The Salt Dependence of A-form RNA Duplexes: Structures and Implications

Yen-Lin Chen, **Lois Pollack**[*]

School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853

Abstract

The biological functions of RNA range from gene regulation through catalysis and depend critically on its structure and flexibility. Conformational variations of flexible, non-base paired components, including RNA hinges, bulges or single-stranded tails, are well documented. Recent work has also identified variations in the structure of ubiquitous, base-paired duplexes found in almost all functional RNAs. Duplexes anchor the structures of folded RNAs, and their surface features are recognized by partner molecules. To date, no consistent picture has been obtained that describes the range of conformations assumed by RNA duplexes. Here, we apply wide angle, solution x-ray scattering (WAXS) to quantify these variations, by sampling length scales characteristic of helical geometries under different solution conditions. To identify the radius, helical rise, twist and length of dsRNA helices we exploit molecular dynamics generated structures, explicit solvent models and ensemble optimization methods. Our results quantify the substantial and salt-dependent deviations of double-stranded (ds) RNA duplexes from the assumed canonical A-form conformation. Recent experiments underscore the need to properly describe the structures of RNA duplexes when interpreting the salt dependence of RNA conformations.

Graphical Abstract

---
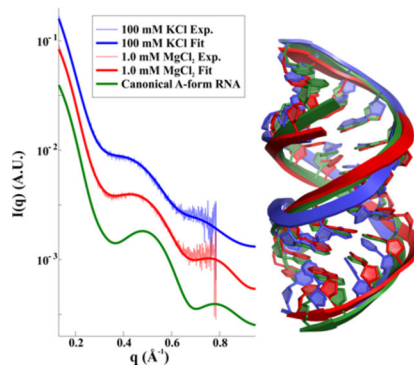
[*]To whom correspondence should be addressed. lp26@cornell.edu Phone: 607-255-8695.
SUPPORTING INFORMATION
Computational Details, Supplementary Data and Figures

INTRODUCTION

RNA's functions, conformations and dynamics are encoded in molecular sequences and regulated by environments and binding partners such as other RNA domains, proteins and metabolites. These functions are intimately related to structure. The most recognizable aspect of RNA structure is the short duplex, stabilized by base pairing between two antiparallel strands. Because RNA duplexes resemble the A-form of dsDNA found in crystallographic studies near the end of last century[1,2], they have been assigned an A-form structure, which is shorter and wider than the canonical B-form of most dsDNA. Interestingly, significant variation from the A-form has been observed in NMR[3] or crystal structures.

The hierarchical nature of RNA structures suggests that dsRNA duplexes serve as important building blocks of more complex RNAs, specifically those with tertiary structures and folds. Efforts to ascertain the degrees of structural freedom have so far focused primarily on the conformations of non-base paired, flexible RNA junctions, bulges and single-stranded (ss) tails, often flanked by dsRNA duplexes, as they would be found in biology. The following studies have substantially extended our knowledge of flexible RNA components or the topology of connectors that join duplexes and contributed insights to our understanding of the RNA folding problem. Chu et. al. investigated the conformational biases of single and double stranded (polyethylene-glycol) junctions by incorporating tertiary contacts on RNA duplexes[4]. Daher et. al. focused on RNA stability and functionality by varying junction or bulge designs[5]. Shi et. al. determined conformations from a bulged RNA duplex from X-ray scattering interferometry[6]. High-throughput thermodynamic analyses of RNA junctions reveal conformational preferences within RNA stems of various lengths[7]. In addition to sequence and architecture, RNA conformational subtleties can also be induced by changing ionic conditions[8]. A main conclusion of this work focusing on flexible RNA regions suggests that they serve to preposition tertiary contacts for engagement.

Here, we identify another important consideration in the orientational positioning of RNA structures: conformational variation of fully base-paired dsRNA duplexes. Recent work reports the structural subtleties of dsRNA[9] and their potential impact on tertiary contact formation. A length variation of dsRNA duplexes was found to be mediated by divalent and multivalent ions[10]. Partial folding of a well-studied RNA domain (P4-P6 from the Tetrahymena ribozyme) and a variant, was attributed to the misalignment of tertiary contacts resulting from non-native helical orientation[11].

Wide-angle X-ray scattering (WAXS) experiments report structural variations on length scales of ~10 Å, shorter than their counterpart small angle x-ray scattering (SAXS), and are straightforward to set up. Because the helical features (e.g. width and periodicity) of dsRNA duplexes are on this length scale, WAXS is an ideal probe of their structures. For example, any disruption of structural periodicity results in measurable changes in WAXS profiles. SAXS or WAXS data can be qualitatively interpreted by focusing on changes of the scattering profiles or quantitatively interpreted by fitting high-resolution structural models to the data. The latter is more informative, but can suffer from many challenges, such as overfitting, lack of validation and failure to account for experimental errors[12].

Unconstrained MD simulations have been applied to provide quantitative structural information about systems containing RNA, providing insightful results in solvent-RNA interactions[13], the varied flexibilities of dsRNA and dsDNA duplexes[14,15] and the salt- and sequence-dependent rigidity of dsRNA[16]. To compensate for imperfect force fields, experimental data have been used as a restraint, to refine conformations from unbiased MD simulations. Many studies have exploited this approach, applying different refinement methods and optimization algorithms[17–22], yet may still be subject to the challenges and problems mentioned[12]. Here, we revise and improve this approach to identify the conformations of dsRNA duplex in different salt conditions, with a focus on addressing these challenges. We begin with MD generated structures of unconstrained duplexes, then use the theory of three-dimensional reference interaction site model (3D-RISM) to model the explicit solvent and ions around the dsRNA duplex for accurate calculation of the small- and wide-angle X-ray scattering (SWAXS) profiles. Finally, we apply ensemble optimization to identify those dsRNA structures which best recapitulate the experimental measurement, using 5-fold cross validation (CV) and tests on different signal-to-noise ratios (S/N) to support the robustness of our results. We find that duplex structures deviate substantially from canonical A-form geometries in a salt dependent manner. The forms assumed by these RNA helices are more diverse in KCl than in $MgCl_2$ solutions. Unlike DNA duplexes, whose conformations appear independent of salt over conditions probed here, the variability of dsRNA duplex *in vitro* must be accounted for when modeling the conformational dynamics of RNA, due to substantial deviations from the assumed rigid and fixed forms of a right-handed A-form helix.

MATERIAL AND METHODS

RNA and DNA Samples and SWAXS Measurements

Two mixed-sequence RNA duplexes of lengths 12 and 25 base-pairs (bp) (RNA12 and RNA25 respectively) were used for this study. The sequence for RNA12 is adopted from one helix of the helix-junction-helix construct of Ref[8] : CCU CCU AAU CGC and its complement. The sequences for RNA25 and DNA25 are adopted from previous studies[9,10,24]: GCA UCU GGG CUA UAA AAG GGC GUC G, and its complement; for DNA, T replaces U.

Sample acquisition and preparation procedures were described in Ref[25]. All samples employed to measure variations as a function of monovalent salt concentration are held in buffers containing 10 mM potassium 3-(N-morpholino) propanesulfonic acid (K-MOPS) and 20 μM Ethylenediaminetetraacetic acid (EDTA). Samples for the divalent salt series contained an additional 100 mM KCl background. The RNA25 sample was buffer exchanged into solutions containing additional salts: 100 mM NaCl, 400 mM KCl and 10.0 mM $MgCl_2$. The RNA12 sample was buffer exchanged into solutions containing 30, 50, 100, 200, 500 mM KCl and 0.25, 0.5, 1.0, 5.0 mM $MgCl_2$. Flow throughs were monitored during dialysis, and the total loss of RNA in both samples was less than 3%. Final RNA concentrations were between 600 and 1100 μM and samples were diluted further, if necessary, at the beamline. Different sample concentrations (100, 150 and 250 μM for

RNA25 and 200, 400, 600 μM for RNA12) were tested to achieve the optimal S/N without radiation damage.

Small-and-wide angle x-ray scattering experiments were conducted at the Cornell High Energy Synchrotron Source (CHESS) beamline G1 using two different experimental setups. The scalar value of the X-ray momentum transfer q is defined as $q = (4\pi/\lambda) \times \sin(2\theta/2)$, where λ is the incident X-ray wavelength and 2θ is the scattering angle. Scattering profiles from RNA25 samples were recorded on an EigerX 1M detector (Dectris AG, Switzerland) with a sample-to-detector distance of about 0.43 meters, to achieve a q range of 0.0312 to 1.7223 Å^{-1}. SAXS and WAXS profiles of RNA12 samples were acquired simultaneously using two PILATUS 100K detectors (Dectris AG, Switzerland) with sample-to-detector distances of 1.7 (SAXS) and 0.45 (WAXS) meters. The SAXS and WAXS profiles were matched in regions of overlapping q: from 0.210 to 0.271 Å^{-1}. These coupled measurements recorded data over a q range from 0.008 to 0.786 Å^{-1}, with a minimum correlation length of $d_{min} = \frac{2\pi}{q_{max}} = 8.00$ Å. The SWAXS profiles were integrated and averaged from two sets of 20, 10-second exposures. Real-time SWAXS analysis was performed with the RAW[26] software package. Measurements for each sample-salt combination were repeated twice within each beamtime and repeated during other beamtimes to ensure reproducibility (see Figs. S8–S9). For the WAXS data, a solid angle correction was used to account for the flat detector panel. We scaled and averaged all the scattering profiles acquired, using CorMap[27] to assess radiation damage and to assure appropriate scaling. We observed systematic shifts at low-q due to inter-particle interactions mediated by ionic strength[24]; data below $q = 0.1$ Å^{-1} were discarded to ensure the absence of these effects. In-house *MATLAB* scripts were employed for off-line, detailed SWAXS data processing.

Extended Ensemble Optimization Method with 5-fold Cross Validation

The Ensemble Optimization Method[17,18] (EOM) has been applied to selectively refine conformations from structural pools modeling SAXS data. EOM has achieved success for flexible systems[20,28–32], bimodal systems[25] and others. However, the following major caveats must be resolved before applying traditional EOM to SWAXS profiles. First, EOM depends explicitly on the program CRYSOL[33] to account for the hydration shells of biomolecules. Although effective for computing the scattering profiles of proteins in the small angle regime, CRYSOL uses simplified description of solvents and does not accurately compute solvent contributions at higher angles. In addition it ignores the ion atmosphere of nucleic acids[12,34–39]. Second, as written, EOM can be applied up to a maximum allowed value, $q = 0.5$ Å^{-1}. Third, the intensity of scattering to wide-angles is typically reduced by two orders of magnitude relative to the small-angle signal, therefore; the traditional χ^2 criterion in the following equation (used to evaluate the quality of an EOM fit) will heavily bias the low-q data.

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^{K} \left(\frac{\mu I(q_j) - I_{exp}(q_j)}{\sigma(q_j)} \right)^2 \qquad (1)$$

In Eq. (1), I and I_{exp} represent the ensemble and experimental SAXS profiles with K data points acquired at q_j; σ and μ represent the experimental error and scaling constant. Finally, as discussed in multiple articles where EOM is implemented to interpret the x-ray scattering profile of different macromolecules, it suffers greatly from the issues of non-uniqueness and overfitting[40], given the low information content (few Shannon channels present) in the SAXS profile[12,41–43]. Improvements are required to extend EOM for analysis of wide-angle data. See Supporting Material for a discussion of parameters that guide our use of EOM.

We resolved the first and second caveats using the 3D-RISM formulation discussed in the Supporting Material to account for explicit solvent and ion models, and to remove EOM's pre-set limit on the q range. The parameters for 3D-RISM calculation for nucleic acids are well explored and documented in Ref[36]. We followed an explicit solvent formulation[34,37,38] to calculate SWAXS profiles of each conformation under a specific ion condition. For interested readers, details about the formulation and algorithm can also be found in Supporting Material. We tackled the third caveat by introducing a new χ^2, defined as follows:

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^{K} \left\{ \frac{log_{10}\left[fI(q_j) + c\right] - log_{10}\left[I_{exp}(q_j)\right]}{\sigma'(q_j)} \right\}^2, \tag{2}$$

where f and c are scaling factor and constant fitting parameter[38,44]. The term σ' represents the propagated experimental error:

$$\sigma'(q_j) = \left| \frac{\sigma(q_j)}{I_{exp}(q_j) \log 10} \right| = \frac{1}{(S/N)_j \log 10} \tag{3}$$

in this equation, $(S/N)_j$ is the signal-to-noise ratio at q_j. Note that Eq. (2) resembles Eq. (13) in Ref[44] except that we choose to apply two constants f and c to the calculated intensity from the ensemble. The uniqueness and overfitting issues remain challenging and are data- and sample-dependent. However, because we incorporate wider-angle data while decreasing our ensemble size with only one free fitting parameter c, the chance of overfitting is significantly reduced. To provide further confidence we introduced a 5-fold cross validation method (discussed in more detail below) to support the uniqueness of the ensembles found.

To carry out the ensemble selection, we implemented a genetic algorithm similar to the traditional EOM method described in Ref[17]. Following the same genetic algorithm nomenclature, there were $N = 5$ genes (SWAXS profiles) in a(n) chromosome (ensemble) and we chose the number of chromosomes (population) to be $K = 50$. The same processes of random mutation and crossing were performed within each generation. We then averaged the curves in every chromosome using $I(q) = \frac{1}{N}\sum_{n=1}^{N} I_n(q)$ and evaluated the fitness scores, χ^2, for all $3K$ chromosomes using Eq. (2). Note that low fitness scores were preferred. At the end of each generation, we sampled K survivors into the next generation from $3K$ chromosomes according to the following softmax function:

$$Pr(k) = e^{-3\chi_k^2} \Bigg/ \sum_{i=1}^{3K} e^{-3\chi_i^2}, \tag{4}$$

where $Pr(k)$, $k = 1,2,3 \dots 3K$, is the probability that the kth chromosome survives this generation and χ_k^2 is the corresponding fitness score. This extended EOM algorithm ran with 10^6 generations and 10^3 repetitions (cycles) and output the best chromosome(s) with the lowest fitness score(s) throughout all generations and cycles.

WAXS signals are weaker than SAXS signals and experimental noise can potentially undermine data modeling and interpretation. We applied a cross-validation strategy to test the eEOM methodology against q-regions with different experimental noise. This cross validation also demonstrates the robustness of our results for RNA12 conformations. Cross-validation methods have been applied to refine real-space electron densities from X-ray diffraction[45] and cryo-EM data[46]. Here, we applied 5-fold cross validation (CV) as follows. We uniformly cut our experimental SWAXS profiles into 5 segments and excluded one segment from full data set each time as the truncated SWAXS profile. Five truncated SWAXS profiles, I_l, where $l = 1,2,3,4,5$, were fed into the extended EOM algorithm, generating 5 sets of ensembles. The errors, E_l, were evaluated using Eq. (2) on the excluded data set and averaged: $\bar{E} = \sum_{l=1}^{5} E_l$. This 5-fold CV was repeated 100 times. We then compared these 5 sets of ensembles with the ensemble(s) obtained from the full data set (denoted as I_l, where $l = 0$). For RNA12 in all the salt concentrations, these 6 ensembles are identical from every repetition, except for 0.5 mM $MgCl_2$ whose CV results were not consistent with significantly larger χ^2 due to poor data quality and are thus dropped from our results and discussions. The flowcharts are shown in Fig. 2. The algorithm was written as modules and functions in *Julia*[47].

RESULTS

Solution X-ray Scattering Experiments and Interpretation

Solution X-ray scattering is a powerful tool for investigating macromolecular conformations *in vitro*. Small-angle X-ray scattering (SAXS), with a resolution of 20–100 Å, typically provides low-resolution structural details[48–50] including radius of gyration (R_g) and compactness. Measurements at wider scattering angles are required to reveal features on shorter length scales. These wide-angle X-ray scattering (WAXS) experiments are performed less extensively, despite the straightforward experimental setup and sharper (3–5 times higher) spatial resolution. At these wide angles, shorter range structural periodicities can be detected, adding to the useful information readily available from low angle data. WAXS data also add additional Shannon channels (information content)[41,42] which significantly reduce the chance of overfitting when models are applied to fit the data. WAXS has successfully been applied to study solution crowding effects[51], conformational changes[10,52], and protein ensembles from MD simulations[37,38,44].

For this work, we carried out SAXS and WAXS (or SWAXS) experiments to measure the structure of RNA12 in buffered solutions containing different concentrations of KCl and

mixed KCl/MgCl$_2$, shown in Fig. 3(a,c) respectively. Three key challenges must be overcome to interpret these data. First, the S/N ratio is low in the wide-angle regime; the WAXS signal is about 10–100 times weaker than SAXS signals. Thus, a higher concentration of sample is needed for measurements that achieve sharper real-space resolution. Second, forward modeling of WAXS data is challenging and computationally expensive. WAXS profiles contain information from ion atmospheres and solvent in addition to the macromolecules. Explicit ions and solvent models with excluded volumes are required to compute theoretical WAXS profiles with good accuracy[34,37]. Finally, in contrast to SAXS data analysis, there are few tools available to extract useful information from the experimental WAXS profiles and the analysis, we believe, is case-dependent at this point.

Here, we address each of the above challenges and demonstrate that we can extract valid and quantitative models whose scattering profiles recapitulate the measured SWAXS curves. We apply the three-dimensional reference interaction site model (3D-RISM)[53] to calculate locations of explicit solvent molecules and ions (both near the solute and in bulk solution) around various RNA12 conformations (see Supporting Materials: 3D-RISM Calculations). With the fully solvated and charge compensated RNA12 duplexes, theoretical SWAXS profiles are computed[34] (see Supporting Materials: Calculations of Theoretical SWAXS Profiles Using Explicit Solvent and Ions). The refined RNA12 ensemble in a particular solution (salt concentrations) is determined using the extended ensemble optimization method (eEOM) described in the Material and Methods to select models that recapitulate measurement.

Fitting Results and Refined Structures

The fitting results for SWAXS curves from the RNA12 sample, using the structural refinement by eEOM, are shown in Fig. 3(a,c). Each curve (color) reflects a different concentration of KCl and MgCl$_2$ in the buffer, and the χ^2 value for each fit is reported in the caption. The curves are plotted on a logarithmic scale and offset for visualization to emphasize the changing features in the wider-angle regime. Good agreement between fit and measurement is observed at q values up to ~0.8 Å^{-1}.

The dominant molecular structures derived from the refined ensembles are shown in Fig. 3(b,d), using the color scheme of Fig. 3(a,c). Only the backbones are drawn. The green structure represents the canonical A-form dsRNA duplex of the same sequence built from online nucleic acid builder (NAB)[54], which serves as the alignment reference for all structures. The ensembles are homogeneous, containing one or at most two very similar conformations (Figs. S2, S3). Importantly, we find that the RNA12 duplex does not assume the A-form under any of conditions tested (with 2–5 Å all-atom RMSD). A closer examination of the structures suggests that the backbone helical geometry of RNA12 varies substantially in KCl containing solutions. In MgCl$_2$ containing solutions, the duplex is better defined with structures that display tighter major and minor grooves regardless of the ionic strengths. Overall, Mg^{2+} ions make RNA12 shorter and more compact compared to K$^+$ ions, and the deviations from canonical A-form are more pronounced at the ends.

Different RNA Duplex Conformations Revealed

To quantify our observations of the RNA12 duplex, we used the *x3dna-dssr* program[55] to extract the helical parameters of the selected ensemble. These SWAXS experiments report crucial parameters that define helical structures, such as total helical extension and total helical twist, in addition to details related to structural periodicity such as the average helical radius and rise. Finer subtleties, for example, base pair geometries and phosphodiester bonds are below the theoretical resolution of our experimental setup, yet they can as well be extracted or inferred[56] from measurements at higher q. Fig. 4 maps this multi-dimensional conformational space onto two 2D parameter spaces: average helical radius vs. rise (Fig. 4(a, b)), and helical extension vs. twist (Fig. 4(c, d)) using the color schemes of Fig. 3 to represent different KCl and $MgCl_2$ concentrations. We chose these parameters because they represent a helical geometry and can be extracted by *x3dna-dssr* program. The canonical A-form is shown as green diamond in these figures. The full structural pool, e.g. all the structures used for eEOM refinement, are shown as small transparent points in the background. Consistent with the dominant helical conformations displayed in Fig. 3, this figure quantitates the deviations of the RNA12 duplexes from the canonical A-form geometry. A wider range of conformations is found in KCl containing solutions, with helical radii as large as 10.04 Å and helical lengths of 27 Å, compared to 9.2 and 31 Å respectively for canonical A-form helix. The helical conformations in KCl do not follow a specific trend, but rather indicate the unconstrained structural dynamics of the short RNA duplex. On the other hand, in $MgCl_2$, RNA12 duplexes are more compact than in KCl. The larger helical radii show that the two strands are not as tightly coupled as in the A-form. Moreover, RNA12 is shorter than an equivalent A-form duplex in $MgCl_2$ containing solutions. To form such short and tight conformations, the duplex has to unwind and assume smaller total twists to avoid steric clashes. Different effects of monovalent and multivalent ions (on a background of monovalent ions) were observed[10] and here, we provide a quantitative interpretation and different picture. Monovalent ions alone cause structural deviations from assumed A-form conformation, whereas added divalent ions further distort the RNA duplex. The schematic cylindrical models of Fig. 4(e) reflect these suggested helical deformations and variations.

Robustness of the Refined Conformations

Challenging, but vitally important questions must be addressed when using optimization algorithms to match experimental data with theoretical models. How confident are we that the results do not suffer from overfitting and remain invariant despite different levels of experimental error? For these data, the number of Shannon channels is between 12–16. With as few as 5 curves in the ensemble and only one free fitting parameter, overfitting is not a concern. It is also important to report the effect of different ensemble sizes used[57]. Increasing the ensemble size from 1 to 4 improves the fit; no further improvement is detected when the size increases above 5. Extended-EOM with larger ensemble sizes (for example, 8, 10, 15, 20 or 50) simply selects the same conformation repeatedly without an improvement in ensemble statistics. Therefore, the ensemble size was set to the smallest integer value to reduce χ^2 defined in Eq. (2) and to avoid overfitting.

Furthermore, we implemented the 5-fold cross validation described in Methods and Materials. Figs. 5(a,b) show comparisons of 5-fold results to the results from full data set while Fig. 5(c) shows the averaged error traces $\bar{E}$ in the validation set from 100 repetitions. Robustness of the eEOM method can be inferred from the stable and favorable value of χ^2 with small standard deviations across all fits. In addition, the quality of the fits implies that the hyperparameters used in eEOM (the number of generations, cycles and curves in an ensemble) are sufficient to converge to a globally optimal ensemble. As an additional confidence check, we synthesized experimental curves with 4 different levels of experimental error. Selected structures are plotted as histograms in Fig. 5(d) with the inset showing the fitted curves on top of synthesized data (offset for clarity). The eEOM refinement remains robust, with consistent conformations across the 4 tests. However, although the dominant conformations are recapitulated, the effect of larger experimental error is to include more mixed conformations in the ensemble which, in turn, results in larger error bars (larger confidence intervals) in the conformational space and geometric parameters. Note that the S/N from our SWAXS experiments is between 3–5, near the lower end of the test range. These simulations demonstrate the statistical robustness of the results.

Comparison With A-RNA25 and B-DNA25

To rule out the possible dominance of end effects in conformational variation for the 12 base pair duplexes, we also measured scattering profiles from 25 base pair RNA and DNA duplexes and compared scattering profiles computed from a canonical A-form RNA25 and a B-form DNA25 with our measured SWAXS curves. These comparisons are shown in Fig. 6. The theoretical SWAXS profiles were computed as described above and are plotted on top of the experimentally acquired profiles with no adjustment (no free fitting parameters). For DNA, the single B-form model effectively captures the features in the experimental curve (subject to our ion and solvent population protocols). In contrast, the profile computed from the pure A-form model of RNA25 disagrees with the experimental data, consistent with the above results on RNA12. Even for these longer helices, the deviations in scattering profiles suggest that RNA duplex structures are distinct from one another, as salt conditions are changed, and from the assumed A-form structure.

DISCUSSION

Agreement with Canonical B-form DNA

Comparisons between experimental and simulated SWAXS curves of canonical B-form DNA were previously reported using a similar 3D-RISM formalism[36]. The major differences with this work are as follows. First, we applied *placevent*[58] to populate the calculated 3D-RISM densities with explicit solvent and ion molecules. Second, we used orientational averaging across 1,500 uniformly distributed vectors in 3D instead of the CUBE method[59]. This method is very precise compared to the classical Debye formula with more than 1,000 orientations at $q_{max} = 1.4 \text{ Å}^{-1}$. Third, we used an explicit solvent background to simulate buffer subtraction. Finally, we do not expect good agreement of full SWAXS profile over the entire q range since intermolecular interactions were observed in all samples as a result of the high nucleic acid concentration needed to obtain a measurable signal in the WAXS regime. This effect alters the scattering at low q values, e.g. below $q = $

0.1 Å^{-1} [24]. We therefore omitted this range in our fits (RNA12) and visualization (RNA25 and DNA25). More detailed information can be found in the Supporting Material.

It is known that the addition of a constant, c, to either the simulated or experimental WAXS profiles improves their agreement. This additive constant compensates for imperfections in buffer subtraction, detector dark currents and possible subtle effects beyond physical formulation. The value indicates the degree of deviation between computed and experimental profiles (given the same conformation, e.g. B-form DNA25). We only applied this additive constant in the case of eEOM refinement of RNA12 conformations, and its value does not exceed 10% of the SWAXS intensity at $q_{max} \cong 0.79$ Å^{-1}. Moreover, even without this constant, the computed WAXS profiles of DNA25 agree with experimental measurement, underscoring the robustness of our computation.

The A-Form Ratio of RNA12

Each distinct helical geometry can be effectively described by the conformations of dinucleotides steps within the duplex. These geometries are determined and classified into C2' or C3'-endo sugar puckers in *x3dna-dssr*. For the canonical A-form, each dinucleotide step assumes the C3'-endo conformation. We analyzed the 11 dinucleotide steps present in the RNA12 conformations in the SWAXS-derived ensemble (the selected structures) and computed the A-form ratios, r_A, defined to be the fraction of nucleotides in A-form. This ratio equals the number of observed A-form dinucleotide steps, N_A, divided by 11.

$$r_A = \frac{N_A}{11}$$

(5)

The salt dependence of ratios is shown in Fig. 7 using the color scheme of Figure 3. An interesting bimodal distribution appears for the RNA12 system in KCl containing solutions. The highest deviation from A form, or the most heterogenous structures, is found near 100 – 200 mM KCl, close to physiological ionic strength. In contrast more A-like structures are found at lower or higher [KCl]. A tighter distribution is found with added MgCl$_2$, though only around 60% of the dinucleotide steps assume the A-form geometry.

A Survey of DsRNA Duplexes

We now place this work in the context of other literature reports of deviations from A-form. With the addition of multivalent ions, compression of the duplex was previously observed in simulation, but was accompanied by an overtwisting of the duplex induced by trivalent Cobalt hexamine ions (Table 1)[10]. Those simulations suggest that a more compact helical conformation arises from decreasing helical rise and helical over-winding. A similar result was reported by independent MD simulations investigating RNA twist-stretch coupling[15]. However, our results suggest that divalent ions give rise to the compaction of RNA12 duplex by slightly *unwinding* the helix and compressing it along the helical axis. This claim is supported by the larger measured helical radius of the RNA12: if the duplex over-winds with respect to the axis to a thinner conformation, the radius should decrease to accommodate a larger inclination of the base pairs (which is the case for canonical A-form RNA). Conveniently, unwinding also prevents some steric clashes given the larger volume within the backbones. This ion-induced compression implies that divalent or multi-valent ions enter

the RNA major grooves, pulling together the "ceiling" and "floor" of the major-groove chamber, while serving as counter-ions to charge neutralize the RNA duplex. Monovalent ions such as K^+ might also access the major groove but may have a smaller impact on compression due to their lower charge density.

There are a few simulations of dsRNA duplexes that report helical geometries. They were mostly conducted using MD simulations and corroborated by comparison to experimental data. Table 1 provides a, to the best of our knowledge, complete list of these works. The first entry in the table arises from this study, reporting diverse helical parameters of our SWAXS refined conformations in solutions containing different monovalent and divalent salt concentrations. This parameter, discussed in more depth in the next section, is readily extractable from the SWAXS profiles.

In addition to the simulation results of Table 1, long or short dsRNA duplex structures were determined using experimental methods such as X-ray crystal diffraction and solution NMR. Some of the solved structures are available through the protein data bank (PDB). We analyzed the PDB structures of 23 isolated dsRNA duplexes of different lengths using *x3dna-dssr* and compare the average helical radius and twist with our results. Figure 8 shows these comparisons, and more details are provided in Table S1. Although many of these duplexes have structures that resemble the canonical A-form, most of the solution and some of the crystal structures have different conformations. On average, these duplexes have larger helical radii and lower average helical twist, suggesting that they are unwound compared to A-form, which has the tightest double strands. It is noteworthy that the parameters of the RNA12 duplexes in $MgCl_2$ resemble crystallized RNA duplexes while RNA12 duplexes in KCl more closely agree with other solution conformations. Note that the crystal/solution environments of these dsRNA duplexes vary and can include heavy atoms, large anion groups or even bound proteins.

Correlation of Radius and Helical Rise with WAXS Features

Solution X-ray scattering maps 3D macromolecular solution structures to 1D scattering profiles by Fourier transform of the ensemble-averaged electron density. Any definition or disruption of a structural periodicity within the theoretical resolution is reflected by features in the scattering profile (assuming that the scattering is dominated by the macromolecule and not the surrounding solvent). Real-space periodicities of the molecule often manifest as local minima and maxima at large q values. Therefore, it is useful to correlate the properties of those features observed in the scattering profiles, such as positions of peaks and troughs and their amplitudes (so called 'contrast') with the determined structural parameters of the RNA12 duplex conformations. We use the entire pool of MD generated structures to establish these correlations. We identify the positions and define the contrast, *CR*, of the extrema as the following

$$CR = \frac{I_{ext}}{I_0} \tag{5}$$

where I_{ext} and I_0 are the scattering intensities at the extremum and $q = 0$. Figs. 9(a, b) demonstrate an empirical correlation between the first maximum position/contrast and

helical radius and between the first minimum position and average helical rise. As expected, the first maximum position shifts toward higher q as the helical radius (the strongest periodicity) decreases. The correlation length within this q regime is between 14 and 20 Å, consistent with the dimensions of the duplex's helical diameter(s). The contrast of the first maximum also decreases linearly with helical radius because a shorter correlation length inevitably reflects finer details in periodicity or disruption of periodicity, such as groove spacing and base arrangements. Additionally, the averaged helical rise correlates positively with the first minimum position. Surprisingly, the larger the helical rise, the higher in q is the position of the first minimum. Interpretations of this type are useful in facilitating model-free interpretations of SWAXS profiles. More data, shown in Figs. S4–S7, suggest that as the salt concentration increases, such correlations between extrema and structural features become relatively insensitive as the curves flatten out at higher KCl concentrations. This effect might be the result of increased numbers of unstructured ions (both anions and cations) in the bulk solution, with only a small portion of cation distribution shaped by the molecule. Effects of this type illustrate the importance of including ion clouds in modeling in this q regime.

CONCLUSION

In summary, we demonstrate the ability to model and extract useful quantitative information from the small- and wide-angle X-ray scattering experiments on RNA duplexes, with an emphasis on the less explored, but information rich wider-angle data. A close examination of the profiles computed from structures, populated with solvent and ions, allows us to link features of the SWAXS curves with real space properties of the duplex. This analysis supports the claim that SWAXS experiments can probe finer details of solution macromolecular conformation with high periodicity. Any definition or disruption of the structural periodicity, even below the theoretical resolution of correlation length,

$d_{min} = \frac{2\pi}{q_{max}}$, can be observed empirically, such as helical rise.

We also developed and applied an extended ensemble optimization method to refine MD generated pools of structures for compatibility with experimental data on short RNA duplexes, exploiting a customized goodness-of-fit, χ^2. The statistical robustness of our model was tested extensively by 5-fold cross validation and on synthesized datasets with different signal-to-noise ratios. Our analysis framework can be easily applied to interpret wider-angle data of other structured monodisperse macromolecular systems. The structures in these refined ensembles suggest that RNA duplexes adopt diverse versus definite helical conformations in solutions containing KCl and MgCl$_2$ respectively. Duplexes in KCl containing solutions have larger diameters than those in solutions containing MgCl$_2$, and both diameters exceed that of duplexes in the canonical A-form. The A-form helical structure of RNA duplexes is found to be the exception, rather than the rule. The analysis suggests the mechanism for shrinkage of the RNA12 duplex and its subsequent tight helical conformation in divalent ions. The duplex unwinds to assume a larger helical diameter that can tolerate additional compression while avoiding steric clashes.

Our results underscore the dynamic nature of the dsRNA duplex. These structures must be considered when accounting for any inter or intramolecular interactions of RNA. For example, RNA folding studies highlight the importance of properly oriented molecules that enable tertiary contact formation. Similar and parallel results demonstrating the sequence-dependent helical conformation of RNA duplexes underscore the fine structural dynamics of these important elements[65]. Beyond folding, RNA duplexes serve various biological functions that rely on precisely sculpted interactions with their surfaces, including the recognition of dsRNA binding domains (dsRBD)[66] such as RNA helicase[67] and multidomain ribonuclease DICER[68,69], CRISPR RNA processing[70], determination of conformations of the post-processed micro RNAs (miRNA)[71,72] and many more. All of these processes rely on understanding the structure(s) of the RNA duplex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Dock-Bregeon AC, Chevrier B, Podjarny A, Moras D, deBear JS, Gough GR, Gilham PT and Johnson JE High resolution structure of the RNA duplex [U(U-A)6A]2. Nature. 335(6188): 375–8 (1988). [PubMed: 2458530]

2. Holbrook SR, Cheong C, Tinoco I and Kim SH Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature. 353(6344): 579–81 (1991). [PubMed: 1922368]

3. Tolbert BS, Miyazaki Y, Barton S, Kinde B, Starck P, Singh R, Bax A, Case DA and Summers MF Major groove width variations in RNA structures determined by NMR and impact of13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. J. Biomol. NMR 47(3):205–19 (2010). [PubMed: 20549304]

4. Chu VB, Lipfert J, Bai Y, Pande VS, Doniach S and Herschlag D Do conformational biases of simple helical junctions influence RNA folding stability and specificity? RNA 15(12): 2195–205 (2009). [PubMed: 19850914]

5. Daher M, Mustoe AM, Morriss-Andrews A, Brooks CL and Walter NG Tuning RNA folding and function through rational design of junction topology. Nucleic Acids Res. 45(16): 9706–9715 (2017). [PubMed: 28934478]

6. Shi X, Walker P, Harbury PB and Herschlag D Determination of the conformational ensemble of the TAR RNA by X-ray scattering interferometry. Nucleic Acids Res. 45(8): e64 (2017). [PubMed: 28108663]

7. Denny SK, Bisaria N, Yesselman JD, Das R, Herschlag D and Greenleaf WJ High-throughput investigation of diverse junction elements in RNA tertiary folding. Cell 174(2):377–390.e20 (2018). [PubMed: 29961580]

8. Sutton JL and Pollack L Tuning RNA flexibility with helix length and junction sequence. Biophys. J 109(12): 2644–2653 (2015). [PubMed: 26682821]

9. Tolokh IS, Pabit SA, Katz AM, Chen Y, Drozdetski A, Baker N, Pollack L and Onufriev AV Why double-stranded RNA resists condensation. Nucleic Acids Res. 42(16): 10823–10831 (2014). [PubMed: 25123663]

10. Pabit SA, Katz AM, Tolokh IS, Drozdetski A, Baker N, Onufriev AV and Pollack L Understanding nucleic acid structural changes by comparing wide-angle x-ray scattering (WAXS) experiments to molecular dynamics simulations. J. Chem. Phys 144 (20): 205102, (2016). [PubMed: 27250330]

11. Chen Y-L, Sutton JL and Pollack L How the conformations of an internal junction contribute to fold an RNA domain. J. Phys. Chem. B 122(49): 11363–11372 (2018). [PubMed: 30285445]

12. Hub JS Interpreting solution X-ray scattering data using molecular simulations. Current Opinion in Structural Biology 49: 18–26 (2018). [PubMed: 29172147]

13. Auffinger P and Westhof E Melting of the solvent structure around a RNA duplex: A molecular dynamics simulation study. Biophys. Chem 95(3): 203–210 (2002). [PubMed: 12062380]

14. Bao L, Zhang X, Shi YZ, Wu YY and Tan ZJ Understanding the relative flexibility of RNA and DNA duplexes: stretching and twist-stretch coupling. Biophys. J 112(6): 1094–1104 (2017). [PubMed: 28355538]

15. Liebl K, Drsata T, Lankas F, Lipfert J and Zacharias M Explaining the striking difference in twist-stretch coupling between DNA and RNA: a comparative molecular dynamics analysis. Nucleic Acids Res. 43(21): 10143–10156 (2015). [PubMed: 26464435]

16. Jin L, Shi YZ, Feng CJ, Tan YL and Tan ZJ Modeling structure, stability, and flexibility of double-stranded RNAs in salt solutions. Biophys. J 115(8): 1403–1416 (2018). [PubMed: 30236782]

17. Bernado P, Mylonas E, Petoukhov MV, Blackledge M and Svergun DI Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc 129(17): 5656–5664 (2007). [PubMed: 17411046]

18. Tria G, Mertens HDT, Kachala M and Svergun DI Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ 2(Pt 2): 207–217 (2015).

19. Róycki B, Kim YC and Hummer G SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure 19(1): 109–116 (2011). [PubMed: 21220121]

20. Plumridge A, Katz AM, Calvey GD, Elber R, Kirmizialtin S and Pollack L Revealing the distinct folding phases of an RNA three-helix junction. Nucleic Acids Res. 46(14): 7354–7365 (2018). [PubMed: 29762712]

21. Shevchuk R and Hub JS Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. PLoS Comput. Biol 13(10): 1–27 (2017).

22. Grishaev AV Biological small angle scattering: techniques, strategies and tips. Singapore, Springer, 2017.

23. Panjkovich A; Svergun DI SASpy: A PyMOL plugin for manipulation and refinement of hybrid models against small angle Xray scattering data. Bioinformatics 2016, 32 (13), 2062–2064. [PubMed: 27153695]

24. Li L, Pabit SA, Meisburger SP and Pollack L Double-stranded RNA resists condensation. Phys. Rev. Lett 106(10): 1–4 (2011).

25. Chen Y-L, Lee T, Elber R and Pollack L Conformations of an RNA helix-junction-helix construct revealed by SAXS refinement of MD simulations. Biophys. J 116(1): 19–30 (2019). [PubMed: 30558889]

26. Hopkins JB, Gillilan RE and Skou S BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J. Appl. Crystallogr 50(Pt 5): 1545–1553 (2017). [PubMed: 29021737]

27. Franke D, Jeffries CM and Svergun DI Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. Nat. Methods 12(5): 419–422 (2015). [PubMed: 25849637]

28. Plumridge A, Meisburger SP and Pollack L Visualizing single-stranded nucleic acids in solution. Nucleic Acids Res. 45(9): e66 (2016).

29. Plumridge A, Meisburger SP, Andresen K and Pollack L The impact of base stacking on the conformations and electrostatics of single-stranded DNA. Nucleic Acids Res. 45(7): 3932–3943 (2017). [PubMed: 28334825]

30. Chen Y, Tokuda JM, Topping T, Meisburger SP, Pabit SA, Gloss LM and Pollack L Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. Proc. Natl. Acad. Sci114(2): 334–339 (2017). [PubMed: 28028239]

31. Chen Y, Tokuda JM, Topping T, Sutton JL, Meisburger SP, Pabit SA, Gloss LM and Pollack L Revealing transient structures of nucleosomes as DNA unwinds. Nucleic Acids Res. 42(13): 8767–76 (2014). [PubMed: 24990379]

32. Mauney AW, Tokuda JM, Gloss LM, Gonzalez O and Pollack L Local DNA sequence controls asymmetry of DNA unwrapping from nucleosome core particles. Biophys. J 115(5): 773–781 (2018). [PubMed: 30072033]

33. Svergun D, Barberato C and Koch MH CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr 28: 768–773 (1995).

34. Park S, Bardhan JP, Roux B and Makowski L Simulated x-ray scattering of protein solutions using explicit-solvent models. J. Chem. Phys 130(13): 134114 (2009). [PubMed: 19355724]

35. Grishaev A, Guo L, Irving T and Bax A Improved fitting of solution x-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. J. Am. Chem. Soc 132(44): 15484–15486 (2010). [PubMed: 20958032]

36. Nguyen HT, Pabit SA, Meisburger SP, Pollack L and Case DA Accurate small and wide angle x-ray scattering profiles from atomic models of proteins and nucleic acids J. Chem. Phys 141(22): 1–15 (2014).

37. Chen PC and Hub JS Interpretation of solution x-ray scattering by explicit-solvent molecular dynamics. Biophys. J 108(10): 2573–2584 (2015). [PubMed: 25992735]

38. Knight CJ and Hub JS WAXSiS: A web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. Nucleic Acids Res. 43(W1): W225–W230 (2015). [PubMed: 25855813]

39. Ivanovi MT, Bruetzel LK, Shevchuk R, Lipfert J and Hub JS Quantifying the influence of the ion cloud on SAXS profiles of charged proteins. Phys. Chem. Chem. Phys 20(41): 26351–26361 (2018). [PubMed: 30303199]

40. Yang S, Salmon L and Al-Hashimi HM Measuring similarity between dynamic ensembles of biomolecules. Nat. Methods 11(5): 552–554 (2014). [PubMed: 24705474]

41. Moore PB Small-angle scattering. Information content and error analysis. J. Appl. Cryst 13, 168–175 (1980).

42. Rambo RP and Tainer JA Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496(7446): 477–481 (2013). [PubMed: 23619693]

43. Spill YG and Nilges M SAS profile correlations reveal SAS hierarchical nature and information content. PLoS One 12(5): e0177309 (2017). [PubMed: 28493950]

44. Chen PC and Hub JS Validating solution ensembles from molecular dynamics simulation by wide-angle x-ray scattering data. Biophys. J 107(2): 435–447 (2014). [PubMed: 25028885]

45. Brünger AT Free R-Value - A novel statistical quantity for assessing the accuracy of crystal-structures. Nature 355(6359): 472–475 (1992). [PubMed: 18481394]

46. Falkner B and Schroder GF Cross-validation in cryo-EM-based structural modeling. Proc. Natl. Acad. Sci 110(22): 8930–8935 (2013). [PubMed: 23674685]

47. Bezanson J, Karpinski S, Shah VB and Edelman A Julia: a fresh approach to numerical computing. SIAM Rev. 59(1): 65–98 (2017).

48. Petoukhov MV and Svergun DI Applications of small-angle x-ray scattering to biomacromolecular solutions. International Journal of Biochemistry and Cell Biology 45(2): 429–437 (2013). [PubMed: 23142499]

49. Kikhney AG; Svergun DI A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Letters 2015, 589 (19A), 2570–2577. [PubMed: 26320411]

50. Li J, Jiao A, Chen S, Wu Z, Xu E and Jin Z Application of the small-angle x-ray scattering technique for structural analysis studies: a review. Journal of Molecular Structure 1165(5): 391–400 (2018).

51. Makowski L, Rodi DJ, Mandava S Minh D, Gore DB and Fischetti RF Molecular crowding inhibits intramolecular breathing motions in proteins. J. Mol. Biol 375(2): 529–546 (2008). [PubMed: 18031757]

52. Malmerberg E, Omran Z, Hub JS, Li X, Katona G, Westernhoff S, Johansson LC, Andersson M, Cammarata M, Wulff M, et al., Time-resolved WAXS reveals accelerated conformational changes in iodoretinal-substituted proteorhodopsin. Biophys. J 101(6): 1345–1353 (2011). [PubMed: 21943415]

53. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J and Kovalenko A Three-dimensional molecular theory of solvation coupled with molecular dynamics in AMBER. J. Chem. Theory Comput 6(3): 607–624 (2010). [PubMed: 20440377]

54. Macke TJ and Case DA Modeling unusual nucleic acid structures. Mol. Model. Nucleic Acids 24: 379–393 (1997).

55. Lu XJ and Olson WK 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat. Protoc 3(7): 1213–1227 (2008). [PubMed: 18600227]

56. Zuo X, Cui G, Merz KM Jr, Zhang L, Lewis FD and Tiede DM X-ray diffraction 'fingerprinting' of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. Proc. Natl. Acad. Sci 103(10): 3534–3539 (2006). [PubMed: 16505363]

57. Frank AT, Stelzer AC, Al-Hashimi HM and Andricioaei I Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. Nucleic Acids Res. 37(11): 3670–9 (2009). [PubMed: 19369218]

58. Sindhikara DJ, Yoshida N and Hirata F Placevent: An algorithm for prediction of explicit solvent atom distribution-application to HIV-1 protease and F-ATP synthase. J. Comput. Chem 33(18): 1536–1543 (2012). [PubMed: 22522665]

59. Virtanen JJ, Makowski L, Sosnick TR and Freed KF Modeling the hydration layer around proteins: applications to small- and wide-angle X-ray scattering. Biophys. J 101(8): 2061–2069 (2011). [PubMed: 22004761]

60. Blanchet C, Pasi M, Zakrzewska K & Lavery R CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. Nucleic Acids Res. 39(Web Server issue): W68–73 (2011). [PubMed: 21558323]

61. Liu J-H, Xi K, Zhang X, Bao L, Zhang X and Tan Z-J Structural flexibility of DNA-RNA hybrid duplex: stretching and twist-stretch coupling. Biophys. J 117(1): 74–86 (2019). [PubMed: 31164196]

62. Marin-Gonzalez A, Vilhena JG, Perez R and Moreno-Herrero F Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics. Proc. Natl. Acad. Sci 114(27): 7049–7054 (2017). [PubMed: 28634300]

63. Lavery R, Zakrzewska K and Sklenar H JUMNA (junction minimisation of nucleic acids). Comput. Phys. Commun 91(1–3): 135–158 (1995).

64. Liu C, Janowski PA and Case DA All-atom crystal simulations of DNA and RNA duplexes. Biochim. Biophys. Acta - Gen. Subj 1850(5): 1059–1071 (2015).

65. Yesselman JD, Denny SK, Bisaria N, Herschlag D, Greenleaf WJ and Das R Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. Proc. Natl. Acad. Sci 116(34): 201901530 (2019).

66. Hentze MW, Castello A, Schwarzl T and Preiss T A brave new world of RNA-binding proteins. Nature Reviews Molecular Cell Biology 19(5): 327–341 (2018). [PubMed: 29339797]

67. Fu Q and Yuan YA Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). Nucleic Acids Res. 41(5): 3457–3470 (2013). [PubMed: 23361462]

68. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD and Doudna JA Structural basis for double-stranded RNA processing by Dicer. Science 311(5758): 195–8 (2006). [PubMed: 16410517]

69. Han J, Lee Y, Yeom KH, Kim YK, Jin H and Kim VN The Drosha – DGCR8 complex in primary microRNA processing. Genes Dev. 18(24): 3016–27 (2004). [PubMed: 15574589]

70. Li H Structural principles of CRISPR RNA processing. Structure 23(1): 13–20 (2015). [PubMed: 25435327]

71. Roth BM, Ishimaru D and Hennig M The core microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a nonspecific RNA-binding protein. J. Biol. Chem 288(37): 26785–26799 (2013). [PubMed: 23893406]

72. Quarles KA, Chadalavada D and Showalter SA Deformability in the cleavage site of primary microRNA is not sensed by the double-stranded RNA binding domains in the microprocessor component DGCR8. Proteins. 83(6): 1165–79 (2015). [PubMed: 25851436]
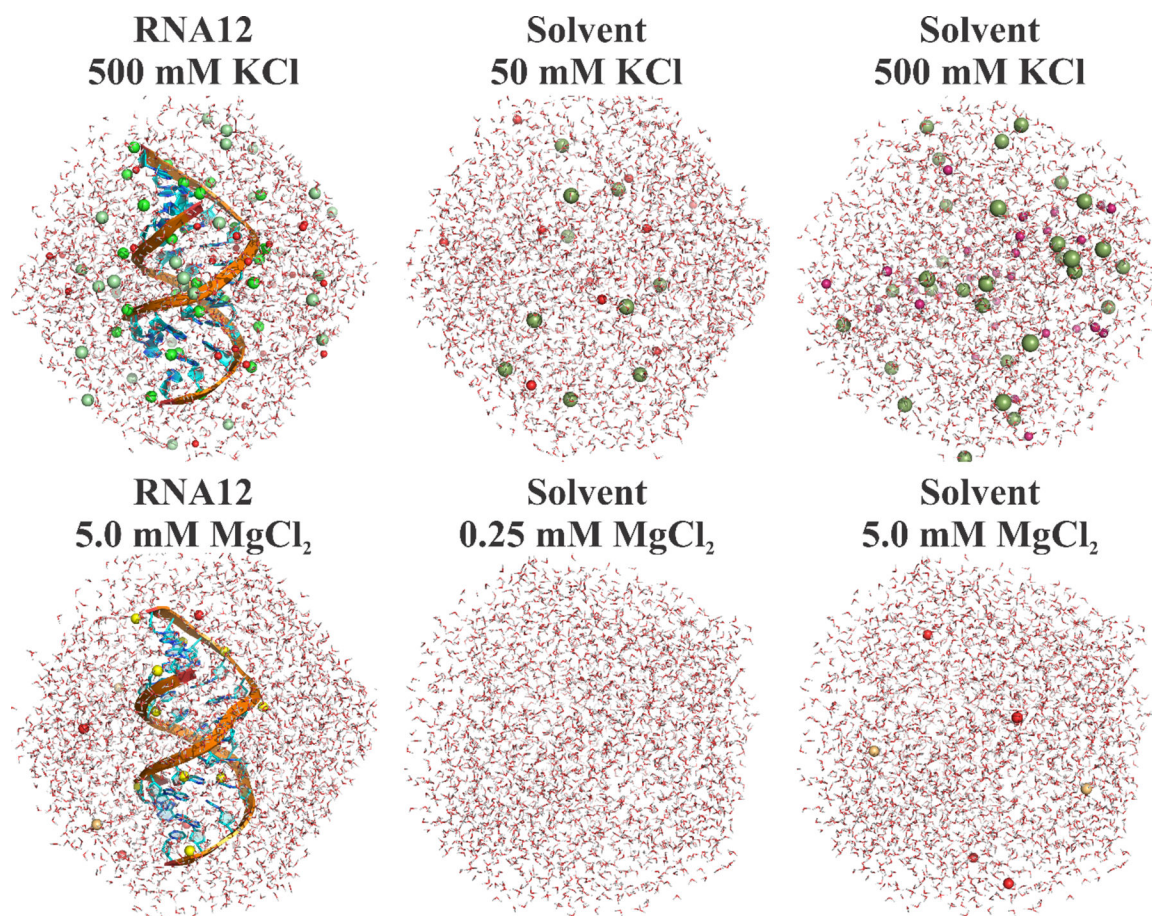
Figure 1.
Explicit solvent and ion models. The left panel shows one RNA12 conformation in 500 mM KCl and 5.0 mM $MgCl_2$. Solvent and ion models in the absence of solute are shown in the middle and right columns, at the quoted salt concentrations. The color schemes are defined as follows. Green: K^+ from 3D-RISM. Dark green: bulk K^+. Yellow: Mg^{2+} ions from 3D-RISM. Dark yellow: bulk Mg^{2+} ions. Red: bulk Cl^- ions. The number of ions directly corresponds to the experimental bulk concentrations. Notice that for 0.25 mM $MgCl_2$, the bulk solvent contains no ions due to the low concentration.

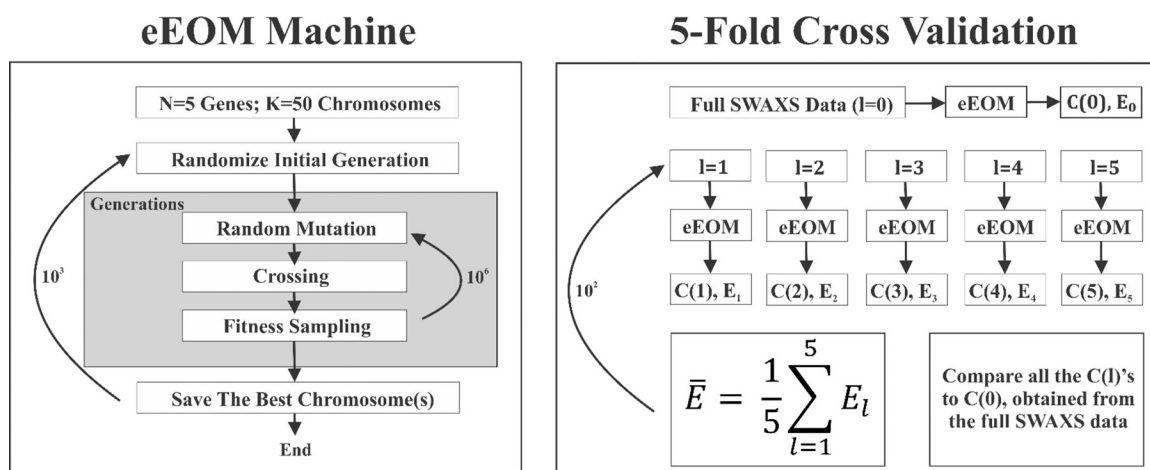## eEOM Machine

5-Fold Cross Validation

Figure 2.

The flowchart for the extended ensemble optimization method (eEOM) and the 5-fold cross validation (CV) strategy, including hyperparameters. The l's are the lth fold of the truncated SWAXS curve and $C(l)$ and E_l are the chromosome and error from the lth CV. See main text for a detailed explanation.
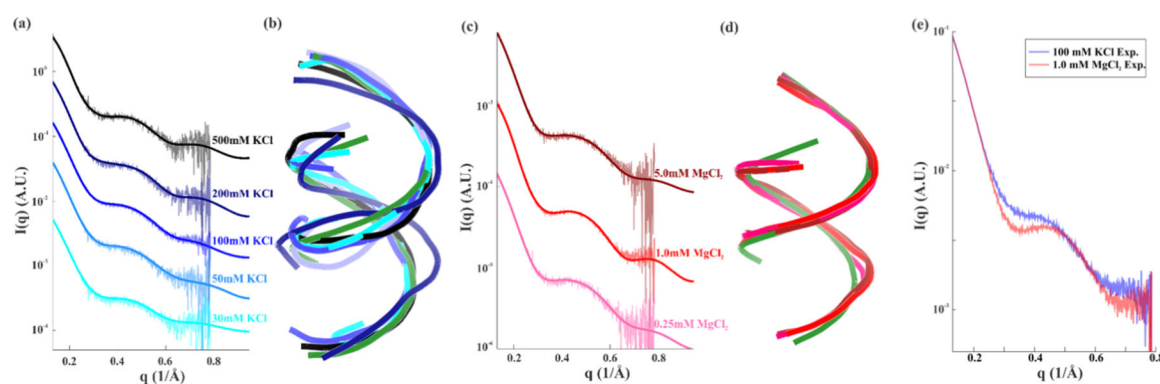
Figure 3.
The fitted results from eEOM and the corresponding RNA12 conformations shown as phosphate backbones. (a) The experimental SWAXS curves and fits for duplexes in solutions containing different concentrations of KCl. The χ^2 value for each condition (from 30 (cyan) to 500 (black) mM) is 1.169, 0.932, 0.946, 1.152 and 1.037. (b) The representative RNA12 conformations in the KCl refinement. (c) The experimental SWAXS curves and fits for duplexes in solutions containing different concentrations of $MgCl_2$. The χ^2 value for each condition (from 0.25 (pink) to 5.0 (dark red) mM) is 1.434, 1.124 and 1.400. (d) The representative RNA12 conformations in the $MgCl_2$ refinement. Note that the SWAXS curves are offset for visualization. The green duplex represents the canonical A-form RNA12 helix. The backbone structures are rendered by PyMOL (Schrodinger Inc., New York, NY). (e) This panel compares experimental SWAXS curves in solutions containing 100 mM KCl and 1 mM $MgCl_2$. The differences measured between the salts is larger than the deviations within either salt series, over the range tested. All data have been deposited in the SASBDB with accession codes: SAS SASDHG2, SASDHH2, SASDHJ2, SASDHK2, SASDHGL2, SASDHM2, SASDHN2, SASDHP2, SASDHQ2, SASDHS2, SASDHR2, SASDHT2, SASDHU2, SASDHV2.
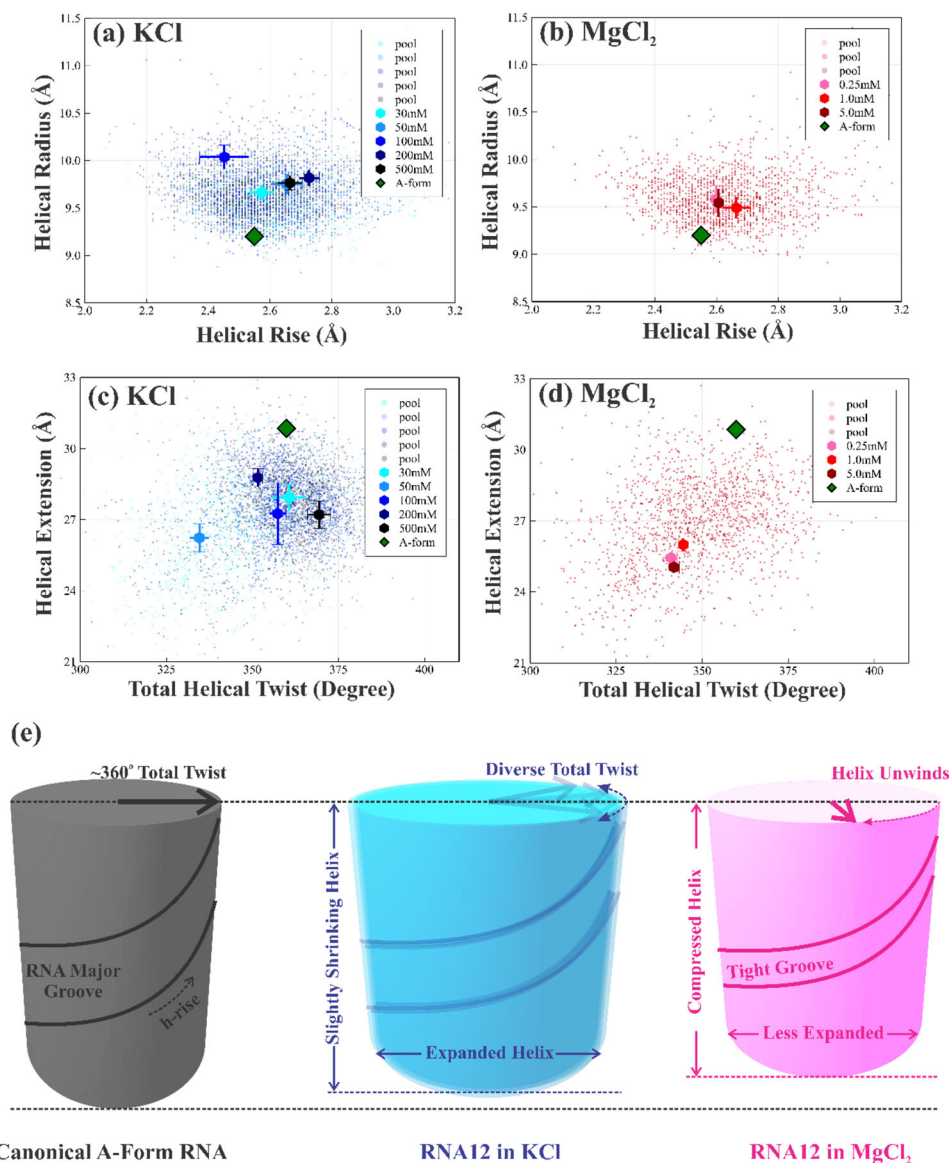
Figure 4.
The mapping of a multidimensional conformational spaces into two, 2-dimensional parameter spaces using helical rise and radius for KCl (a) and $MgCl_2$ (b) and total helical extension and twist for KC l(c) and $MgCl_2$ (d) with the same color schemes as in Fig. 3. Parameters for all structures in the MD pools are shown as small transparent points in the background. (e) The schematic summary of the RNA12 duplex in canonical A-form, KCl and $MgCl_2$ as suggested by the helical parameters in (a)-(d). Note that the cartoon illustration is not drawn to scale. Solution conformations of RNA duplexes differ from the A-form structure. Results of our refinements suggest that higher-valent cations unwind and further compress the duplex.
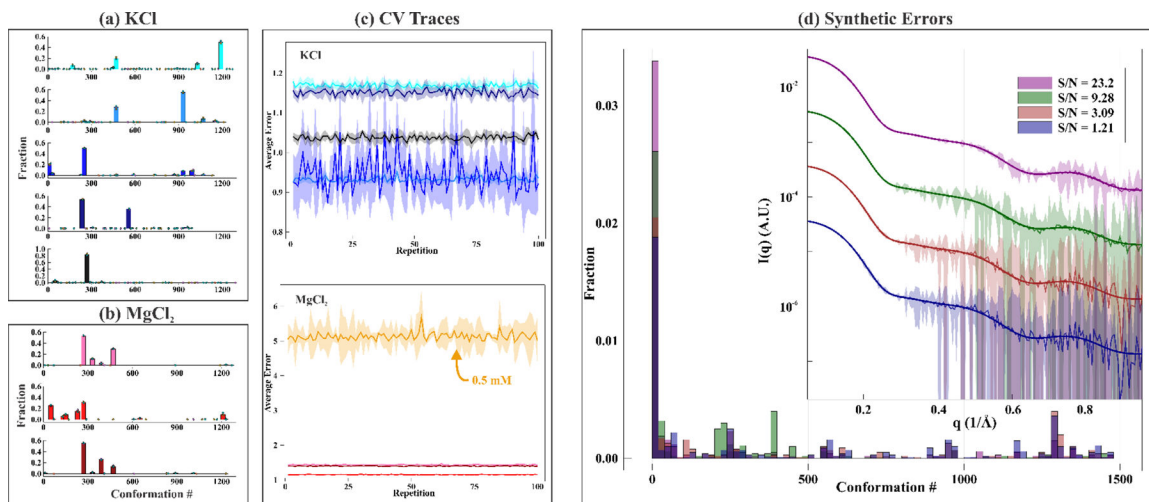
Figure 5.
The results from 5-fold cross validation and from varying tests of experimental errors. (a) The refined conformations of the full SWAXS data (bars) and every fold of validation test (dots) from solutions containing 30 (top) to 500 (bottom) mM KCl. Figs. S10–S14 provide additional information about the robustness of the 5-fold cross validation analysis. (b) The same plot as (a) for solutions containing $MgCl_2$ at concentrations of 0.25 (top) to 5.0 (bottom) mM. (c) The traces of $\bar{E}$ and one standard deviation from 100 CV runs for all the salt concentrations explored, using previously defined color scheme. Note that the range of the y-axis differs for KCl and $MgCl_2$ to show the non-convergent behavior and the poor fit obtained for the 0.5 mM $MgCl_2$ condition, whose interpretation is dropped in the main text. With the exception of this point, the others have a good average error on the validation set, and display the consistent refined conformations shown in (a) and (b), which bolsters the robustness of our results. (d) The refined RNA12 conformations extracted from data synthesized with added experimental noise to achieve a S/N ratio of 23.2, 9.28, 3.09 and 1.21. The inset displays the fit of our eEOM procedure to this fabricated, noisy data set. The representative RNA conformations are well-recapitulated despite the different S/N levels, even when the signal is comparable in size to the noise. As expected, larger noise introduces a mix of conformations into the refined ensemble, rendering a larger confidence interval in the conformational spaces. The S/N level of our SWAXS experiments is about 3–5.
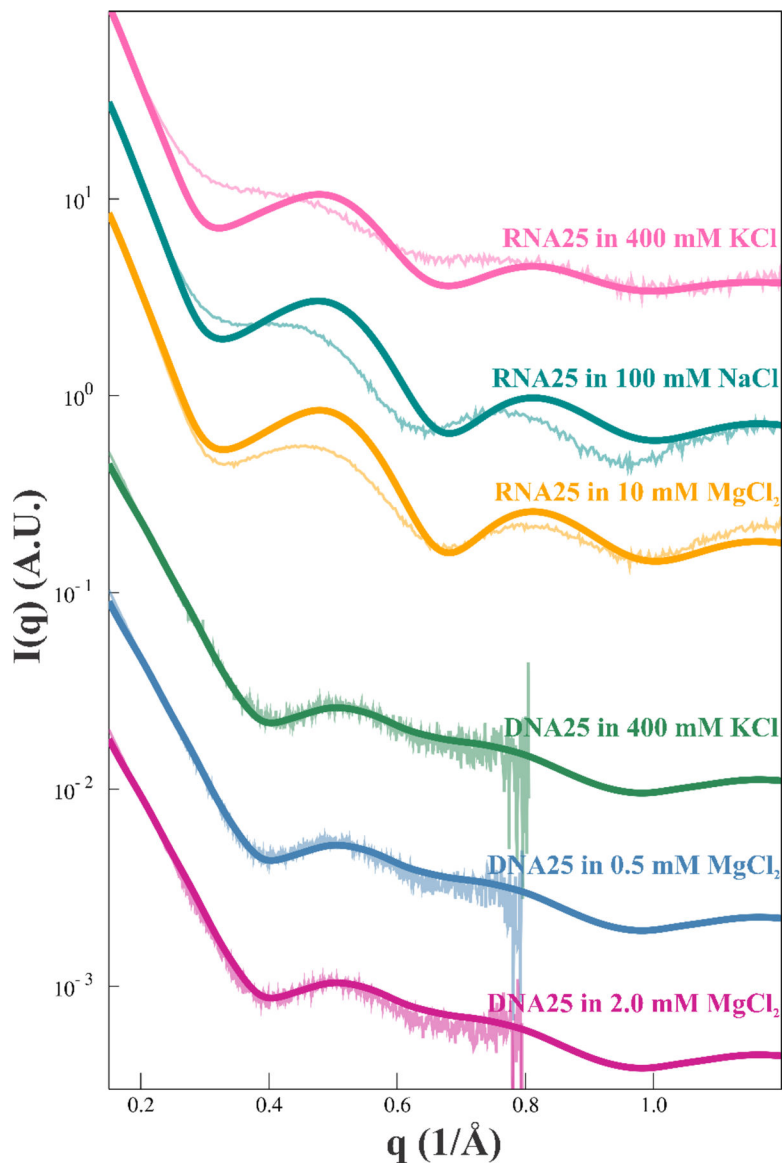
Figure 6.
Direct comparison of experimental SWAXS measurements of 25-bp RNA and DNA duplexes to calculated SWAXS curves from a canonical A-form RNA25 and B-form DNA25 following the same procedures as the RNA12. These comparisons suggest that the dsDNA assumes a canonical B-form duplex under a broad range of solution conditions. In contrast, and consistent with our findings on the shorter duplex, the 25 bp RNA duplex has significant structural variations in different solutions and does not adopt a canonical A-form. The obvious deviations suggest that RNA duplex is highly dynamic, with ion-dependent conformations.
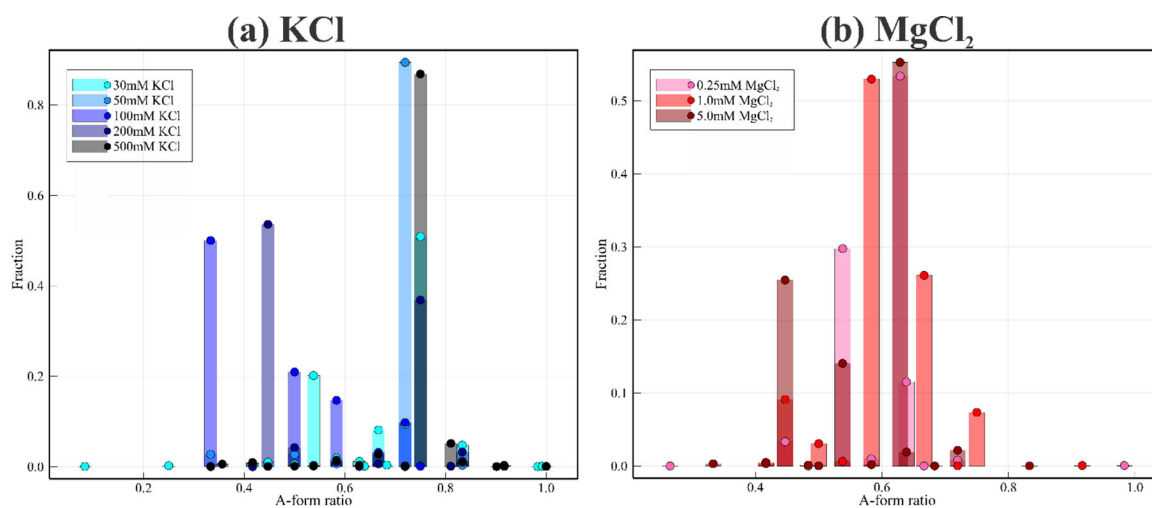
(a) KCl

(b) MgCl$_2$

Figure 7.
The A-form ratios of the refined ensembles using 11 dinucleotide steps for KCl (a) and
MgCl$_2$ (b) containing solutions at the quoted concentrations. The perfect A-form duplex has
an A-form ratio of 1.0. The normalized fractions indicate the abundance of A-form structure
in the derived ensembles. Note the broader distribution of structures in KCl, and the overall
(average) lower A -ratio in MgCl$_2$ which reflects the duplex compression in divalent salt.
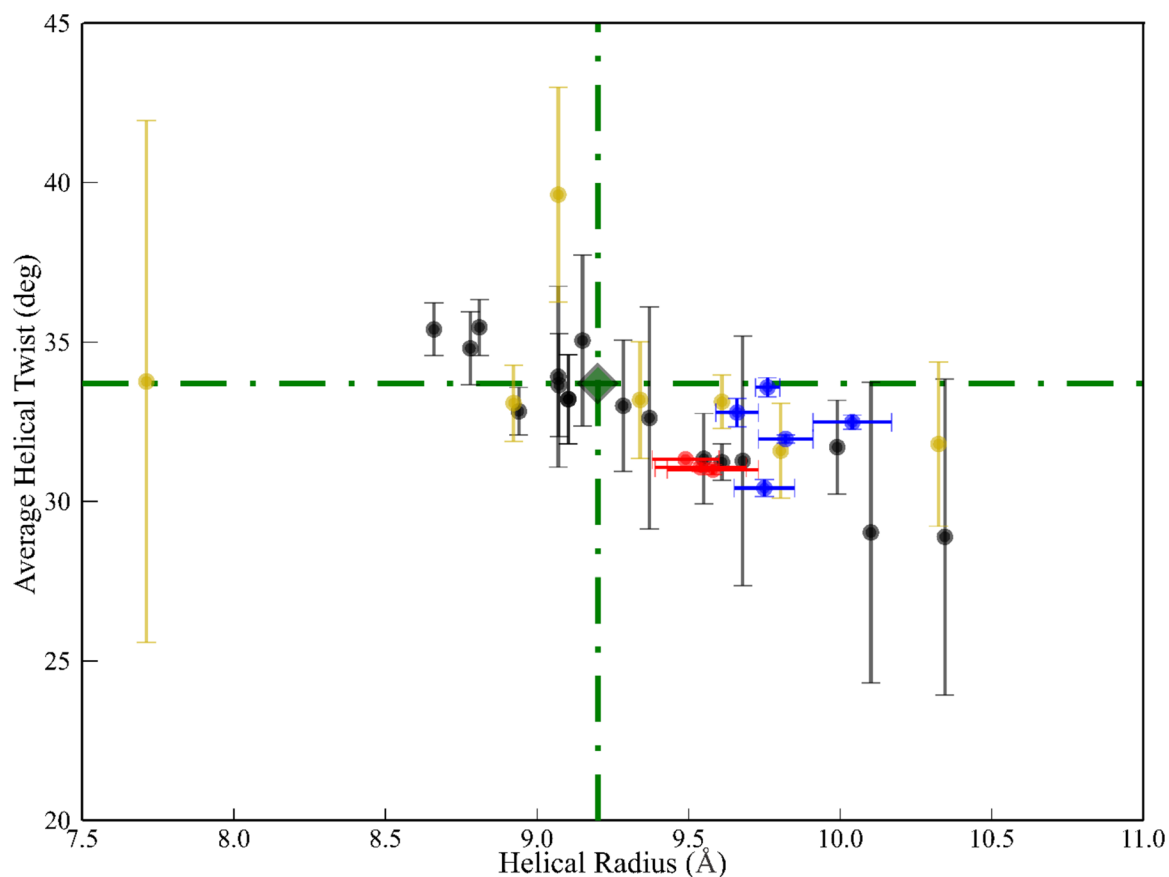
Figure 8.

Comparison of twist and radius of 23 dsRNA duplexes from the online protein data bank (pdb) with our results. These 23 dsRNA duplexes have different lengths and are found in both crystals (black) and in solution (yellow) with various agents, ions and even bound proteins. More information about these 23 duplex structures can be found in Table S1. Our refined RNA12 duplex data in KCl (blue) and $MgCl_2$ (red) are also shown. The canonical A-form RNA duplex is shown as green diamond. Most (but not all) of PDB derived duplex structures have larger helical radii and are more unwound than the A-form.
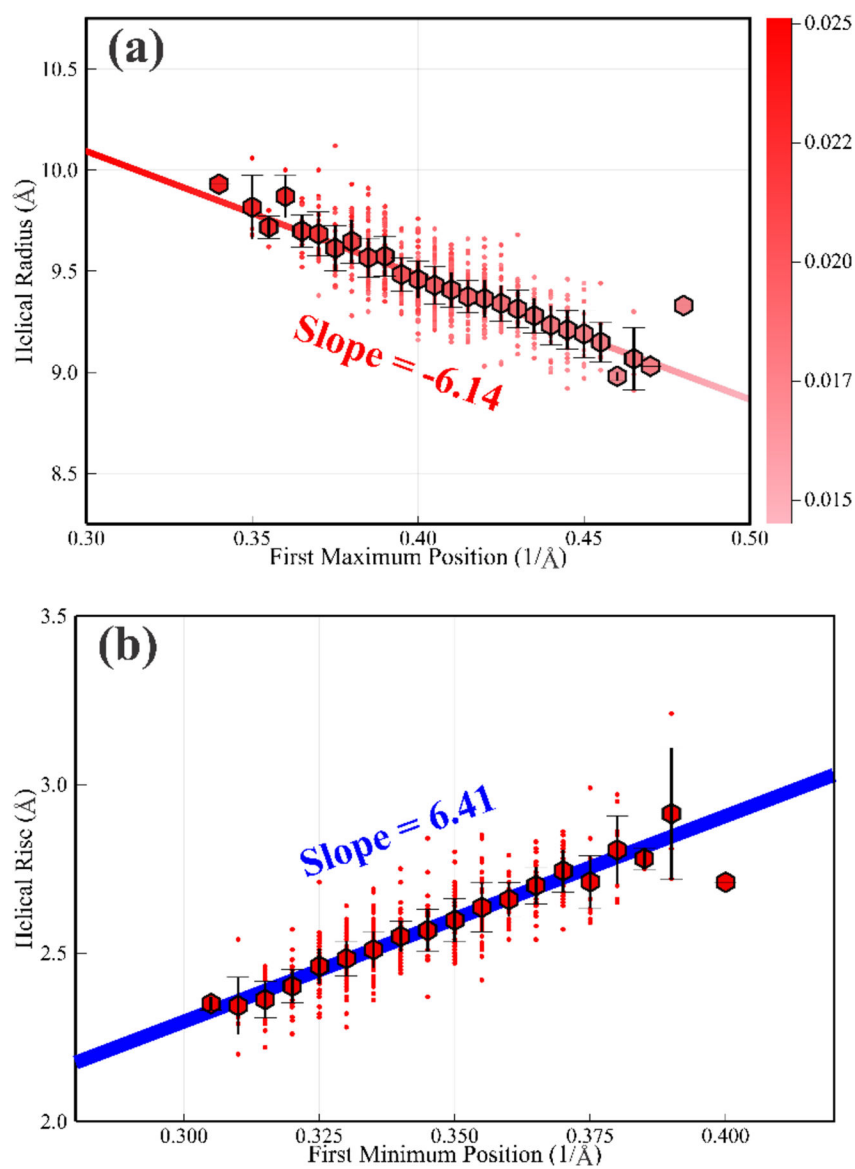
Figure 9.
The empirical correlation between the features in the calculated SWAXS curves and geometrical parameters of the duplex. (a) A linear relationship is found between the position and contrast (defined in the text) of the first maximum in the scattering profile, and the helical radius. Since helical radius gives rise to the strongest periodicity in the RNA12 duplex, the smaller the radius, the higher q the maximum shifts to with less contrast due to more structural contributions. (b) A linear relationship is found between the position of the first minimum in the profile and the average helical rise. Although the length scale of the rise in the dinucleotide step is below our resolution, the periodicity still appears in multi-nucleotide step, which falls in the resolution of our q-range.

Table 1.

A comparison of these determinations of helical radius, helical rise (in Å) and helical twist (in degrees) with previously published MD studies of dsRNA duplexes. Note that helical parameters were extracted using either *curves+*[60] or *x3dna-dssr*[55].

System	Method	Helical Radius	Helical Rise	Helical Twist
Canonical A-form	X-ray Crystal Diffraction	9.2	2.56	33.7
RNA12 in 30 mM KCl		9.66 ± 0.07	2.57 ± 0.04	32.79 ± 0.45
RNA12 in 50 mM KCl		9.75 ± 0.10	2.65 ± 0.05	30.42 ± 0.26
RNA12 in 100 mM KCl		10.04 ± 0.13	2.45 ± 0.08	32.49 ± 0.22
RNA12 in 200 mM KCl	SWAXS experiment + eEOM	9.82 ± 0.09	2.72 ± 0.03	31.96 ± 0.12
RNA12 in 500 mM KCl	(This work)	9.76 ± 0.04	2.66 ± 0.04	33.58 ± 0.30
RNA12 in 0.25 mM $MgCl_2$		9.58 ± 0.15	2.60 ± 0.01	31.00 ± 0.16
RNA12 in 1.0 mM $MgCl_2$		9.49 ± 0.11	2.66 ± 0.04	31.32 ± 0.02
RNA12 in 5.0 mM $MgCl_2$		9.54 ± 0.15	2.61 ± 0.01	31.07 ± 0.14
RNA16[61]	MD with ff99bsc0 + bsc1 and ff99bsc0+χOL3	10.14 ± 0.58	3.33 ± 0.07	31.02 ± 0.88
Canonical A-form		N/A	2.56	32.70
RNA25 in NaCl	Unconstrained MD 380 ns[10]	N/A	2.74	31.08
RNA25 in NaCl/CoHex		N/A	2.40	33.40
RNA40 and RNA16 [*†14]	MD with ff99bsc0	N/A	2.77 ± 0.15	30.57 ± 1.07
	MD with ff99bsc0+χOL3	N/A	2.79 ± 0.14	31.03 ± 0.86
RNA16 [*62]	Unconstrained MD 1 ms	N/A	2.68	32.08
$(GC)_{12}$ [*]		N/A	2.67	34.7
$(GC)_{12}$ deform 1 [*]	MD and JUMNA [‡15,63]	N/A	2.89	32.2
$(GC)_{12}$ deform 2 [*]		N/A	2.51	38.1
RNA14 with only U/A [*64]	MD Crystal simulation	N/A	N/A	31.4
	MD Solution simulation	N/A	N/A	29.8

[*] Helical parameters calculated by curves+.

[†] The analyses excluded the bp near the ends.

[‡] Values are extracted using final parameters after energy minimization in Table 1 of Ref.[15].