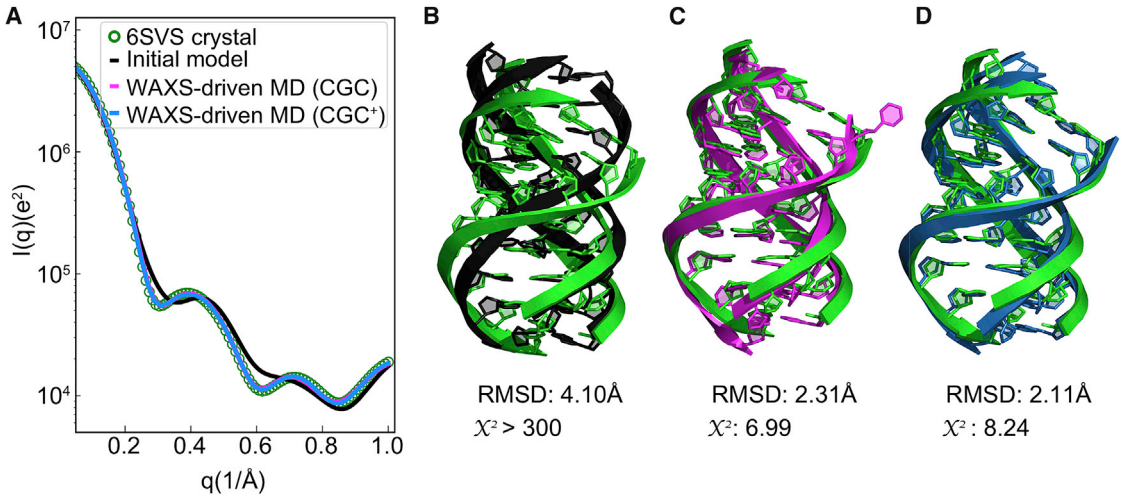


Figure 3. Illustration of the WAXS data-driven approach against a known triplex (PDB: 6SVS)

The crystal structure serves to compute the WAXS profile that is used to drive an ideal triplex model.

(A) WAXS profiles of the crystal structure (green), initial model (black) and best fit models after WAXS-driven MD simulations with (blue) and without (purple) protonation of CGC triplet.

(B–D) The comparison of structures overlaid to the 6SVS structure. The RMSD (Å) and the χ^2 metric for each condition are shown.

Computational method generates WAXS-consistent RNA ensembles for structurally unknown triplex

The success of WAXS-driven MD in refining the structure of the known triplex RNA motivates our study of structures that have not yet been characterized. To accomplish this, we begin with the short triplex (ST-TPX). After the starting conformation is selected, our WAXS-driven approach guides the MD toward regions of conformational space where the computed profiles find better agreement with the experimental data. Figure 4A (black) shows results for the ST-TPX RNA in NaCl. The process, as applied to the other RNAs studied, yields similar improvements in fit and is reported in Figure S5. The structural search reaches our set criteria ($\chi^2 < 2$) within 40 ns (Figure 4B). Once the search converges to a structure consistent with experimental signal the error stays around the same value, suggesting no major change in WAXS profiles. To provide ensembles for analysis, we construct a pool from the conformations that fall below the threshold. These pools in each condition are used to derive equilibrium properties and distributions. Throughout the course of the conformational search stage of the simulation, the error and the structural changes are simultaneously monitored (Figures 4B and 4C). Interestingly, significant deviations from the starting structure are localized to specific regions; for example, near residues 19–22 and 41–52 in ST-TPX. Changes are "tracked" by monitoring structural deviations in these regions that correspond to the tetraloop and TFO position of the ST-TPX, respectively. As a control, we employ unbiased MD starting from the initial RNA structure and from the best structure selected by WAXS-driven MD (Figure 4A, red versus cyan green). The agreement of unbiased MD is poor in both cases, highlighting the necessity of our approach. Specifically, the spontaneous drift of the structural pool from the best structure suggests discrepancies in the force field in their representation of triple-helix RNAs. In contrast, computed WAXS profiles of duplex structures evolve toward better agreement with experiments when sampling starts from the conformations close to the experimentally measured ones (Figures S5A and S5C). Thus, our current force field appears to adequately capture duplex topology, but needs adjustments when modeling higher order structures.

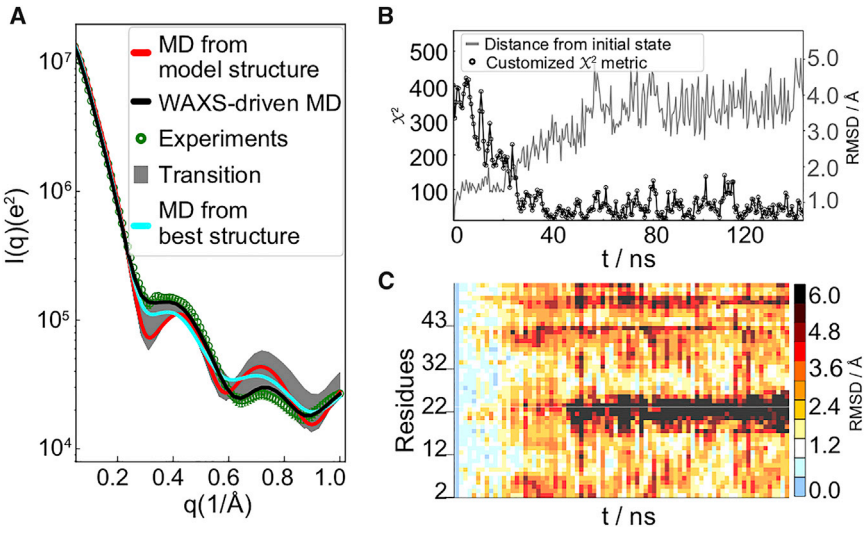


Figure 4. Illustration of WAXS data-driven approach of ST-TPX in buffered solution of 200 mM NaCl

(A) The experimental data that are the target of the WAXS-driven MD simulations are shown as green circles. The red curve is the ensemble average of the unbiased simulation results when the initial model is taken as the starting structure. The cyan curve is the ensemble average of the unbiased simulation when the initial structure is the best fit model. Neither of them represents a single structure, rather both represent ensemble averages. The progression of scattering profiles during the data-driven simulation is shown in gray, denoted as transition in the figure and the result of the data-driven simulation is shown in black. The WAXS evolution of the remaining constructs is shown in Figure S5.

(B) The time variation of χ^2, as well as the overall RMSD change are shown, using the initial frame as reference. The distance from the initial state is given in Ångstroms. The asymptotic amplitude and turning point of the two variables respectively used as a qualitative measure of the convergence of the fits to a 3D model. Note that this convergence is not a rigorous statistical measure.

(C) The time evolution of the RMSD at the level of individual residues is shown, using the initial structure as the reference. The color bar provides the distance from initial conformation. Note the relatively higher structural deviations displayed near residues 19–22 and 41–52 residues, which correspond to the tetraloop and TFO of ST-TPX, respectively.

Pairwise distances correlate with WAXS profiles

Key to the success of this method is understanding how WAXS features correlate with RNA structural features. Structures sampled by MD simulation serve as a training set to find relationships between features in the computed WAXS profile and 3D structural details. We first examined the correlation between residue pair distances d_{ij} and WAXS amplitude at each discrete point q_k, $I(q_k)$ (see detailed description in Pairwise distance correlation analysis). Correlation maps are used to display these relationships. Figure 5 summarizes our analysis for ST-TPX. A comparable analysis for LT-TPX is detailed in Figure S9.

The normalized pairwise distances of Figure 5 show both positive and negative correlations with SAXS features. A positive correlation implies an increase in $I(q)$ with an increase in normalized deviation of distance pair at a particular q value, with the opposite true for negative correlations. Here, we show only the highest correlations in magnitude for three q ranges (0.3–0.4, 0.6–0.75, and 0.8–0.95 $Å^{-1}$) (Figure 5).

These analyses demonstrate that the TFO conformations play a significant role in shaping the scattering profiles (Figure 5). As an example, the experimental profile

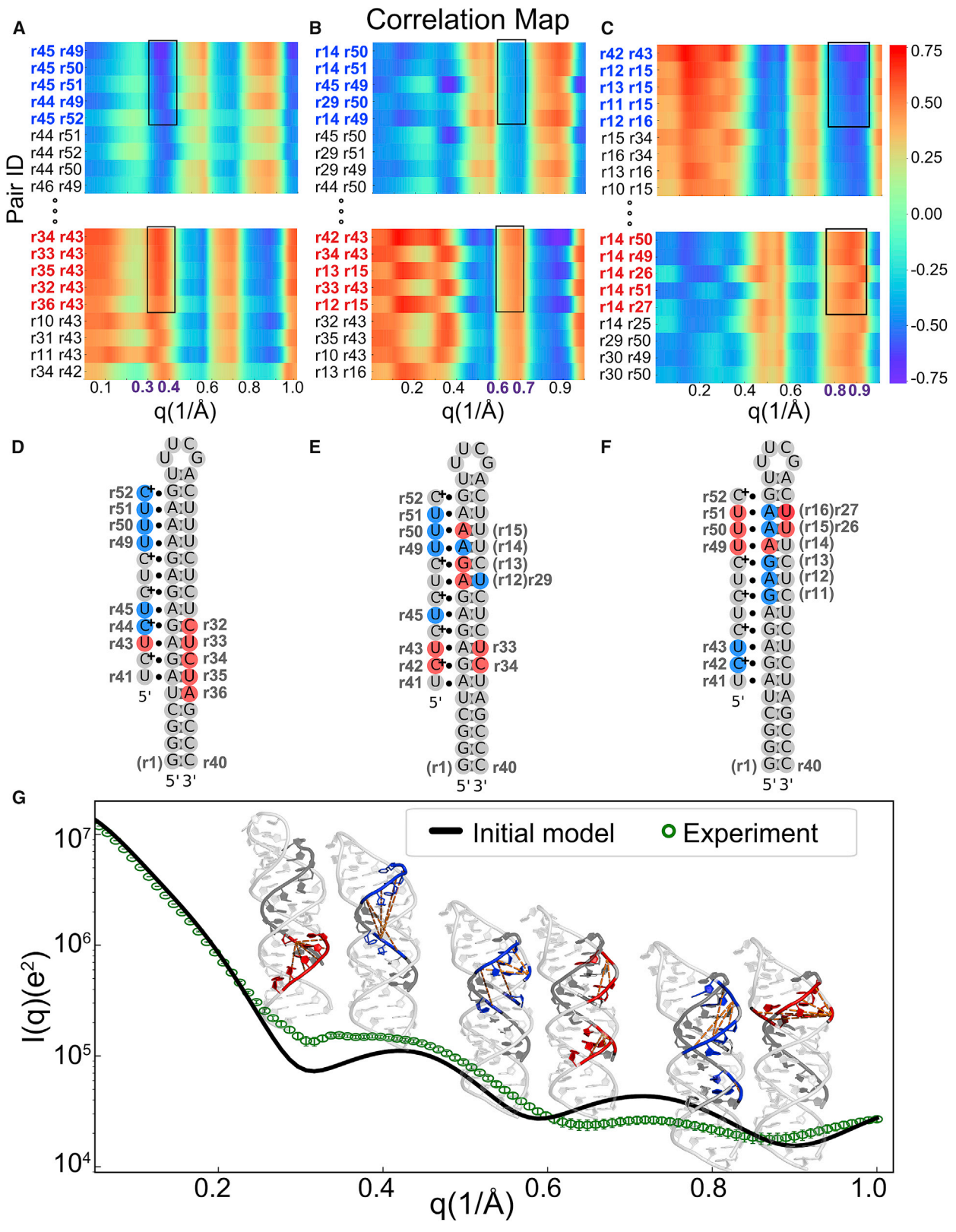

Correlation Map
A
B
C
Pair ID
r45 r49
r45 r50
r45 r51
r44 r49
r45 r52
r44 r51
r44 r52
r44 r50
r46 r49
r34 r43
r33 r43
r35 r43
r32 r43
r36 r43
r10 r43
r31 r43
r11 r43
r34 r42
0.1 0.3 0.4 0.6 0.8 1.0
q(1/Å)
r14 r50
r14 r51
r45 r49
r29 r50
r14 r49
r45 r50
r29 r51
r29 r49
r44 r50
r42 r43
r34 r43
r13 r15
r33 r43
r12 r15
r32 r43
r35 r43
r10 r43
r13 r16
0.2 0.4 0.6 0.7 0.9
q(1/Å)
r42 r43
r12 r15
r13 r15
r11 r15
r12 r16
r15 r34
r16 r34
r13 r16
r10 r15
r14 r50
r14 r49
r14 r26
r14 r51
r14 r27
r14 r25
r29 r50
r30 r49
r30 r50
0.2 0.4 0.6 0.8 0.9
q(1/Å)
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
D
E
F
5'
5' 3'
G
Initial model
Experiment
I(q)(e²)
q(1/Å)

Figure 5. Correlation analysis of measured WAXS profiles with structures sampled by MD simulations of ST-TPX in 200 mM NaCl

(A–C) These partial correlation maps, extracted from the full correlation map (Figure S7), display the distance between pairs of phosphate atoms associated with the indicated residues, computed from MD structures (Equation 4), against the normalized deviation of the computed WAXS profile from the experimental measurement (Equation 3). These maps identify structural regions that potentially contribute to deviations in specific q regions. Regions with the highest deviations in a given q range, e.g., the residue pairs with high correlations ($|\rho| > 0.5$ and five largest values), are highlighted within black rectangles. We focus here on structural regions that create significant deviations in profiles near $q \approx 0.35 Å^{-1}$ (A), $q \approx 0.65 Å^{-1}$ (B), and $q \approx 0.85 Å^{-1}$ (C).

(D–F) The regions identified in (A–C) are mapped onto the secondary structure and real space models of the RNAs. Residue pairs that contribute to changing the scattering profile near $q \approx 0.35 Å^{-1}$ (D), $q \approx 0.65 Å^{-1}$ (E), and $q \approx 0.85 Å^{-1}$ (F), respectively, are highlighted in color.

(G) connects distinct structural variation with changes in the scattering profiles. A comparable analysis applied to LT-TPX is provided in Figure S9.

deviates significantly from the unbiased MD prediction near $q = 0.35 Å^{-1}$. The correlation map shown in Figure 5A provides insight into the structural changes that can bring the experimental and predicted curve into closer agreement at this q value. Specifically, a decrease in distance between pairs highlighted with blue blocks, as well as a flipping out of the end of the TFO, which increases the pairwise distances shown by the red blocks, suffice to increase the computed scattering intensity at this q value, smearing the obvious minimum that is present in the standard MD curve. This change corresponds to a compression of the TFO along its length (decreased distance between the blue-colored residues in Figure 5D), as well as an increase in the distance between its 5′ end and the duplex. Together, these effects enhance the signal intensity at $q = 0.35 Å^{-1}$ to bring the computed and experimental curves into agreement (Figures 5A, 5D, and 5G). Similarly, the discrepancy between unbiased MD and measurement seen at around $q = 0.65 Å^{-1}$ can be compensated by breaking the symmetry in the duplex-TFO pairwise distance (Figures 5B and 5E). This analysis suggests that variations in the structure may not always be reflected at a unique q range, rather, the auto-correlations among WAXS features (see Figure S6A for ST-TPX) suggests that structural variations may induce changes at multiple q regions. Despite the complexity that this multiple dependency can induce, these cross-correlations do ensure consistency. For example, the WAXS features at around $q \approx 0.35 Å^{-1}$ positively correlate with that of $q \approx 0.65 Å^{-1}$; however, the same feature negatively correlates with intensity at $q \approx 0.85 Å^{-1}$ (Figure S6A). Therefore, a closer duplex-TFO contact amplifies the WAXS amplitude at $q \approx 0.65 Å^{-1}$ (Figures 5B, 5E, and 5G, blue blocks) and, at the same time, reduces the $I(q)$ at $q \approx 0.85 Å^{-1}$ (Figures 5C, 5F, and 5G, red blocks). Meanwhile, stem expansion near the loop increases the amplitude at $q \approx 0.85 Å^{-1}$ (Figures 5C, 5F, and 5G). Thus, the agreement of our computed profile with experiment over the entire q range provides support for all of the selected changes.

The above discussion highlights our experimental sensitivity to the conformational state of the TFO. Additional information about changes in the duplex topology that result from TFO binding are also available: features in the WAXS profile correlate with the major-groove width. The positive correlation at $q \approx (0.8–0.95 Å^{-1})$, especially residue pairs of r14-r26 and r14-r27 (Figures 5C and 5F), suggests a major groove widening, which is detailed in the next section. Despite the above-described success in articulating the structures of the duplex and the TFO, the correlation maps do not provide robust guidance (high values of correlation) for the changes in the tetraloop region (Figure S7). This is not surprising because of the highly dynamic nature of the loops revealed by MD simulations,[46] as well as the short length scales that characterize interactions within the loops. Data acquired at higher q values, beyond the scope of the present study, will be required to inform about the structure(s) of this region. The most representative structures from the WAXS-driven simulation are shown in Figures 6B, 6C, and 6D.

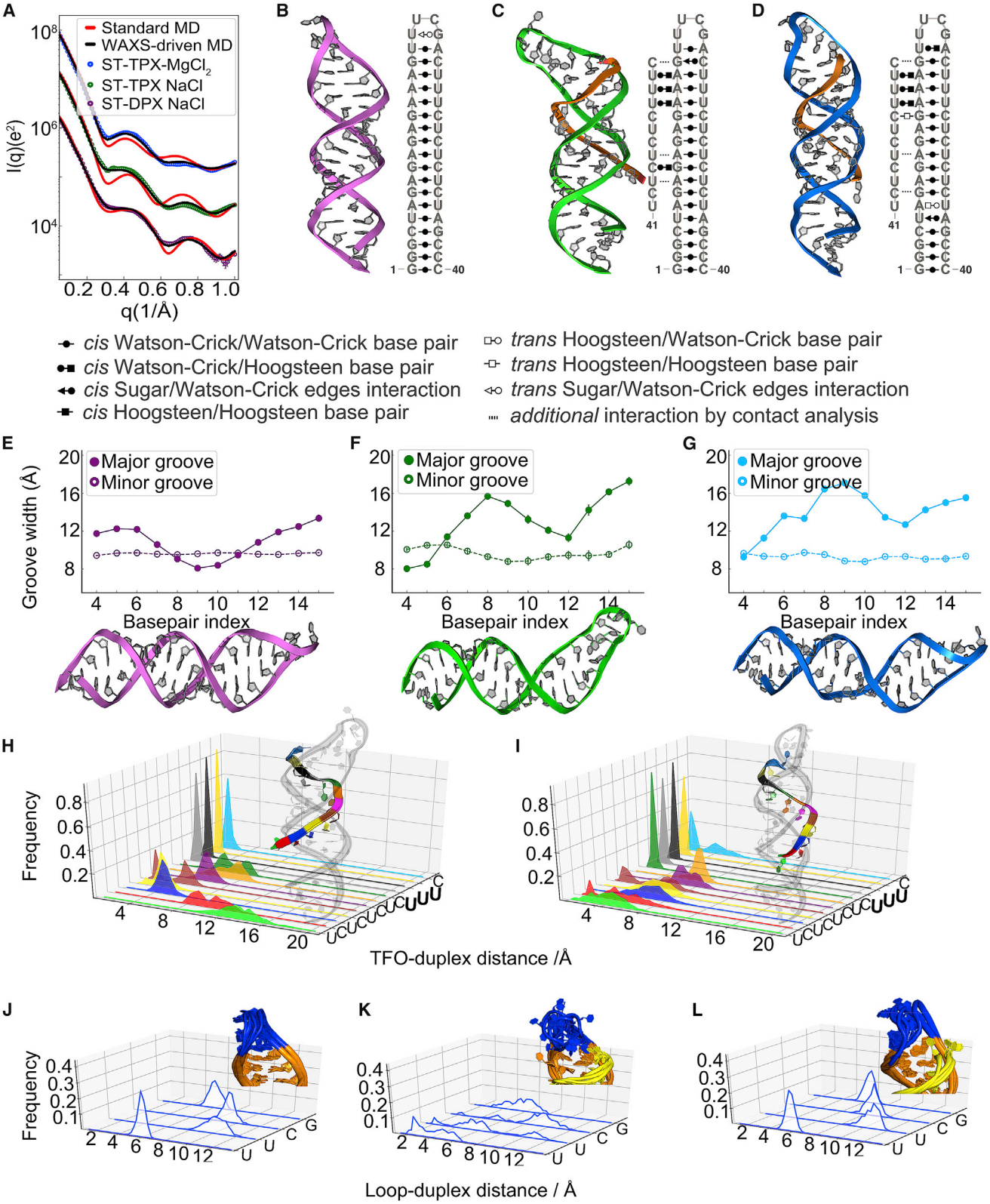

A
B
C
D
Standard MD
WAXS-driven MD
ST-TPX-MgCl2
ST-TPX NaCl
ST-DPX NaCl
I(q)(e²)
q(1/Å)
cis Watson-Crick/Watson-Crick base pair
cis Watson-Crick/Hoogsteen base pair
cis Sugar/Watson-Crick edges interaction
cis Hoogsteen/Hoogsteen base pair
trans Hoogsteen/Watson-Crick base pair
trans Hoogsteen/Hoogsteen base pair
trans Sugar/Watson-Crick edges interaction
additional interaction by contact analysis
E
F
G
Major groove
Minor groove
Groove width (Å)
Basepair index
H
I
Frequency
TFO-duplex distance /Å
J
K
L
Loop-duplex distance / Å

Figure 6. Structural analysis of ST-DPX/TPX under different conditions

(A) The experimental scattering profiles (circles), computed WAXS profiles from standard MD simulations of model structure (red) and WAXS-driven simulations (black) are pictured for three constructs.

(B–D) From the conformational ensembles generated by WAXS-driven MD, we show the most representative structure of (B) ST-DPX in NaCl, (C) ST-TPX in NaCl, and (D) ST-TPX in $MgCl_2$, and alongside each snapshot, we report the corresponding secondary structures using Leontis-Westhof notation,[47] and dashed line annotation denotes weak hydrogen bonding interactions for other non-canonical base pair formations. For equilibrium properties we consider the whole ensemble. The ensemble was generated from conformations that fall below the threshold $\chi^2 < 2$ value.

(E–G) Average groove width (GW) of the duplex part of structures reveals a dependence on salt: (E) ST-DPX in NaCl, (F) ST-TPX in NaCl, and (G) ST-TPX in $MgCl_2$, respectively.

(H and I) The contact analysis measuring the relative positions of the TFO and duplex of ST-TPX reveals detailed structural information about TFO binding in (H) NaCl and (I) $MgCl_2$.

(J–L) The contact analysis between tetraloop and duplex of (J) ST-DPX in NaCl and ST-TPX in (K) NaCl, and (L) $MgCl_2$ reveals an unexpected salt dependence of the loop within triplex structures.

Duplex and triplex structures and how the TFO binding modulates RNA structural ensemble

We now turn to a detailed discussion of the structures of the ST-DPX and ST-TPX constructs, which is uniquely enabled by the agreement of computed scattering profiles with experimental measurements. First, we note that unbiased MD simulations require only small modifications to accurately recapitulate the structure of the ST-DPX construct (Figure 6A), that is, the WAXS-derived forces did not lead to appreciable change in structure during the WAXS-driven simulation runs. In contrast, structural pools from unbiased MD simulations show much more deviation when modeling the triple-stranded constructs (Figures 6B–6D). In these latter constructs, the final MD-driven result, shown as simulation snapshots and corresponding secondary structures in Figures 6C and 6D, reveal significant variations in the tetraloop and TFO, which may be difficult for unbiased MD to capture. Once the simulations have been driven to agree with the experiment, the real space projection of the WAXS profiles afforded by all-atom MD simulations provide a platform to investigate the TFO-induced structural changes with atomic detail.

To accurately describe and interpret the structural changes, we separately consider the three distinct components of the RNA structure: (1) base-paired duplex stem (without TFO and loop), (2) TFO, and (3) tetraloop regions.

As a first step in providing a description of the duplex structure, as well as the ability to easily describe changes resulting from TFO binding, we focus on the dimensions of the groove widths. Figures 6E–6G displays the widths of both the minor and major grooves for three short constructs. Figure 6E reports the average groove widths computed from the pool for the ST-DPX construct, revealing a consistently sized minor groove and some variations in the major groove width, especially near the ends of the base paired regions. Figures 6F and 6G show the significant TFO-induced changes to the major groove dimensions and therefore to the duplex structure. Most significantly, the major groove widens once the TFO is bound. This change is dramatic near the middle of the duplex and near the tetraloop (beyond base pair 14).

We now focus on the positioning and conformations of the TFO in the pool for the triplex construct in different salt conditions. To quantify our findings, we measured and report the pair distances between TFO and duplex nucleotides (details in Residue-residue contact analysis). These distance distributions allow for a comparison of the sequence specificity of TFO binding across salt conditions and reflect the structural heterogeneity of Hoogsteen base triples. Most interestingly, the U• A• U base triples (UUU end, at the "top" in the figure) form close contacts to the duplex (Figures 6H and 6I). This tight binding shows no variation with the salts used, suggesting a

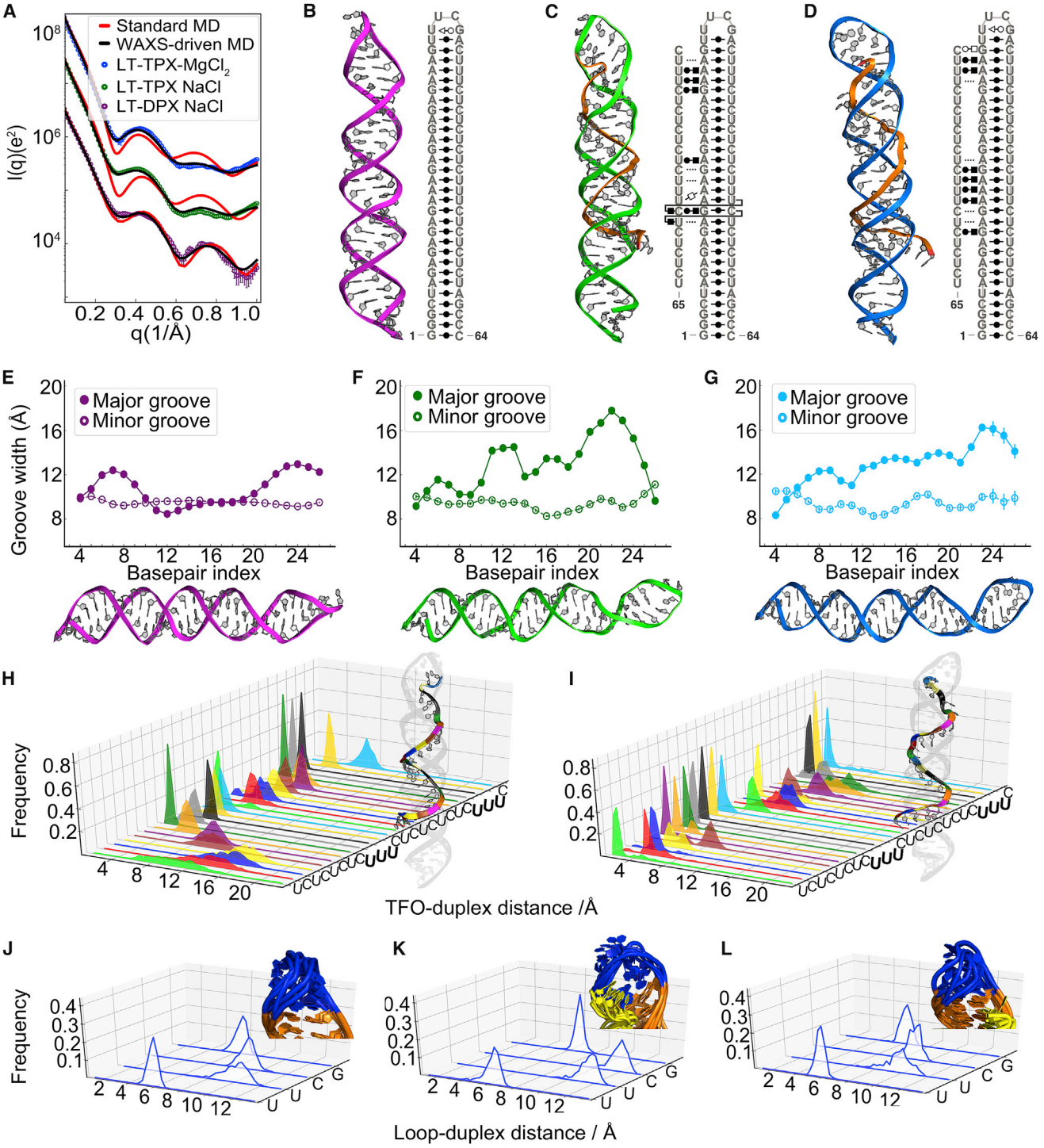


Figure 7. Structural analysis of LT-DPX/TPX

(A) The experimental scattering profiles (circle), computed WAXS profiles from standard MD simulations of model structure (red) and WAXS-driven simulations (black) are pictured for these longer constructs.

(B–D) As in Figure 6, we show the most representative structure of (B) LT-DPX in NaCl and LT-TPX in (C) NaCl and (D) $MgCl_2$, and, alongside each snapshot, we report the corresponding secondary structures using the same nomenclature as Figure 6. For equilibrium properties we consider the whole ensemble. The ensemble was generated from conformations that fall below the threshold $\chi^2 < 2$ value.

Figure 7. *Continued*

(E–G) Average groove width (GW) of the duplex part of structures in different salt conditions reveals a more distinct pattern than observed in the shorter complex for: (E) LT-DPX in NaCl, (F) LT-TPX in NaCl, and (G) LT-TPX in $MgCl_2$, respectively.

(H and I) The contact analysis between TFO and duplex of LT-TPX in (H) NaCl and (I) $MgCl_2$, and (J–L) between tetra-loop and duplex of (J) LT-DPX in NaCl and LT-TPX in (K) NaCl and (L) $MgCl_2$ show smaller variations relative to the shorter construct in NaCl.

short-range non-electrostatic force governing contact formation in these tandem uracil units. This observation confirms earlier work and crystallographic studies.[48–50] While this "upper end" of the TFO appears to be fixed by the tandem uracil units, the contacts at the TFO lower end appear more dynamic. This effect is more pronounced for RNA in Na^+ in comparison to Mg^{2+} salt, and likely reflects the weakened electrostatic screening afforded by monovalent relative to divalent ions.

Finally, we consider the structure(s) of the loop region in our best-fit ensembles. As seen in the TFO contacts, a significant salt variation of this region is observed (Figures 6J–L). Here, the dynamics are depicted both as structures and as computed distances. Both representations illustrate the dynamic features of this flexible region. It is interesting to compare the loop structure in the ST-DPX and the ST-TPX complexes. Relative to the former, the loop structure in the latter tends to twist upon the introduction of the TFO (Figures 6J and 6K). In addition, the ST-TPX loop appears more dynamic in Na^+ than in Mg^{2+} (Figures 6K and 6L). The added strand increases the charge density around the loop, which likely leads to stronger repulsion that partially melts this region. With higher valence Mg^{2+}, the canonical form of the tetraloop is restored (Figure 6L).

All together, we reveal that triplex formation requires global modifications of the duplex geometry in both salt conditions studied. Salt-dependent effects are manifested in the structural variations of the terminating loop as well as in the variations of structure of the TFO, especially near its lower end.

Short and long triplex sequences show similar structural variations

To extend our findings, we measured and simulated systems consisting of longer, loop-terminated duplexes, with and without TFOs. We examined a longer stem-loop construct (LT-DPX) without and with the addition of a longer TFO (LT-TPX). The triplex-forming constructs were examined in solutions containing either Na or Mg ions. Results for the LT-TPX show that the key deviations of MD predictions from WAXS measurements occur in regions that reflect structural coupling between the duplex and the TFO, consistent with findings from the ST-TPX system (Figure S9). As in our above-described studies with the shorter motifs, we applied correlation analysis to bring the predictions for all of the longer constructs into agreement with experiment. Following the application of WAXS-driven MD, the representative best-fit structures of each case are shown in Figures 7B–7D. Conclusions from simulations of the structures of LT-DPX (Figure 7E) are comparable with those derived from the ST-DPX system (Figure 6E). In particular, increases in the major groove width are observed near the two ends of the helical regions, relative to near the center of the duplex. The minor groove widths remain relatively constant and unchanged from the shorter construct. For these longer constructs, the addition of the TFO induces a major groove widening, as seen in the shorter strands; however, it is more emphasized in LT-TPX where a nearly monotonic increase from bottom of the stem to the tetraloop is observed (Figures 7F and 7G).

The binding mode of the longer TFO shows sequence specificity and salt dependence. Consistent with observations on the shorter triplex, the uracil bases of this

Table 1. Structural parameters of RNA constructs from experiments and simulations

	6SVS crystal[27]	ST/LT-DPX [Na^+]	ST/LT-TPX [Na^+]	ST/LT-TPX [Mg^{2+}]
Shift (Å)	− 0.27	0.03±0.11	0.05±0.15	0.10±0.14
Slide (Å)	− 2.32	− 1.52±0.13	− 1.82±0.17	− 1.87±0.16
Rise (Å)	3.17	3.33±0.05	3.41±0.09	3.43±0.10
Tilt (°)	− 2.57	0.30±0.80	0.46±1.18	− 0.11±1.00
Roll (°)	3.40	4.87±1.33	2.32±1.33	0.83±1.47
Twist (°)	29.76	30.99±0.66	29.30±0.94	31.06±1.07
x-displacement (Å)	− 5.14	− 3.62±0.32	− 3.82±0.90	− 3.60±0.39
y-displacement (Å)	0.28	0.01±0.29	0.18±0.66	− 0.04±0.44
Helical rise (Å)	2.79	2.96±0.11	3.09±0.15	3.26±0.17
Inclination (°)	7.18	8.80±2.48	4.23±3.02	1.12±3.00
Helical twist (°)	30.49	32.27±0.59	30.79±1.11	32.24±1.10
Propeller (°)	− 6.06	− 12.55±2.47	− 5.43±3.48	− 4.75±2.96
P … P (Å)	6.32	6.14±0.05	6.19±0.08	6.25±0.05
Major groove (Å)	13.27	9.94±0.89	13.69±0.75	13.99±0.95
Glycosidic torsional angle	anti	anti	anti	anti
Sugar pucker	C3′-endo	C3′-endo	C3′-endo	C3′-endo

TFO bind tightly to the duplex. The two tandem units of uracil share this trend (Figures 7H and 7I) and the deeply buried major groove binding of these tandem units is salt independent. Due to its longer length, the end of this TFO (perspective view along the y axis) is more susceptible to excursions and shows higher dynamical fluctuations, especially in Na salt. However, in these longer constructs the tetraloop exhibits less heterogeneity in NaCl (Figure 7K) when compared with ST-TPX. Finally, in parallel with the ST-TPX, the structural heterogeneity of the TFO terminus is restricted in Mg^{2+} (Figure 7L).

We also looked into the differences in the global structural parameters of RNA duplexes alone and in triplexes. Table 1 summarizes the comparison of average duplex features from simulations and experiments. The base pair step parameters are comparable among the structures regardless of buffer condition and sequence, indicating the minimal effect of the TFO and ions on the adjacent base pairs. Remarkably, in contrast to the aforementioned findings, local base pair parameters, such as propeller, capture the conformational change of the duplex region upon the formation of triplex. What is striking in this analysis is the widening of the major groove, which is observed in both the crystal structure and our TPX models. This implies that expanding the major groove of the RNA duplex, i.e., the modification on the duplex geometry, is necessary to accommodate the TFO during the triplex-forming process. Moreover, the phosphorus-base distance (P … P) of ~6.0 Å and the dominance of C3′-endo sugar pucker indicate the retained A-form conformation of duplex regions.

Average helical parameters of RNA ensembles generated from WAXS-driven MD were compared with those of a realistic U•A-U rich triplex.[27] Only the regions designed to be triplex were taken into account.

Spatial cation distributions of duplex and triplex structure are different

A great benefit of these atomically detailed MD simulations is the additional information gleaned about how the spatial distributions of cations respond to accommodate TFO binding. These studies directly address the question of how a negatively charged third strand displaces tightly bound cations from a deep and negatively charged groove. Figure 8 shows the MD-derived average cation occupancy around

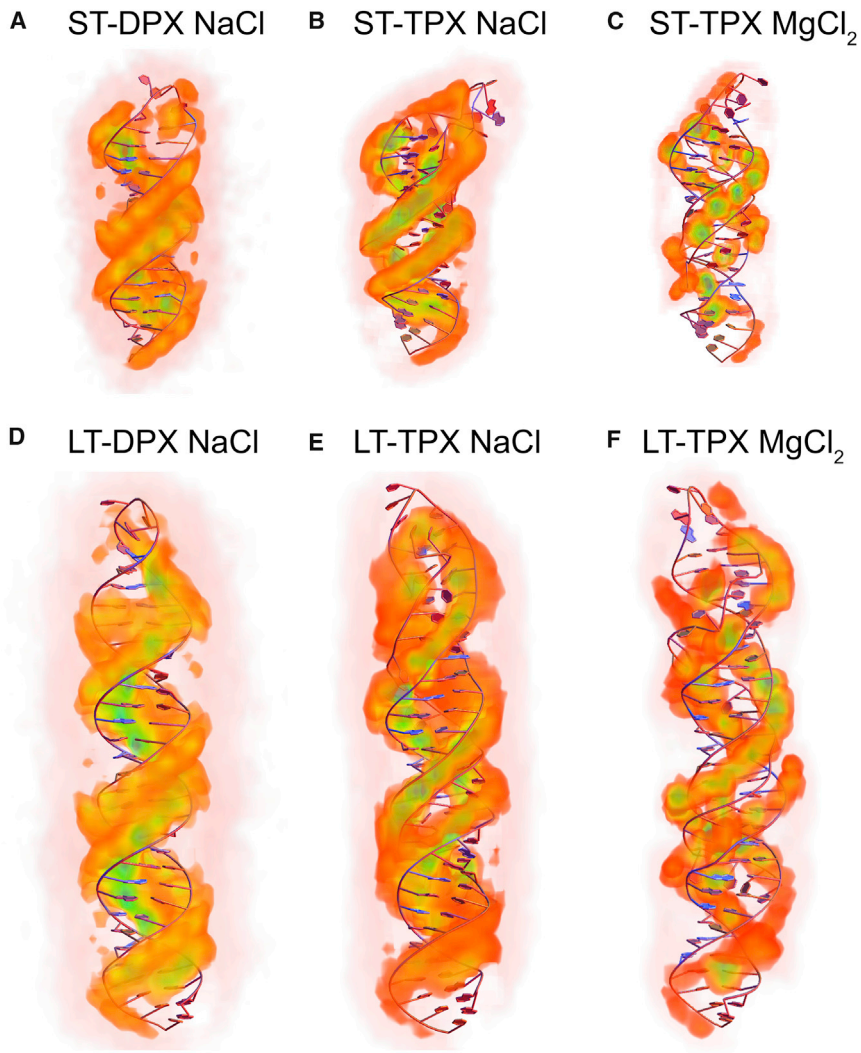


Figure 8. Counterion density profiles from simulations

(A–F) (A) ST-DPX in NaCl, ST-TPX in (B) NaCl and (C) $MgCl_2$, (D) LT-DPX in NaCl, LT-TPX in (E) NaCl and (F) $MgCl_2$, respectively. Increasing ion density is indicated with color: white (low) to red to green (high).

the RNA structures in all of the systems studied. Marked differences are observed between cation distributions around duplex and triplex RNAs. This information is most relevant when the same salt is used, in this case NaCl. Consistent with our earlier works,[51,52] duplex geometries show major groove binding of cations (Figures 8A and 8D), with localization of Na^+ cations to the phosphate backbone. Most surprisingly, the binding of TFO to the major groove does not exclude Na^+ cations from the major groove, rather, it leads to stronger localization of cations into upper and lower layers juxtaposed to the third phosphate backbones (Figures 8B and 8E). The tetraloop region also showed an increase in cation occupancy in the case of the triplex structure in the stem region. This increased localization may result from the higher charge density created by the third strand. Similarly, hexa-hydrated Mg^{2+} cations also bind to the major groove of TPX structures despite the presence of TFO. Localization of divalent ions is more discrete (Figures 8C and 8F), in sharp contrast to the case of Na^+ where binding is territorial and diffusive.

DISCUSSION

Combining atomically detailed simulations with experimental data can be a challenging task. Many factors impact the accuracy of the interpretation of the results,[53–56] such as the resolution of the experimental data, the accurate reflection in the measurement of well-defined structural correlations in the molecules, the type and magnitude of the errors, and finally the accuracy of the back calculation.[57] Computational challenges are also important, especially selection of a good starting model, assessment of overlap between the experimental and computational ensemble(s), and the role of force field and coupling methods on the ensemble generated need to be investigated more in depth to develop a robust computational strategy. Here, we proposed and demonstrated a strategy for solving the solution structure of RNA molecules possessing stacks of base triples in a triplex motif. This framework integrates SaS MD with WAXS-driven forces to direct the simulation into agreement with the experimental data. We applied this framework to solve the structures of biologically important, but structurally uncharacterized RNA triplex motifs. Structural variations, dynamic arrangements, as well as ion effects were accessible to our analysis, and were examined in depth. From these straightforward comparisons, we derived an improved understanding of RNA triplexes at the atomic level. For example, we observed that a like-charged third strand of RNA is readily accommodated by an already formed duplex by widening the major groove, sequence-dependent contact formation, and redistribution and localization of charge-compensating cations. The new charge distribution pattern identified here, in addition to the structural importance of tight U•A-U base triples from tandem uracil residues, likely contributes to function. Our coupled studies also underscore the importance of divalent ions in stabilizing flexible motifs, tightening loops, and lashing down free ends. These interactions may impact long-range molecular recognition and signal transduction in currently unrecognized ways. The application of additional experimental methods, such as UV melting or kinetic analyses of single or combined mutations in the base triplex, can potentially reveal the energetics of these interesting interactions. Studies performed under different ionic conditions could provide additional insight into the vital role of charge-compensating counterions in stabilizing these higher-order structures. These new degrees of freedom inform our understanding of RNA tertiary structure and inspire ongoing studies to more fully characterize the detailed electrostatic environment of these highly charged molecules, which uniquely enables their interactions with select cellular partners.

Our method fills an important void by integrating state-of-the-art wide-angle solution X-ray scattering experiments with MD simulations to investigate ensembles of RNA structures. Although demonstrated for stable RNA triplexes, this approach has the potential to solve structures containing other tertiary motifs. Immediate future applications will extend to dynamic systems that contain flexible regions and smaller numbers of stabilizing base triples.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to the lead contact, Lois Pollack (lp26@cornell.edu).

Materials availability

This study did not generate new unique materials or reagents.

Data and code availability

The data generated during this study are available from the corresponding authors upon reasonable request. All data associated with the reported analyses are included in the paper and/or the supplemental information, and in the Small Angle Scattering Biological Data Bank (SASBDB) (https://www.sasbdb.org/) with the following access codes: SASDKP5 SASDKS5 SASDKV5 SASDKY5 SASDKR5 SASDKU5 SASDKX5 SASDK26.

RNA triple helix motif design

We designed short and long RNA hairpin constructs, called ST-DPX and LT-DPX, by terminating 17-bp and 29-bp duplex stems, respectively, with a UUCG tetraloop. The stems consist of sequence-specific 12-bp and 24-bp binding sites for two TFOs, of 12 and 24 nucleotides in length, called TFO12 and TFO24, forming major-groove RNA triple helical structures, denoted ST-TPX and LT-TPX. The schematic secondary structures of the aforementioned RNA molecules are shown in Figure 1A. See section RNA triple helix motif design in the supplemental Appendix for a detailed description.

RNA triple helix sample preparation

Both ST-DPX and LT-DPX RNA molecules were synthesized using four batches of ten T7 reactions at 37°C from dsDNA templates purchased from IDT (Coralville, IA). We used ethanol precipitation at −80°C to stop the T7 reactions and condense the RNA samples with impurities. The RNA samples were first purified using a Mono Q 5/50 GL anion exchange column (GE Healthcare, Chicago, IL) at pH 7.0 and then annealed in buffers containing 40 mM NaCl, 20 mM sodium 3-(N-morpholino)propanesulfonic acid, 50 μM EDTA (pH 7.0). Half of each ST-DPX and LT-DPX sample was then annealed with the protonated TFO12 and TFO24 strands in 200 mM NaCl, 1.0 mM $MgCl_2$, 20 mM sodium 2-(N-morpholino)ethanesulfonic acid, 50 μM EDTA (pH 5.5) to form ST-TPX and LT-TPX, respectively. ST-TPX and LT-TPX were anion-exchange purified and buffer exchanged to either 200 mM NaCl or 5 mM $MgCl_2$ for solution X-ray scattering experiments. A detailed description of sample preparation and characterization can be found in section RNA triple helix sample preparation of the supplemental Appendix.

Solution X-ray scattering

All solution X-ray scattering data were acquired at the 16-ID (LiX) beamline of the National Synchrotron Light Source II (NSLS II) at Brookhaven National Laboratory.[58] Solution samples (60 μL) and their matching buffers were manually loaded and individually measured in a continuous flow mode with five 2-s exposures. Solution X-ray scattering data with a q range of 0.005 to 3.2 $Å^{-1}$ were recorded using a Pilatus 1M (small angle) and two Pilatus 300K (wide angle) detectors (Dectris, Switzerland) in vacuum. On-site frame-to-frame data screening was done using the *py4xs* package[58] to ensure matching of pre- and post-sample buffers and to ensure the absence of radiation damage. For all measurements, the signal-to-noise ratio (SNR) was above 30.0 at q = 1.0$Å^{-1}$. The SASBDB access codes for experimental profiles used in this study are reported in Table S3.

Calculation of solution X-ray scattering profiles

SWAXS profiles were calculated following the scattering theory from excess electron densities.[44,59,60] We used a probability isosurface to enclose the ion and solvent shells near the RNA molecule with a distance of d = 12 Å from the RNA surface. This distance was determined empirically. From 20 ns long solvent alone system (details in supplemental Appendix, section WAXS computing from MD trajectory)

we select equally spaced frames for every 50 ps and used to estimate the electron density of the bulk solvent in each case. The orientational integral of the scattering amplitudes over 4π solid angle was computed numerically, using 1,750 uniformly distributed q vectors on a sphere,[61] spanning a q range from 0.0 to 1.0Å^{-1} with 201 points.

The customized statistical metric

The agreement between experimentally and computationally acquired solution X-ray scattering profiles was quantified using the customized χ^2_{log} metric:

$$\chi^2_{log} = \frac{1}{n-1}\sum_{i=1}^{n}\left\{\frac{\log_{10}[I_{exp}(q_i)] - \log_{10}[I_{com}(q_i)]}{\sigma'(q_i)}\right\}^2, \qquad \text{(Equation 1)}$$

where n is the number of q points and $I_{com}(q_i)$ and $I_{exp}(q_i)$ are the computed and experimental intensities at q_i, respectively. In Equation 1, $\sigma'(q_i)$ denotes the propagated error at q_i: $\sigma'(q_i) = \sigma(q_i)/[I_{exp}(q_i)\log 10]$ with $\sigma(q_i)$ being the experimental error. χ^2_{log} reduces the weight on the small-angle data and is statistically similar to linear χ^2 given the large SNR (>10.0).[62]

Molecular modeling and simulation methods

Initial models of the ST-TPX, LT-TPX, ST-DPX, and LT-DPX were built using 3D-NuS,[45] Nucleic Acid Builder,[63] and 3dRNA[64] programs. The constructed models were energy minimized using the GROMACS 5.0.5 package[65] with the Amber χOL3 force field.[66,67] Subsequently, the constructs were solvated with water and ions to enable explicit water simulations. To mimic experiments, we prepared the RNAs in solutions containing ions at 200 mM free Na^+ and 5 mM free Mg^{2+} concentrations. We used χOL3 parameters for nucleic acids, Smith and Dang parameters for ions,[68] and modeled water using TIP3P.[69] Details of the general simulation method are summarized in section General MD simulation set up of the supplemental Appendix.

Sample and select approach

A 300-ns-long room temperature (300K) production runs were employed to compute the initial WAXS profiles. Starting from the last frame of the MD simulation we performed a 100-ns-long HT (T = 340K) simulation for each system. The sampled conformations are clustered based on their RMSD using the gromos algorithm with a cutoff of 2.0 Å[70] (see details in supplemental information). WAXS profiles were computed using the method described above for selected cluster centers. We evaluate the fit of each cluster using the metric defined in (Equation 1). The cluster that gives rise to the smallest χ^2 is selected for the WAXS data-driven simulations.

WAXS-driven MD simulations

The incorporation of experimental WAXS data $[I_{exp}(q)]$ to MD simulations was achieved by a Hamiltonian including a hybrid energy $E_{hybrid} = E_{FF} + E_{WAXS}$ term, where E_{FF} is the energy from the MD force field, while E_{WAXS} represents the energetic penalty if the real-time computed amplitude $[I_{comp}(q,t)]$ deviates from the target scattering curve $[I_{exp}(q)]$ of the experiment. The real-time WAXS curves are calculated on-the-fly from the simulation using the method described in Calculation of solution X-ray scattering profiles and in Chen et al.[41] The energy penalty to the deviation between the simulated and experimental WAXS profiles started with a logarithmic non-weighted coupling term,

$$E_{WAXS}^{(log)}(t) = \alpha(t) k_c \frac{k_B T}{n_q} \sum_{i=1}^{n_q} \{\log[I_{com}(q_i,t)] - \log[I_{exp}(q_i)]\}^2. \quad \text{(Equation 2)}$$

Here, n_q is the number of intensity points spanning the specific range of scattering vectors q. The coefficient k_c adjusts the weight of the WAXS potential E_{WAXS} compared with the force field potential E_{FF}. The parameter $\alpha(t)$ is a time-dependent function that allows the gradual introduction of external coupling at the start of the simulation. Later, a non-weighted coupling was employed to reduce the curve deviation at higher q regions, followed by uncertainty-weighted coupling for globally fitting the WAXS signatures. Further details of the WAXS-driven MD approach can be found in the section Coupling experimental data with MD simulations and WAXS-driven simulation set up of the supplemental Appendix.

Pairwise distance correlation analysis

To evaluate the correlation between structural features and WAXS signals, we construct correlation maps (Equation 11 of He et al.[39]). We computed the normalized deviation between logarithmic intensities of simulated and experimental curves. To monitor structural changes, we used normalized deviation of pairwise distances. These distances were measured by picking the phosphate (P) atom of each residue, to which the solution X-ray scattering is most sensitive. The logarithmic form of the intensities allowed us to focus on structural characteristics beyond $d_{min} = 2\pi/q_{max} \approx 5$Å. The normalized deviation at angle q_l was computed by

$$\delta_k(q_l) = \frac{\log[I_{com,k}(q_l)] - \log[I_{exp}(q_l)]}{\sigma'(q_l)}, \quad \text{(Equation 3)}$$

where, $I_{com,k}(q_l)$ and $I_{exp}(q_l)$ are intensities from computation and experiment, respectively. Accordingly, the structural fluctuation of the k-th conformation was assessed using:

$$\delta_k(d_{ij}) = \frac{d_k(p_{ij}) - d_0(p_{ij})}{d_0(p_{ij})}, \quad \text{(Equation 4)}$$

where, $d_k(p_{ij})$ is the pairwise distance between residue i and residue j of the k-th conformation, and $d_0(p_{ij})$ is the averaged pair distance between residue i and residue j in the best-fit ensemble.

Data analysis

The conformations that display $\chi^2_{log} < 10$ (Equation 1) were collected and used for all subsequent analysis. To investigate Loop < – > Duplex and TFO< – > Duplex positioning, we carried out contact analysis (see Residue-residue contact analysis in the supplemental Appendix). To study duplex topology, we estimated the groove width using the 3-DNA program.[71] Later, we align the trajectory to a reference conformation and analyze the cation distribution. The structure of the cation cloud around the helices was analyzed using:

$$\varphi(\mathbf{r}) = \left\langle \frac{1}{V} \sum_i \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle, \quad \text{(Equation 5)}$$

where $\mathbf{r} \equiv (x, y, z)$, describes an arbitrary cation position in Cartesian coordinates, $\varphi(x, y, z)$ is the average cation number density, V is the volume of simulation box, and $\delta(x)$ is the Kronecker delta. The sum is performed over the cation index i, and $\langle \ldots \rangle$ represents the ensemble average.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2022.100971.

ACKNOWLEDGMENTS

This work was supported by NIH grant R35-GM122514 to L.P. Support for work performed at the CBMS beam line LIX (16ID) at NSLS-II is provided by NIH (P30 GM133893, S10 OD012331, and BER- BO 070). NSLS-II is supported by DOE (BES-FWP-PS001). Computational research was carried out on the High Performance Computing resources at New York University Abu Dhabi. S.K. and W.H. are supported by an AD181 faculty research grant. The authors thank Shirish Chodankar and Lin Yang for experimental assistance and Jorge Naranjo for his computational support.

AUTHOR CONTRIBUTIONS

Y.-L.C. performed the experiments and analyzed the data. W.H. performed and analyzed the simulations. S.K. and L.P designed the study. All authors contributed to integrating experimental with computational data and to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 20, 2021
Revised: April 9, 2022
Accepted: June 9, 2022
Published: July 11, 2022

REFERENCES

1. Crooke, S.T., Witztum, J.L., Bennett, C.F., and Baker, B.F. (2018). Rna-targeted therapeutics. Cell Metabol. *27*, 714–739. https://doi.org/10.1016/j.cmet.2018.03.004.

2. Pardi, N., Hogan, M.J., Porter, F.W., and Weissman, D. (2018). mRNA vaccines — a new era in vaccinology. Nat. Rev. Drug Discov. *17*, 261–279. https://doi.org/10.1038/nrd.2017.243.

3. Butcher, S.E., and Pyle, A.M. (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. Accounts Chem. Res. *44*, 1302–1311. https://doi.org/10.1021/ar200098t.

4. Abu Almakarem, A.S., Petrov, A.I., Stombaugh, J., Zirbel, C.L., and Leontis, N.B. (2011). Comprehensive survey and geometric classification of base triples in RNA structures. Nucleic Acids Res. *40*, 1407–1423. https://doi.org/10.1093/nar/gkr810.

5. Conrad, N.K. (2013). The emerging role of triple helices in RNA biology. Wiley Interdisciplinary Reviews: RNA *5*, 15–29. https://doi.org/10.1002/wrna.1194.

6. Brown, J.A. (2020). Unraveling the structure and biological functions of RNA triple helices. WIREs RNA *11*, e1598. https://doi.org/10.1002/wrna.1598.

7. Kim, N.-K., Zhang, Q., Zhou, J., Theimer, C.A., Peterson, R.D., and Feigon, J. (2008). Solution structure and dynamics of the wild-type pseudoknot of human telomerase rna. J. Mol. Biol. *384*, 1249–1261. https://doi.org/10.1016/j.jmb.2008.10.005.

8. Cash, D.D., Cohen-Zontag, O., Kim, N.-K., Shefer, K., Brown, Y., Ulyanov, N.B., Tzfati, Y., and Feigon, J. (2013). Pyrimidine motif triple helix in the kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. Proc. Natl. Acad. Sci. USA *110*, 10970–10975. https://doi.org/10.1073/pnas.1309590110.

9. Jiang, J., Chan, H., Cash, D.D., Miracco, E.J., Ogorzalek Loo, R.R., Upton, H.E., Cascio, D., O'Brien Johnson, R., Collins, K., Loo, J.A., et al. (2015). Structure of tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. Science *350*, aab4070. aab4070–aab4070. https://doi.org/10.1126/science.aab4070.

10. Nguyen, T.H.D., Tam, J., Wu, R.A., Greber, B.J., Toso, D., Nogales, E., and Collins, K. (2018). Cryo-EM structure of substrate-bound human telomerase holoenzyme. Nature *557*, 190–195. https://doi.org/10.1038/s41586-018-0062-x.

11. Theimer, C.A., Blois, C.A., and Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol. Cell *17*, 671–682. https://doi.org/10.1016/j.molcel.2005.01.017.

12. Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318–323. https://doi.org/10.1038/nature21079.

13. Haack, D.B., Yan, X., Zhang, C., Hingey, J., Lyumkis, D., Baker, T.S., and Toor, N. (2019). Cryo-EM structures of a group II intron reverse splicing into DNA. Cell *178*, 612–623.e12. e12. https://doi.org/10.1016/j.cell.2019.06.035.

14. Brown, J.A., Bulkley, D., Wang, J., Valenstein, M.L., Yario, T.A., Steitz, T.A., and Steitz, J.A. (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat. Struct. Mol. Biol. *21*, 633–640. https://doi.org/10.1038/nsmb.2844.

15. Torabi, S.-F., Vaidya, A.T., Tycowski, K.T., DeGregorio, S.J., Wang, J., Shu, M.-D., Steitz, T.A., and Steitz, J.A. (2021a). RNA stabilization by a poly(a) tail 3'-end binding pocket and other modes of poly(a)-RNA interaction. Science *371*, eabe6523.

16. Torabi, S.-F., Chen, Y.-L., Zhang, K., Wang, J., DeGregorio, S.J., Vaidya, A.T., Su, Z., Pabit, S.A., Chiu, W., Pollack, L., and Steitz, J.A. (2021b). Structural analyses of an RNA stability element interacting with poly(a). Proc. Natl. Acad. Sci. USA *118*. e2026656118. https://doi.org/10.1073/pnas.2026656118.

17. Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A., and Steitz, J.A. (2010). Poly(a) tail recognition by a viral RNA element through assembly of a triple helix. Science *330*, 1244–1247. https://doi.org/10.1126/science.1195858.

18. Pham, V.V., Salguero, C., Khan, S.N., Meagher, J.L., Brown, W.C., Humbert, N., de Rocquigny, H., Smith, J.L., and D'Souza, V.M. (2018). HIV-1 tat interactions with cellular 7sk and viral TAR RNAs identifies dual structural mimicry. Nat. Commun. *9*, 4266. https://doi.org/10.1038/s41467-018-06591-6.

19. Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Groß, M., et al. (2012). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res. *73*, 1180–1189. https://doi.org/10.1158/0008-5472.can-12-2850.

20. Conrad, N.K., Mili, S., Marshall, E.L., Shu, M.-D., and Steitz, J.A. (2006). Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. Mol. Cell *24*, 943–953. https://doi.org/10.1016/j.molcel.2006.10.029.

21. Conrad, N.K., Shu, M.-D., Uyhazi, K.E., and Steitz, J.A. (2007). Mutational analysis of a viral RNA element that counteracts rapid RNA decay by interaction with the polyadenylate tail. Proc. Natl. Acad. Sci. USA *104*, 10412–10417. https://doi.org/10.1073/pnas.0704187104.

22. Wang, J., Reiss, K., Shi, Y., Lolis, E., Lisi, G.P., and Batista, V.S. (2021). Mechanism of inhibition of the reproduction of SARS-CoV-2 and ebola viruses by remdesivir. Biochemistry *60*, 1869–1875. https://doi.org/10.1021/acs.biochem.1c00292.

23. Liberman, J.A., Suddala, K.C., Aytenfisu, A., Chan, D., Belashov, I.A., Salim, M., Mathews, D.H., Spitale, R.C., Walter, N.G., and Wedekind, J.E. (2015). Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. Proc. Natl. Acad. Sci. USA *112*, E3485–E3494. https://doi.org/10.1073/pnas.1503955112.

24. Huang, L., Wang, J., Wilson, T.J., and Lilley, D.M. (2017). Structure of the guanidine iii riboswitch. Cell Chemical Biology *24*, 1407–1415.e2. https://doi.org/10.1016/j.chembiol.2017.08.021.

25. Huang, L., and Lilley, D.M.J. (2018). Structure and ligand binding of the SAM-v riboswitch. Nucleic Acids Res. *46*, 6869–6879. https://doi.org/10.1093/nar/gky520.

26. Devi, G., Zhou, Y., Zhong, Z., Toh, D.-F.K., and Chen, G. (2015). Rna triplexes: from structural principles to biological and biotech applications. Wiley Interdisciplinary Reviews: RNA *6*, 111–128. https://doi.org/10.1002/wrna.1261.

27. Ruszkowska, A., Ruszkowski, M., Hulewicz, J.P., Dauter, Z., and Brown, J.A. (2020). Molecular structure of a u· au-rich rna triple helix with 11 consecutive base triples. Nucleic Acids Res. *48*, 3304–3314. https://doi.org/10.1093/nar/gkz1222.

28. Salsbury, A.M., and Lemkul, J.A. (2021). Recent developments in empirical atomistic force fields for nucleic acids and applications to studies of folding and dynamics. Folding Dynamics *67*, 9–17. https://doi.org/10.1016/j.sbi.2020.08.003.

29. Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R.A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., et al. (2018). Rna structural dynamics as captured by molecular simulations: a comprehensive overview. Chem. Rev. *118*, 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.

30. Mlýnský, V., Kührová, P., Kühr, T., Otyepka, M., Bussi, G., Banas, P., Šponer, J., and Sponer, J. (2020). Fine-tuning of the amber rna force field with a new term adjusting interactions of terminal nucleotides. J. Chem. Theor. Comput. *16*, 3936–3946. https://doi.org/10.1021/acs.jctc.0c00228.

31. Tan, D., Piana, S., Dirks, R.M., and Shaw, D.E. (2018). Rna force field with accuracy comparable to state-of-the-art protein force fields. Proc. Natl. Acad. Sci. USA *115*, E1346–E1355. https://doi.org/10.1073/pnas.1713027115.

32. Yang, S., Parisien, M., Major, F., and Roux, B. (2010). RNA Structure Determination Using SAXS Data 114, pp. 10039–10048.

33. Bottaro, S., Bussi, G., Kennedy, S.D., Turner, D.H., and Lindorff-Larsen, K. (2018). Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. Sci. Adv. *4*, eaar8521. https://doi.org/10.1126/sciadv.aar8521.

34. Shi, H., Rangadurai, A., Assi, H.A., Roy, R., Case, D.A., Herschlag, D., Yesselman, J.D., and Al-Hashimi, H.M. (2020). Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. Nat. Commun. *11*, 1–14.

35. Bottaro, S., Bussi, G., and Lindorff-Larsen, K. (2021). Conformational ensembles of noncoding elements in the sars-cov-2 genome from molecular dynamics simulations. J. Am. Chem. Soc. *143*, 8333–8343. https://doi.org/10.1021/jacs.1c01094.

36. Dagenais, P., Desjardins, G., and Legault, P. (2021). An integrative NMR-SAXS approach for structural determination of large RNAs defines the substrate-free state of a trans-cleaving neurospora varkud satellite ribozyme. Nucleic Acids Res. *49*, 11959–11973.

37. Kappel, K., Zhang, K., Su, Z., Watkins, A.M., Kladwang, W., Li, S., Pintilie, G., Topkar, V.V., Rangan, R., Zheludev, I.N., et al. (2020). Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. Nat. Methods *17*, 699–707. https://doi.org/10.1038/s41592-020-0878-9.

38. Bernetti, M., Hall, K.B., and Bussi, G. (2021). Reweighting of molecular simulations with explicit-solvent saxs restraints elucidates ion-dependent RNA ensembles,. Preprint at arXiv. https://doi.org/10.48550/arXiv.2103.04964.

39. He, W., Chen, Y.-L., Pollack, L., and Kirmizialtin, S. (2021). The structural plasticity of nucleic acid duplexes revealed by waxs and md. Sci. Adv. *7*, eabf6106. https://doi.org/10.1126/sciadv.abf6106.

40. He, W., Henning-Knechtel, A., and Kirmizialtin, S. (2022). Visualizing RNA structures by SAXS-driven MD simulations. Frontiers in Bioinformatics *2*. https://doi.org/10.3389/fbinf.2022.781949.

41. Chen, P.c., and Hub, J. (2015). Interpretation of solution x-ray scattering by explicit-solvent molecular dynamics. Biophys. J. *108*, 2573–2584. https://doi.org/10.1016/j.bpj.2015.03.062.

42. Chen, Y.-L., and Pollack, L. (2019). Salt dependence of a-form RNA duplexes: Structures and implications. J. Phys. Chem. B *123*, 9773–9785.

43. Chen, Y.-L., and Pollack, L. (2020). Machine learning deciphers structural features of RNA duplexes measured with solution X-ray scattering. IUCrJ *7*, 870–880.

44. Knight, C.J., and Hub, J.S. (2015). WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. Nucleic Acids Res. *43*, W225–W230. https://doi.org/10.1093/nar/gkv309.

45. Patro, L.P.P., Kumar, A., Kolimi, N., and Rathinavelan, T. (2017). 3D-NuS: a web server for automated modeling and visualization of non-canonical 3-D imensional nucleic acid Structures. J. Mol. Biol. *429*, 2438–2448. https://doi.org/10.1016/j.jmb.2017.06.013.

46. Mráziková, K., Mlýnský, V., Kührová, P., Pokorná, P., Kruse, H., Krepl, M., Otyepka, M., Banas, P., Šponer, J., and Sponer, J. (2020). Uucg rna tetraloop as a formidable force-field challenge for md simulations. J. Chem. Theor. Comput. *16*, 7601–7617. https://doi.org/10.1021/acs.jctc.0c00801.

47. Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of rna base pairs. RNA *7*. S1355838201002515–512. https://doi.org/10.1017/s1355838201002515.

48. Arnott, S., and Bond, P.J. (1973). Structures for poly(u).poly(a).poly(u) triple stranded polynucleotides. Nat. New Biol. *244*, 99–101. https://doi.org/10.1038/newbio244099a0.

49. Arnott, S., Bond, P., Selsing, E., and Smith, P. (1976). Models of triple-stranded polynucleotides with optimised stereochemistry. Nucleic Acids Res. *3*, 2459–2470. https://doi.org/10.1093/nar/3.10.2459.

50. Hoogsteen, K. (1959). The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta Crystallogr. *12*, 822–823. https://doi.org/10.1107/s0365110x59002389.

51. Kirmizialtin, S., and Elber, R. (2010). Computational exploration of mobile ion distributions around RNA duplex. J. Phys. Chem. B *114*, 8207–8220. https://doi.org/10.1021/jp911992t.

52. Kirmizialtin, S., Pabit, S.A., Meisburger, S., Meisburger, S.P., Pollack, L., and Elber, R. (2012). RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations. Biophys. J. *102*, 819–828. https://doi.org/10.1016/j.bpj.2012.01.013.

53. Vendruscolo, M. (2018). Principles of protein structural ensemble determination. Biophys. J.

114, 388a–389a. https://doi.org/10.1016/j.bpj.2017.11.2149.

54. Bottaro, S., and Lindorff-Larsen, K. (2018). Biophysical experiments and biomolecular simulations: a perfect match? Science *361*, 355–360. https://doi.org/10.1126/science.aat4010.

55. Chen, P.-c., Shevchuk, R., Strnad, F.M., Lorenz, C., Karge, L., Gilles, R., Stadler, A.M., Hennig, J., and Hub, J.S. (2019). Combined small-angle x-ray and neutron scattering restraints in molecular dynamics simulations. J. Chem. Theor. Comput. *15*, 4687–4698. https://doi.org/10.1021/acs.jctc.9b00292.

56. Lindorff-Larsen, K. (2020). Combining biophysical experiments and biomolecular simulations. Biophys. J. *118*, 321a. https://doi.org/10.1016/j.bpj.2019.11.1804.

57. Bernetti, M., and Bussi, G. (2021). Comparing state-of-the-art approaches to back-calculate saxs spectra from atomistic molecular dynamics simulations. Eur. Phys. J. B *94*, 180. https://doi.org/10.1140/epjb/s10051-021-00186-9.

58. Yang, L., Antonelli, S., Chodankar, S., Byrnes, J., Lazo, E., and Qian, K. (2020). Solution scattering at the life science x-ray scattering (LiX) beamline. J. Synchrotron Radiat. *27*, 804–812. https://doi.org/10.1107/s1600577520002362.

59. Park, S., Bardhan, J.P., Roux, B., and Makowski, L. (2009). Simulated x-ray scattering of protein solutions using explicit-solvent models. J. Chem. Phys. *130*, 134114. https://doi.org/10.1063/1.3099611.

60. Chen, P.c., and Hub, J. (2014). Validating solution ensembles from molecular dynamics simulation by wide-angle x-ray scattering data. Biophys. J. *107*, 435–447. https://doi.org/10.1016/j.bpj.2014.06.006.

61. Ponti, A. (1999). Simulation of magnetic resonance static powder lineshapes: a quantitative assessment of spherical codes. J. Magn. Reson. *138*, 288–297. https://doi.org/10.1006/jmre.1999.1758.

62. Chen, Y.-L., and Pollack, L. (2020). In vitro electron density refinement from solution x-ray scattering in the wide-angle regime. Preprint at arXiv. https://doi.org/10.48550/arXiv.2012.13370.

63. Case, D.A., Cheatham, T.E., III, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The amber biomolecular simulation programs. J. Comput. Chem. *26*, 1668–1688. https://doi.org/10.1002/jcc.20290.

64. Zhao, Y., Huang, Y., Gong, Z., Wang, Y., Man, J., and Xiao, Y. (2012). Automated and fast building of three-dimensional rna structures. Sci. Rep. *2*, 734. https://doi.org/10.1038/srep00734.

65. Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theor. Comput. *4*, 435–447. https://doi.org/10.1021/ct700301q.

66. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the amber ff99sb protein force field. Proteins: Struct., Funct., Bioinf. *78*, 1950–1958. https://doi.org/10.1002/prot.22711.

67. Zgarbová, M., Otyepka, M., Šponer, J., Šponer, J., Mládek, A., Mládek, A., Banáš, P., Cheatham, T.E., Jurečka, P., and Jurečka, P. (2011). Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J. Chem. Theor. Comput. *7*, 2886–2902. https://doi.org/10.1021/ct200162x.

68. Smith, D.E., and Dang, L.X. (1994). Computer simulations of nacl association in polarizable water. J. Chem. Phys. *100*, 3757–3766. https://doi.org/10.1063/1.466363.

69. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935. https://doi.org/10.1063/1.445869.

70. Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W.F., and Mark, A.E. (1999). Peptide folding: when simulation meets experiment. Angew. Chem. Int. Ed. *38*, 236–240. https://doi.org/10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.0.co;2-m.

71. Lu, X.-J., and Olson, W.K. (2008). 3dna: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat. Protoc. *3*, 1213–1227. https://doi.org/10.1038/nprot.2008.104.