Predicting RNA Structure and Dynamics with Deep Learning and Solution Scattering

Edan Patt[1], Scott Classen[2], Michal Hammel[2*], Dina Schneidman-Duhovny[1*]

[1]School of Computer Science and Engineering, The Hebrew University of Jerusalem
[2]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

Correspondence: mhammel@lbl.gov, dina.schneidman@mail.huji.ac.il, edan.patt@mail.huji.ac.il, sclassen@lbl.gov

Abstract

While novel deep learning and statistics-based techniques predict accurate structural models for proteins and non-coding RNA, describing their macromolecular conformations in solution is still challenging. Small-angle X-ray scattering (SAXS) in solution is an efficient technique to validate structural predictions by comparing the experimental SAXS profile with those calculated from predicted structures. There are two main challenges in comparing SAXS profiles to RNA structures: the structures often lack cations necessary for stability and charge neutralization, and a single structure inadequately represents the conformational plasticity of RNA. We introduce Solution Conformation Predictor for RNA (SCOPER) to address these challenges. This pipeline integrates kinematics-based conformational sampling with the innovative deep-learning model, IonNet, designed for predicting Mg^{2+} ion binding sites. Validated through benchmarking against fourteen experimental datasets, SCOPER significantly improved the quality of SAXS profile fits by including Mg^{2+} ions and sampling of conformational plasticity. We observe that an increased content of monovalent and bivalent ions leads to decreased RNA plasticity. Therefore, carefully adjusting the plasticity and ion density is crucial to avoid overfitting experimental SAXS data. SCOPER is an efficient tool for accurately predicting the solution state of RNAs and providing atomistic models of their structures.

The method is available from: https://github.com/dina-lab3d/IonNet
Our pipeline is available for use as a web server: https://bilbomd.bl1231.als.lbl.gov/

Introduction

In recent years, novel and unexpected roles of non-coding RNAs have been discovered in multiple processes, such as signaling, cancer, development, and stress response (1, 2). Structural characterization of their solution conformations is critical to understanding their functional role (3). RNA flexibility challenges traditional structural characterization techniques, such as X-ray crystallography and NMR spectroscopy. While progress has been made with RNA structure determination by cryo-Electron Microscopy, challenges due to instability, heterogeneity, and small size limit widespread application (4). Consequently, RNA structures comprise only ~3% of all structures in the Protein Data Bank (PDB) (5). Small-angle X-ray scattering (SAXS) can rapidly provide in-solution structural information on biological macromolecules, describing the size, shape, and dynamics (6–9). This in-solution structural technique is experiencing a revival primarily due to improvements in data collection technologies and computational algorithms (8). Unlike X-ray crystallography and NMR spectroscopy, SAXS is a fast and reliable technique performed under dilute conditions, thus requiring minimal amounts of RNA samples. The technique has provided reliable data on particles ranging from small RNA duplexes at 20 kDa to the large 70S ribosome at 2700 kDa (9–12). SAXS can be an invaluable tool for the structural biologist, supplementing the traditional high-resolution techniques; yet, the method has limitations for RNA that merit attention.

Direct comparison of an RNA structure with a SAXS profile requires an accurate calculator of theoretical SAXS profiles from atomistic models. Current SAXS profile calculators do not accurately match the experimental SAXS RNA data within the noise (13). The difference between the theoretical and experimental SAXS profiles is attributed to the ion-induced changes in the hydration layer (14) and the conformational diversity of RNA (15, 16). The results from SAXS studies of counterion interactions with rigid RNA duplexes have greatly improved our understanding of highly charged RNA's folding into compact structures (17). Ions, commonly Mg^{2+}, also contribute signal to the SAXS profiles with the X-ray scattering length density of Mg^{2+} almost two times larger than H_2O. However, the ions are not predicted by structure modeling algorithms and can be easily missed by experimental structure determination methods. In addition to the localization of ions, we need to be able to predict their solution conformations for functional studies of RNA molecules. For flexible RNAs to describe conformational heterogeneity, multi-state models (two or more conformations and their weights) can be required (18, 19).

There are several methods to sample the conformational ensemble from a starting RNA structure, including Molecular Dynamics simulations (18), Monte Carlo-based methods, such as SimRNA (20), Normal Modes (14, 21, 22), and robotics-inspired motion planning approach (KGSRNA (23)). However, there is only one predictor of metal ion positions in RNA, MetalionRNA (24), based on a statistical analysis of ions in the PDB. MetalionRNA relies on a distance and angle-dependent anisotropic potential describing interactions between metal ions and RNA atom pairs. The statistical potential was calculated using ~100 structures when it was first published. With the availability of novel geometric deep-learning methods for protein structures and a larger dataset of RNA structures, there is an opportunity to develop more accurate tools that predict metal ion positions. More recently, works leveraging large amounts of available data have approached this problem using deep learning, such as (25), which uses 3D convolutional neural networks to comb the 3D structure of RNA as a 3D image. While this approach seems promising, 3D convolutions are not invariant to the arbitrary orientation of the RNA structure. 3D

convolutions are also notoriously slow to compute and require large amounts of data to train while also dealing with many sparse voxels that hinder the training process. The newest version of Deepmind's AlphaFold (AlphaFold3 (26)) released with it a powerful new model that can generate 3D RNA complexes along with predicted ion binding sites however, their binding site prediction has no reported accuracy measures.

We developed a SAXS-based Conformation Predictor for RNA (SCOPER), which takes an initial RNA structure and a SAXS profile as input and outputs a single or multi-state model with improved fit to the SAXS profile and Mg^{2+} ions added to RNA conformations. We used geometric deep learning to train a model named IonNet for Mg^{2+} ion placement based on the atomic neighborhood. We use SAXS profiles to select subsets of ion positions that best fit experimental SAXS profiles to identify further the most probable ion positions from the predicted ones. To address the conformational flexibility of RNA molecules, we sample multiple conformations using KGSRNA, a motion planning algorithm, followed by the prediction of Mg^{2+} ion positions for each sampled conformation. Finally, we determine single or multi-state models that fit the data within the noise. We benchmarked SCOPER's capability with 14 experimental SAXS datasets, including experimental data of three RNAs with known crystal structures (P4P6, SAM-riboswitch, and LYS-riboswitch) (27–29).

Methods

Summary. The input to our method is a PDB format structure of the RNA and a SAXS profile. The structure can be obtained experimentally or using prediction tools, such as DeepRNAFold (30), RNAComposer (31), or SimRNA (20). The method proceeds in four main stages (Fig. 1). First, we generate 1,000 RNA conformations from the input structure using KGSRNA that perturbs the structure while preserving base pairing (23). Second, IonNet predicts potential Mg^{2+} positions for each generated conformation. Third, the input SAXS profile selects a combination of Mg^{2+} ion positions that best fit the data using the goodness of fit parameter χ^2. Finally, if a single conformation with Mg^{2+} ions does not fit the SAXS profile within the noise, we apply MultiFoXS to find multi-state models to validate conformational plasticity (32). If a SAXS profile is unavailable, it is still possible to use IonNet as a standalone tool, using the code on github, to obtain the most probable Mg^{2+} ion positions using stages two and three only. Below, we describe stages two and three that we have developed specifically for this pipeline.

Prediction of ion positions with IonNet

IonNet training. IonNet is a deep learning model that can classify a probe located on an RNA surface as either an Mg^{2+} ion or a water molecule based on the atomic neighborhood of the probe (atoms within a 8Å radius). To train the model, we relied on ~1,000 PDB structures that contain ~41,000 Mg^{2+} ions to serve as positive examples. Water molecules and their atomic neighborhoods were used as negative examples (Supplementary Data). We describe these neighborhoods as graphs, where nodes correspond to RNA atoms, and edges represent the distances between the nodes. Our model performs its classification using Graph Neural Networks (GNNs) leveraging graph attention (33) and graph convolution (34) layers to extract information from the input graphs (Fig. S1, Table S1). Four-fold cross-validation was performed with IonNet, resulting in an AUROC mean of 0.89 (Fig. 2A). We provide a complete overview of IonNet's training and accuracy metrics in the Supplementary Data (Table S3, Fig S6-11).

IonNet inference. We generate surface probes to predict ion positions in a given RNA structure (Fig. 2B) using Connolly's surface method (35). A neighborhood of RNA atoms (within 8Å) is extracted for each such probe. Each neighborhood is passed to IonNet and is classified as either an Mg^{2+} ion or a water molecule neighborhood. Iterative clustering selects the predicted binding sites with the highest confidence. By the end of this stage, IonNet suggests a number of plausible Mg^{2+} binding-site locations. For example, IonNet could identify probes in the vicinity of 11 out of 12 experimentally observed ions for the P4P6 structure (Fig. 2C). Overall, out of 5,500 probes, IonNet identified 84 as Mg^{2+} ions. If we use a strict cut-off (0.9 instead of 0.5) for Mg^{2+} ion prediction, 6 out of the 12 experimentally observed ions are correctly predicted, with 30 predicted Mg^{2+} ion positions (Fig. 2D).

Combinatorial SAXS-based selection of ions

SAXS profile fitting. We computed the profiles using the Debye formula with the FoXS program (13, 36). For a given structural model, there are three adjustable parameters: excluded volume (c_1), hydration layer density (c_2), and the scaling factor (c) that are optimized to fit the experimental profile as measured by the χ^2 score:

$$\chi^2 = \frac{1}{S}\sum_{i=1}^{S}\left(\frac{I_{exp}(q_i) - cI(q_i, c_1, c_2)}{\sigma(q_i)}\right)^2$$

where $I_{exp}(q)$ is the experimental profile, which is a function of the momentum transfer $q = 4\pi sin\theta/\lambda$, where 2θ is the scattering angle, λ is the wavelength of the incident X-ray beam, $\sigma(q)$ is the experimental error of the measured profile, and S is the number of points in the profile. $I(q, c_1, c_2)$ is the computed profile, given by the Debye formula (13), with c_1 and c_2 optimized to minimize the χ^2 score.

Enumeration of Mg^{2+} subsets

In the pipeline's 3rd stage, the subset of Mg^{2+} ions that minimizes the χ^2 score is selected from a set of predicted Mg^{2+} ion positions by IonNet. These subsets are enumerated using a branch-and-bound algorithm. First, we calculate the χ^2 score for the RNA structure with a single ion. Second, we enumerate all possible subsets with two ions (a branch step) and keep K best scoring ones (a bound step) for the next iteration, when a third ion will be added. We continue enumerating the ion subsets of size N based on subsets of size $N - 1$ until the score can no longer be improved. To speed up the calculation of the SAXS profile and the corresponding χ^2 score for each subset, we precompute the RNA profile and the profile of each Mg^{2+} ion relative to the RNA. This preprocessing enables a rapid summation of the relevant sub-profiles for each subset of Mg^{2+} ions and χ^2 score calculation.

SAXS data collection

To validate our pipeline, we collect experimental SAXS data from a monomeric RNA state free of contaminants (higher oligomeric states and aggregation). We applied Size Exclusion Chromatography coupled with small angle X-ray scattering (SEC-SAXS) (Tables 1 and 2). All data, except RNA #10 and #14, were collected at SIBYLS beamline 12.3.1 at Advanced Light Source with recently reported developed SEC-SAXS-MALS modality (37). X-ray wavelength was set at λ=1.127 Å, and the sample to detector distance was 2100 mm, resulting in scattering vectors, q, ranging from 0.01 $Å^{-1}$ to 0.4 $Å^{-1}$. 60-95

μL of annealed RNAs with a concentration between 1 and 3 mg/ml were prepared in the SEC running buffer (Table S2). The Shodex KW802.5 column was equilibrated with a running buffer with a flow rate of 0.65 mL/min. Each sample was injected in an SEC column, and two-second X-ray exposures were recorded continuously for 24 min. RNA #10 and #14 were previously collected at the SIBYLS beamline and reported (12, 29). Program RAW (38) was used for further SEC-SAXS processing, including buffer subtractions and merging SAXS frames across the elution peak (Fig. S2). For RNA #8, we further applied the EFA approach (39) to deconvolute the SAXS signal of monomeric RNA from the larger oligomeric species and misfolded species. The final merged SAXS curves were further used for Guinier analysis using the program RAW (38) and computing P(r) functions by the program GNOM (40). The MW_{SAXS} was calculated by Volume of Correlation (41) and compared to the molecular weight estimated by SEC-MALS (Table S2). The SEC-SAXS were deposited to the SIMPLE SCATTERING database https://simplescattering.com/ (deposition IDs are listed in Table S2).

Results

Impact of ionic strength on SAXS fitting.

Fast calculation of SAXS profiles from protein structures usually involves modeling the hydration shell with implicit solvent models (13, 42–45). In these models, the density of the hydration layer can be adjusted to optimize the fit to the experimental SAXS profile (13, 42). The changes in the protein's hydration layer, typically influenced by the protein surface net charge, result from the varying ionic strength of the buffer used in SAXS experiments (13). However, in the nucleic acid world, higher ionic strength increases the rigidity of nucleic acid structures (15) and can alter the overall arrangement of RNA segments (46, 47). The ionic strength-dependent plasticity of RNA poses a greater difficulty in quantifying changes in the RNA hydration layer using SAXS. Nevertheless, here, we look at the impact of salt concentration on the hydration layer model in the FoXS SAXS calculator (13) and RNA plasticity.

We measured SAXS profiles for a small RNA stem-loop (#3 in the benchmark) and P4P6 RNA (#13 in the benchmark) at different ionic strengths while keeping the Mg^{2+} concentration constant at 5 mM. The experimental P(r) functions showed increased RNA rigidity by smoothening oscillations at higher ionic strengths (Fig. 3A, B).

Initially, we attempted to fit the experimental SAXS profiles with a single profile calculated for each one of the 1,000 conformations sampled by KGSRNA (without Mg^{2+} ions), adjusting the fitting parameters (c_1 - excluded volume density and c_2 - hydration layer density; Table 1). As expected, the best-fitting conformer (with the lowest χ^2) did not match the SAXS profiles within the noise (Table 1), indicating that adding Mg^{2+} ions or fitting a multi-state model (32) that models RNA plasticity is necessary to match the experimental data.

The fit quality improved for the best-fitting conformer when Mg^{2+} ions were added (Table 1, Fig. S4). Additionally, the hydration layer density c_2 parameter did not increase with increasing ionic strength (Table 1), suggesting that c_2 adjustment may compensate for the lack of RNA plasticity. A similar improvement in the fit quality was observed with the two-state model without Mg^{2+} ions. However, the c_2 parameters have extreme values (maximal density of 2.0 for P4P6 and minimal density of -0.5 for the

stem-loop, Table 1), suggesting that the adjustment of the hydration layer may compensate for the lack of Mg^{2+} ions. Although the fit quality for both approaches is similar (Fig. 3C, Table 1), the derived models differ significantly (Fig. 3D) and further indicate that hydration layer adjustment can lead to overfitting experimental data.

Next, we investigated whether single or multi-state models with added Mg2+ ions fit the SAXS data well without optimizing the hydration layer. The single best-fitting conformer with Mg^{2+} ions and default c_1/c_2 parameters slightly worsened the fit (Table 1). However, for stem-loop RNA, the pipeline selected the same conformation across all three salt concentrations with similar placement of Mg^{2+} ions (Fig. S4C). It also placed Mg2+ ions in P4P6 at similar positions, with slightly more open conformation selected for the 50 mM KCl condition, which is consistent with increased RNA plasticity at low salt concentration (Fig. S4D). Finally, modeling RNA plasticity by multi-state models with Mg^{2+} ions resulted in an excellent fit for all salt concentrations without needing c_1/c_2 adjustment (Table 1).

Therefore, to avoid data overfitting by simultaneous RNA conformational sampling, adjustment of the hydration layer, and placement of Mg^{2+} ions, we set the c_1 and c_2 parameters to 1.0 in our further pipeline testing.

SAXS profile fitting

Our dataset consisted of 14 RNA samples with experimental SAXS profiles. The starting structures were obtained from DeepFoldRNA (#1-4, #6-8) (30), RNAComposer (#5, #9, #11) (31), and X-ray crystallography structures (#10 PDB 2GIS, #13, PDB 1GID, #14 PDB 3D0U) (28, 29, 48). We used different programs to predict our initial structures mainly due to time constraints, where RNAComposer produced initial fitting structures faster than DeepFoldRNA. DeepFoldRNA was used when the RNAComposer model fitted the SAXS data poorly. Our pipeline is expected to benefit from future improvements in RNA structure prediction that will provide more accurate starting structures (49). To avoid overfitting using the excluded volume and hydration layer parameters, both c_1 and c_2 were set to 1.0 in all profile calculations using both FoXS and MultiFoXS. The fit of most starting structures to their corresponding SAXS profile was poor (Table 2, Fig. 4). We predicted Mg^{2+} ion positions using IonNet. Adding Mg^{2+} ions to the starting structure only marginally improved the fit (Table 2, Fig. 4, Fig. 5). However, in the case of the accurately determined starting structure of the SAM-riboswitch (#10) obtained from X-ray crystallography, the addition of Mg^{2+} ions resulted in significant improvement (Table 2, Fig. 4). The SAM-riboswitch is a well-folded globular RNA where the addition of a relatively large number of placed Mg^{2+} ions contributes to the improvement of the SAXS profile fit. Significant improvement in SAXS fit could also be observed for some small RNAs (#1, #3, #5) with simpler folds and expected limited plasticity. Thus, the starting predicted structure represents the solution state, and placing Mg^{2+} ions improves the SAXS fit. However, selecting a more accurate conformation is key to matching the experimental SAXS profiles except for the RNA-riboswitch (#10).

In all of our benchmark cases, the single conformation with the best fit (lowest χ^2) out of the 1,000 sampled by KGSRNA could fit the data significantly better than the initial structure (Table 2, Fig. 4, Fig. 5). Addition of Mg^{2+} ions to these conformations further improved the fit (Table 2, Fig. 4, Fig. 5). Additional improvement in the fit to SAXS profiles could be achieved using multi-state models that mimic RNA plasticity (Table 2). Particularly, the multi-segment RNAs where the larger movement of segments could be expected (#11, #13, and #14) (46) showed significant fit improvement with a

multi-state model. For example, the SAXS profile of the P4P6 had the best fit with four conformations (Fig. S3) with a χ^2 =1.17 vs. 1.48 for a single conformation. However, the different placements of Mg^{2+} in each conformer (Fig. 3) suggest that placing Mg^{2+} ions in combination with multi-state modeling can lead to the overfitting of experimental SAXS data. On the other hand, it may also suggest that the movement of individual P4P6 segments in solution can result in more loosely defined coordinates for Mg^{2+} ions. Generally, it is difficult to validate through SAXS whether the Mg^{2+} ions observed in the P4P6 crystal structure remain ordered in the dynamic solution state. Smaller RNAs with single-stranded regions (#3) also show improved fit with a multi-state model. The multistate model of the #3 RNA mimics the plasticity of the unpaired 5'-end, resulting in an χ^2 value of 1.22 compared to 1.52 for a single conformation (Table 2).

These results indicate that our pipeline can predict Mg^{2+} binding sites that significantly reduce the χ^2 score of the RNA structure (Figs. 4-5, Table 2). Although the improvements to the fit by adding Mg^{2+} ions were rather marginal compared to finding a suitable conformation, these additions are not insignificant (Table 2, Figs. 4-5). We conducted our profile fit calculations over all 1,000 conformations generated by KGSRNA. We found that running IonNet on all 1,000 conformations can be redundant if a single conformation fits the data within the noise. The best-fitting conformation without Mg^{2+} ions tends to be one of the best-fitting conformations with added ions (Fig. S5). If a user lacks parallel computing power, a safe assumption is that our pipeline can be run only on the best-scoring structures sampled by KGSRNA with little cost to the optimality of the output. However, if multi-state modeling is needed to fit the data within the noise, one needs to run IonNet for all the conformations.

Correlation between RNA size and a number of potential Mg^{2+} binding sites

We found a positive correlation (Pearson coefficient of 0.67) between the number of atoms in an RNA structure and the amount of predicted Mg^{2+} ions for the structure (Fig. 6). This is an expected outcome as the more surface area the structure has, it would seem likely that there would be more Mg^{2+} binding sites to stabilize the structure.

SCOPER Webapp

We implemented a straightforward and easy-to-use web-based application to make the SCOPER pipeline available to the scientific community, particularly those without the knowledge or resources to install, configure, and run the Python code directly from the GitHub repository. The SCOPER pipeline has been added as an option within our existing BILBOMD job runner framework, available at https://bilbomd.bl1231.als.lbl.gov/. Users are required to create an account before submitting jobs. The required inputs are the initial RNA structure in PDB format and the experimental SAXS curve containing three columns (q in $Å^{-1}$ unit, intensities, and experimental error.). The SCOPER web server implements the following pipeline steps (Fig. 1): (i) conformational sampling by KGSRNA, (ii) selection of the best conformer using FoXS, (iii) prediction of possible locations of Mg^{2+} ions using IonNet, and (iv) selection of the best Mg^{2+} ion placements by fitting the SAXS data using MultiFoXS. Users are notified by email when their jobs are complete. Results are then available to view directly within the web app (Fig. 7). They are also available to download as a compressed file containing the original uploaded RNA PDB file and the SAXS fit files, along with the output from the SCOPER pipeline.

Discussion

SCOPER is a novel pipeline that integrates a deep learning model for placing Mg^{2+} ions into the RNA models or experimental structure, with SAXS-based validation to determine dynamic RNA conformation in solution. The starting RNA structure can be obtained using any RNA modeling software where adding Mg^{2+} ions in RNA is important for validating RNA structures in solution by experimental SAXS. For this purpose, we have developed IonNet, a novel, deep learning-based model to predict Mg^{2+} binding sites. Results show that these binding sites are predicted with high accuracy. Support for additional ions commonly found as structure stabilizers in the vicinity of RNA, such as Na^{+} and K^{+} can be added with sufficient training data in future works.

The final ion positions and multi-state models are selected based on a fit to SAXS data. However, data overfitting is possible when predictions of ions used in conjunction with experimental SAXS select a subset of Mg^{2+} ions that only improve the fit. To reduce overfitting, we recommend not varying adjustable SAXS fitting parameters, where adjusting the hydration layer may lead to the incorrect placement of Mg^{2+}. However, with only three SAXS datasets (S-adenosylmethionine riboswitch #10, P4P6 #13, and lysine riboswitch #14) that have an experimentally verified structure, we cannot be sure that our pipeline entirely prevents overfitting. We suggest using a high threshold (0.3-0.5 for our model) to obtain only the most confident IonNet predictions.

IonNet can be improved further by considering Mg^{2+} coordination explicitly when classifying binding sites. Most Mg^{2+} ions are normally coordinated by oxygen and nitrogen atoms (50). In future work, we plan an additional data processing step that classifies the coordination neighborhood based on distances to oxygen and nitrogen atoms. This classification can help improve the model's precision even further.

We show that an accurate RNA SAXS calculator relies on the presence of Mg^{2+} ions in the model. The often used extreme adjustment of the hydration layer, in the absence of Mg^{2+} ions, can lead to the wrong selection of RNA conformer by SAXS fitting. Optimizing the predicted RNA conformer is the most important step in delivering a good agreement with experimental SAXS. Although the prediction of Mg2+ ions in the RNA structure delivers a secondary improvement in the SAXS fit, it increases confidence in the RNA structure validation. This is particularly important when novel RNA structure prediction, like Alphafold 3 (26), needs to be validated.

Overall, SCOPER's results prove that it can find more accurate RNA structures and suggest their plasticity that better represents the solution state. The fact that the multistate model did not significantly improve the SAXS fit in most cases suggests that accounting for the Mg^{2+} in the RNA model is important to properly fit the experimental SAXS data. Here, we show that precisely calculated theoretical SAXS profiles can be used to validate RNA structure predictions.

Data Availability: The SEC-SAXS were deposited to the SIMPLE SCATTERING database (simplescattering.com). Additionally, the depositions contain the final merged SAXS curve and the final atomistic models used to calculate the SAXS fit (see Table S2). The RNA samples #4, #5, #7, #8, #9, #11, #12 were collected under a proprietary agreement and unavailable in the SIMPLE SCATTERING database. The method is available from https://github.com/dina-lab3d/IonNet. For the latest release, please use our Zenodo link dina-lab3D/IonNet: Zenodo Release A SCOPER pipeline is available as a web server from https://bilbomd.bl1231.als.lbl.gov/.

Funding:
This research was partly supported by National Cancer Institute grants for Structural Biology of DNA Repair (SBDR) CA092584 to M.H, The U.S.-Israel Binational Science Foundation (BSF) 2016070. SAXS data collection at SIBYLS is funded through NIGMS grant P30 GM124169-01, ALS-ENABLE, and the IDAT program of the US Department of Energy Office of Biological and Environmental Research. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

References:

1. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Human Molecular Genetics*, **15**, R17–R29.

2. Esteller,M. (2011) Non-coding RNAs in human disease. *Nature Reviews Genetics*, **12**, 861–874.

3. Novikova,I.V., Hennelly,S.P., Tung,C.-S. and Sanbonmatsu,K.Y. (2013) Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs. *Journal of Molecular Biology*, **425**, 3731–3746.

4. Ma,H., Jia,X., Zhang,K. and Su,Z. (2022) Cryo-EM advances in RNA structure determination. *Signal Transduct Target Ther*, **7**, 58.

5. Berman,H.M. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235–242.

6. Crist,B. (1983) A review of: *'SMALL ANGLE X-RAY SCATTERING'*, edited by O. Glatter and O. Kratky (Universitat Graz, Austria) New York: Academic Press, 1982, 515 pp. $89.50. ISBN 0-12-286280. *Chemical Engineering Communications*, **22**, 377–378.

7. Feigin,L.A. and Svergun,D.I. (1987) Structure Analysis by Small-Angle X-Ray and Neutron Scattering. 10.1007/978-1-4757-6624-0.

8. Putnam,C.D., Hammel,M., Hura,G.L. and Tainer,J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, **40**, 191–285.

9. Hammel,M. (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *European Biophysics Journal*, **41**, 789–799.

10. Svergun,D.I., Burkhardt,N., Skov Pedersen,J., Koch,M.H.J., Volkov,V.V., Kozin,M.B., Meerwink,W., Stuhrmann,H.B., Diedrich,G. and Nierhaus,K.H. (1997) Solution scattering structural analysis of the 70 s Escherichia coli ribosome by contrast variation. II † . A model of the ribosome and its RNA at 3.5 nm resolution 1 †Paper I in this series is the accompanying paper, Svergun et al. (1997) 1Edited by M. F. Moody. *Journal of Molecular Biology*, **271**, 602–618.

11. Lipfert,J. and Doniach,S. (2007) Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. *Annual Review of Biophysics and Biomolecular Structure*, **36**, 307–327.

12. Rambo,R.P. and Tainer,J.A. (2010) Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA*, **16**, 638–646.

13. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.

14. Kirmizialtin,S., Pabit,S.A., Meisburger,S.P., Pollack,L. and Elber,R. (2012) RNA and Its Ionic Cloud: Solution Scattering Experiments and Atomically Detailed Simulations. *Biophysical Journal*, **102**, 819–828.

15. Chen,H., Meisburger,S.P., Pabit,S.A., Sutton,J.L., Webb,W.W. and Pollack,L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences*, **109**, 799–804.

16. Drozdetski,A.V., Tolokh,I.S., Pollack,L., Baker,N. and Onufriev,A.V. (2016) Opposing Effects of Multivalent Ions on the Flexibility of DNA and RNA. *Physical Review Letters*, **117**.

17. Chen,Y. and Pollack,L. (2016) SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip. Rev. RNA*, **7**, 512–526.

18. Chen,Y.-L., Lee,T., Elber,R. and Pollack,L. (2019) Conformations of an RNA Helix-Junction-Helix Construct Revealed by SAXS Refinement of MD Simulations. *Biophys. J.*, **116**, 19–30.

19. Dethoff,E.A., Chugh,J., Mustoe,A.M. and Al-Hashimi,H.M. (2012) Functional complexity and regulation through RNA dynamics. *Nature*, **482**, 322–330.

20. Boniecki,M.J., Lach,G., Dawson,W.K., Tomala,K., Lukasz,P., Soltysinski,T., Rother,K.M. and Bujnicki,J.M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.

21. Zacharias,M. and Sklenar,H. (2000) Conformational deformability of RNA: a harmonic mode analysis. *Biophys. J.*, **78**, 2528–2542.

22. Sabei,A., Baia,T.G.C., Saffar,R., Martin,J. and Frezza,E. Internal Normal Mode Analysis applied to RNA flexibility and conformational changes. 10.1101/2022.11.30.518608.

23. Fonseca,R., van den Bedem,H. and Bernauer,J. (2015) KGSrna: Efficient 3D Kinematics-Based Sampling for Nucleic Acids. *Lecture Notes in Computer Science*, 10.1007/978-3-319-16706-0_11.

24. Philips,A., Milanowska,K., Lach,G., Boniecki,M., Rother,K. and Bujnicki,J.M. (2012) MetalionRNA: computational predictor of metal-binding sites in RNA structures. *Bioinformatics*, **28**, 198–205.

25. Zhou,Y. and Chen,S.-J. (2022) Graph deep learning locates magnesium ions in RNA. *QRB Discov*, **3**.

26. Abramson,J., Adler,J., Dunger,J., Evans,R., Green,T., Pritzel,A., Ronneberger,O., Willmore,L., Ballard,A.J., Bambrick,J., *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 10.1038/s41586-024-07487-w.

27. Battle,D.J. and Doudna,J.A. (2002) Specificity of RNA-RNA helix recognition. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 11676–11681.

28. Montange,R.K. and Batey,R.T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172–1175.

29. Garst,A.D., Héroux,A., Rambo,R.P. and Batey,R.T. (2008) Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.*, **283**, 22347–22351.

30. Pearce,R., Omenn,G.S. and Zhang,Y. *De Novo* RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials from Deep Learning. 10.1101/2022.05.15.491755.

31. Biesiada,M., Purzycka,K.J., Szachniuk,M., Blazewicz,J. and Adamiak,R.W. (2016) Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol. Biol.*, **1490**, 199–215.

32. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Research*, **44**, W424–W429.

33. Veličković,P., Cucurull,G., Casanova,A., Romero,A., Liò,P. and Bengio,Y. (2017) Graph Attention Networks. 10.48550/arXiv.1710.10903.

34. Rossi,E., Monti,F., Bronstein,M. and Liò,P. (2019) ncRNA Classification with Graph Convolutional Networks. 10.48550/arXiv.1905.06515.

35. Connolly,M.L. (1983) Analytical molecular surface calculation. *Journal of Applied Crystallography*, **16**, 548–558.

36. Schneidman-Duhovny,D., Hammel,M. and Sali,A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research*, **38**, W540–W544.

37. Rosenberg,D.J., Hura,G.L. and Hammel,M. (2022) Size exclusion chromatography coupled small angle X-ray scattering with tandem multiangle light scattering at the SIBYLS beamline. *Methods Enzymol.*, **677**, 191–219.

38. Hopkins,J.B., Gillilan,R.E. and Skou,S. (2017) *BioXTAS RAW*: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, **50**, 1545–1553.

39. Meisburger,S.P., Taylor,A.B., Khan,C.A., Zhang,S., Fitzpatrick,P.F. and Ando,N. (2016) Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. *J. Am. Chem. Soc.*, **138**, 6506–6516.

40. Semenyuk,A.V. and Svergun,D.I. (1991) GNOM – a program package for small-angle scattering data processing. *Journal of Applied Crystallography*, **24**, 537–540.

41. Rambo,R.P. and Tainer,J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477–481.

42. Svergun,D., Barberato,C. and Koch,M.H.J. (1995) CRYSOL– a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.

43. Liu,H., Morris,R.J., Hexemer,A., Grandison,S. and Zwart,P.H. (2012) Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallogr. A*, **68**, 278–285.

44. Poitevin,F., Orland,H., Doniach,S., Koehl,P. and Delarue,M. (2011) AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models. *Nucleic Acids Res.*, **39**, W184–9.

45. Virtanen,J.J., Makowski,L., Sosnick,T.R. and Freed,K.F. (2011) Modeling the hydration layer around proteins: applications to small- and wide-angle x-ray scattering. *Biophys. J.*, **101**, 2061–2069.

46. Bai,Y., Chu,V.B., Lipfert,J., Pande,V.S., Herschlag,D. and Doniach,S. (2008) Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble. *J. Am. Chem. Soc.*, **130**, 12334–12341.

47. Bruetzel,L.K., Gerling,T., Sedlak,S.M., Walker,P.U., Zheng,W., Dietz,H. and Lipfert,J. (2016) Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-ray Scattering. *Nano Lett.*, **16**, 4871–4879.

48. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) CRYSTAL STRUCTURE OF A GROUP I RIBOZYME DOMAIN: PRINCIPLES OF RNA

PACKING. 10.2210/pdb1gid/pdb.

49. Das,R., Kretsch,R.C., Simpkin,A.J., Mulvaney,T., Pham,P., Rangan,R., Bu,F., Keegan,R.M., Topf,M., Rigden,D.J., *et al.* (2023) Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins*, **91**, 1747–1770.

50. Zheng,H., Shabalin,I.G., Handing,K.B., Bujnicki,J.M. and Minor,W. (2015) Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection. *Nucleic Acids Res.*, **43**, 3789–3801.

Structure/ salt concentration	w.o Mg^{2+}				w. Mg^{2+}			w. Mg^{2+} c_1=1.0 / c_2 =1.0		
	Best single conformer		2-state model		Best single conformer			Best single conformer		Multi-state model
	χ^2	c_2	χ^2	c_2	χ^2	c_2	#ions	χ^2	#ions	χ^2
P4P6-50	1.90	2.50	1.63	2.0	1.49	2.00	14	2.45	20	1.42 w. Mg^{2+} 1.52 w.o Mg^{2+}
P4P6-150	1.96	0.56	1.39	2.0	1.63	0.85	16	1.74	17	1.39 w. Mg^{2+} 1.84 w.o Mg^{2+}
P4P6-300	1.78	0.45	1.31	2.0	1.40	0.84	18	1.48	20	1.17 w. Mg^{2+} 1.21 w.o Mg^{2+}
stem-loop-50	1.90	-0.14	1.50	-0.5	1.67	-0.46	11	1.99	9	1.82 w. Mg^{2+} 1.45 w.o Mg^{2+}
stem-loop-150	1.98	-0.08	1.51	-0.5	1.64	-0.50	11	1.96	9	1.72 w. Mg^{2+} 1.37 w.o Mg^{2+}
stem-loop-300	1.65	0.48	1.42	-0.5	1.40	0.55	9	1.52	10	1.22 w. Mg^{2+} 1.08 w.o Mg^{2+}

Table 1. Concentration results

RNA	Starting Structure		Best Scoring Structure		MultiFoXS		Number of States
	w.o Mg^{2+}	w. Mg^{2+}	w.o Mg^{2+}	w. Mg^{2+}	w.o Mg^{2+}	w. Mg^{2+}	
1	110.73	73.15	36.79	12.65	26.90	11.95	3
2	42.67	42.67	5.12	2.92	4.64	2.92	3
3	18.95	10.52	1.76	1.52	1.08	1.22	6
4	8.70	8.70	3.04	2.47	2.94	2.41	3
5	14.91	9.90	5.95	3.86	4.27	3.35	3
6	2.18	1.97	2.02	1.94	1.87	1.89	3
7	1.92	1.55	1.03	1.00	0.96	0.95	6
8	4.65	3.35	3.15	2.43	2.71	2.37	4
9	26.79	25.56	7.42	5.72	6.84	5.56	2
10	17.25	3.33	5.60	1.77	4.00	1.77	2
11	7.35	7.12	2.51	2.24	2.14	2.04	4
12	5.26	3.84	3.34	2.66	3.34	2.66	2
13	26.49	23.24	1.86	1.48	1.21	1.17	4
14	7.87	7.83	2.62	1.62	0.69	0.69	4

Table 2. χ^2 values for our fitting experiments (corresponding to plots in Fig. 4.)

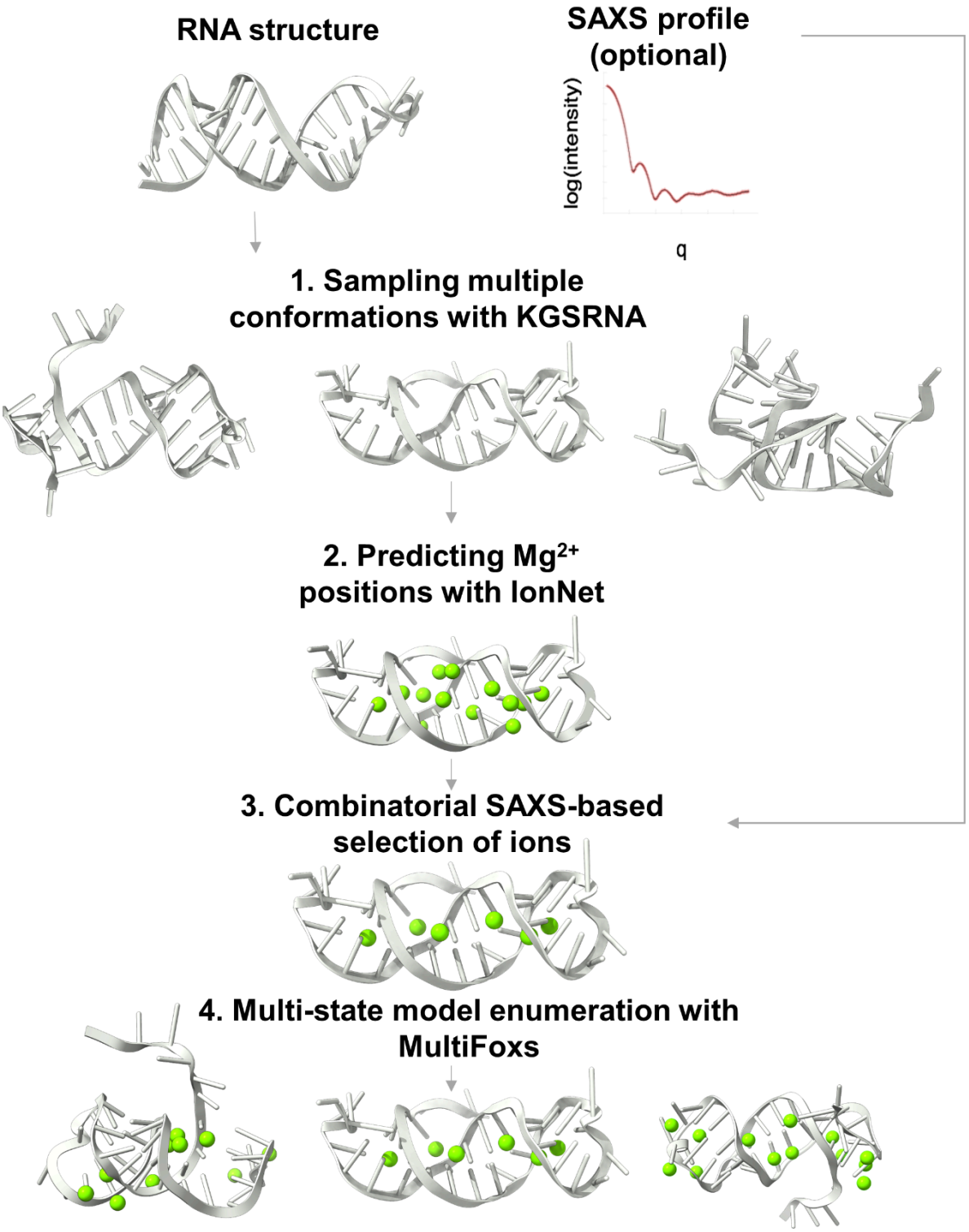


Figure 1. Visualization of SCOPER pipeline.

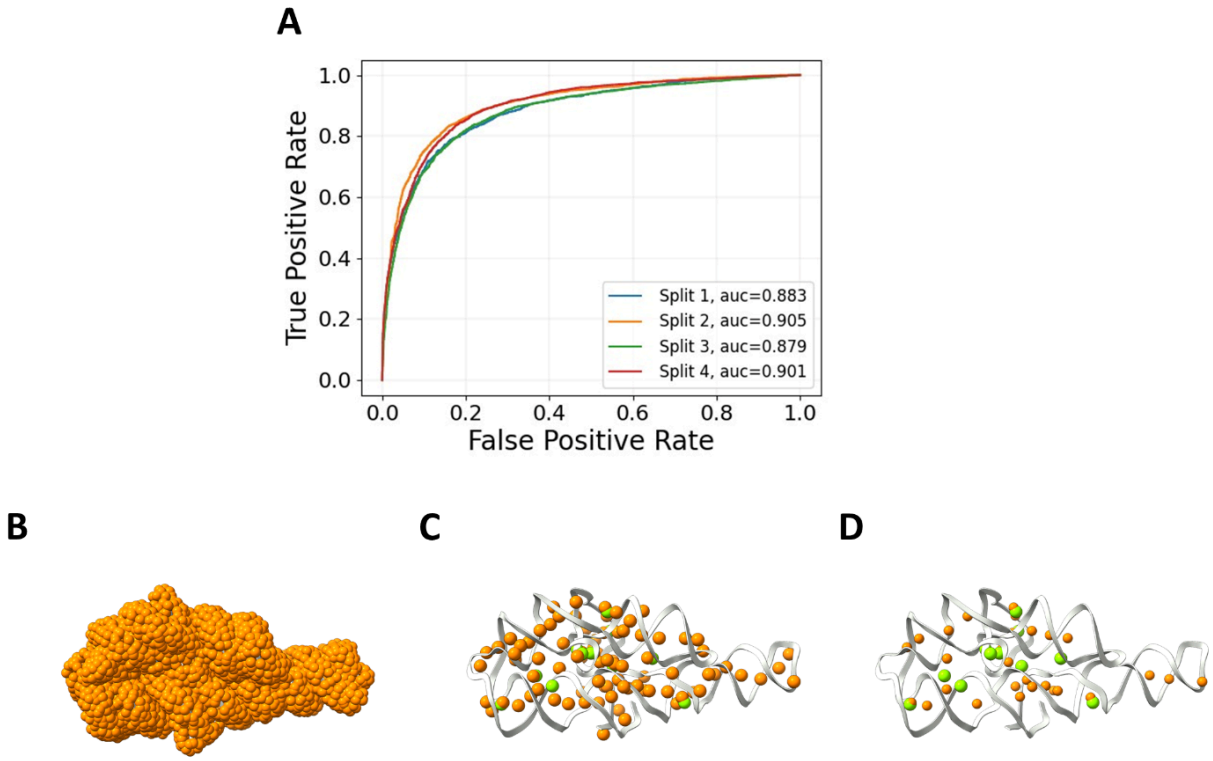


Figure 2: IonNet performance. A. Four-fold cross-validation of our best model over the whole test set. **B-D**. IonNet prediction of Mg^{2+} ions for the P4P6 structure (PDB 1GID). Ions are predicted from the set of 5,500 probes (orange) covering the whole structure (B) with high-confidence predictions above a threshold of 0.5 with ~80 predictions (C) and 0.9 with ~25 predictions (D). The experimentally observed Mg^{2+} ions are in green (12 in total).

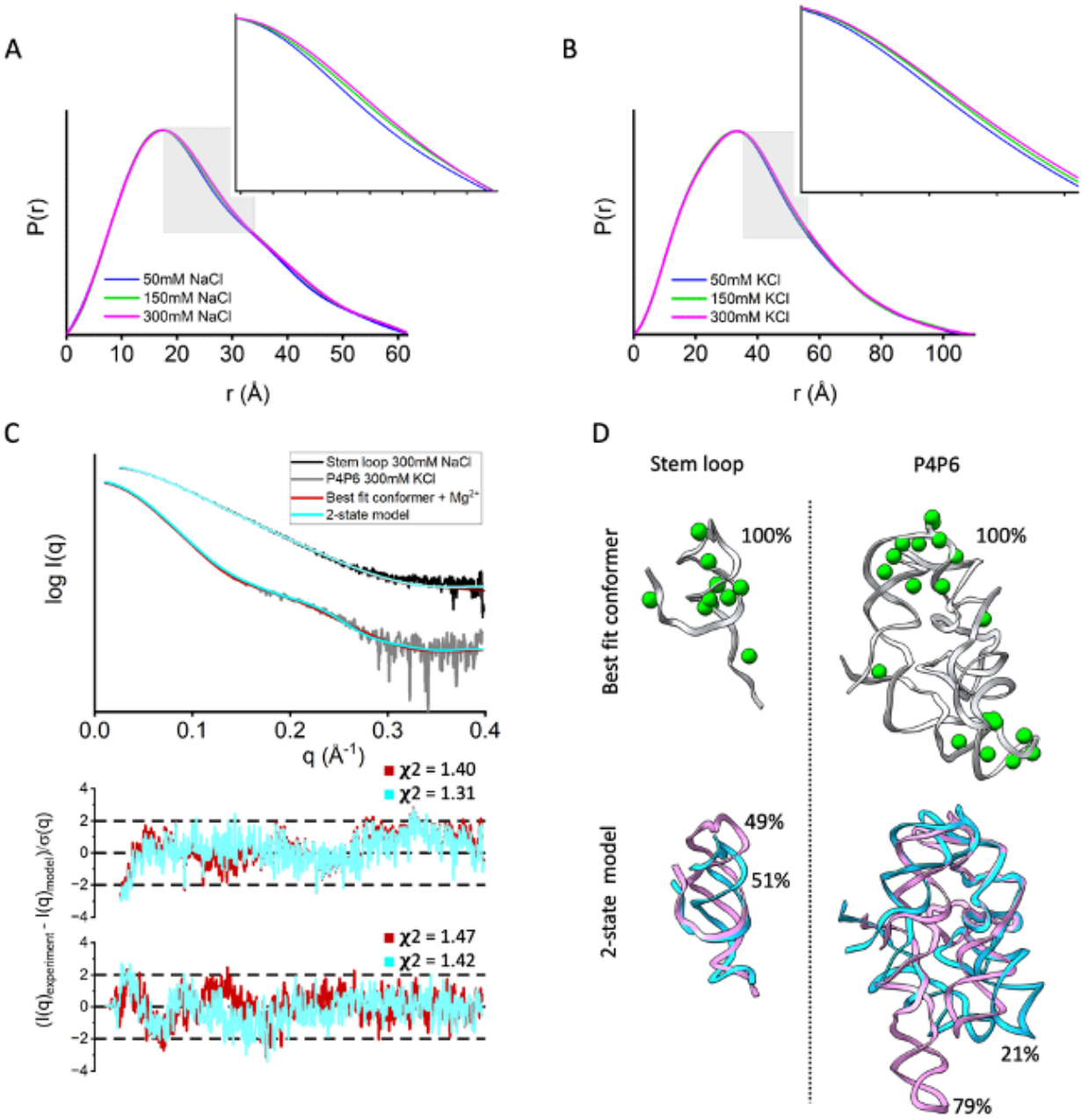
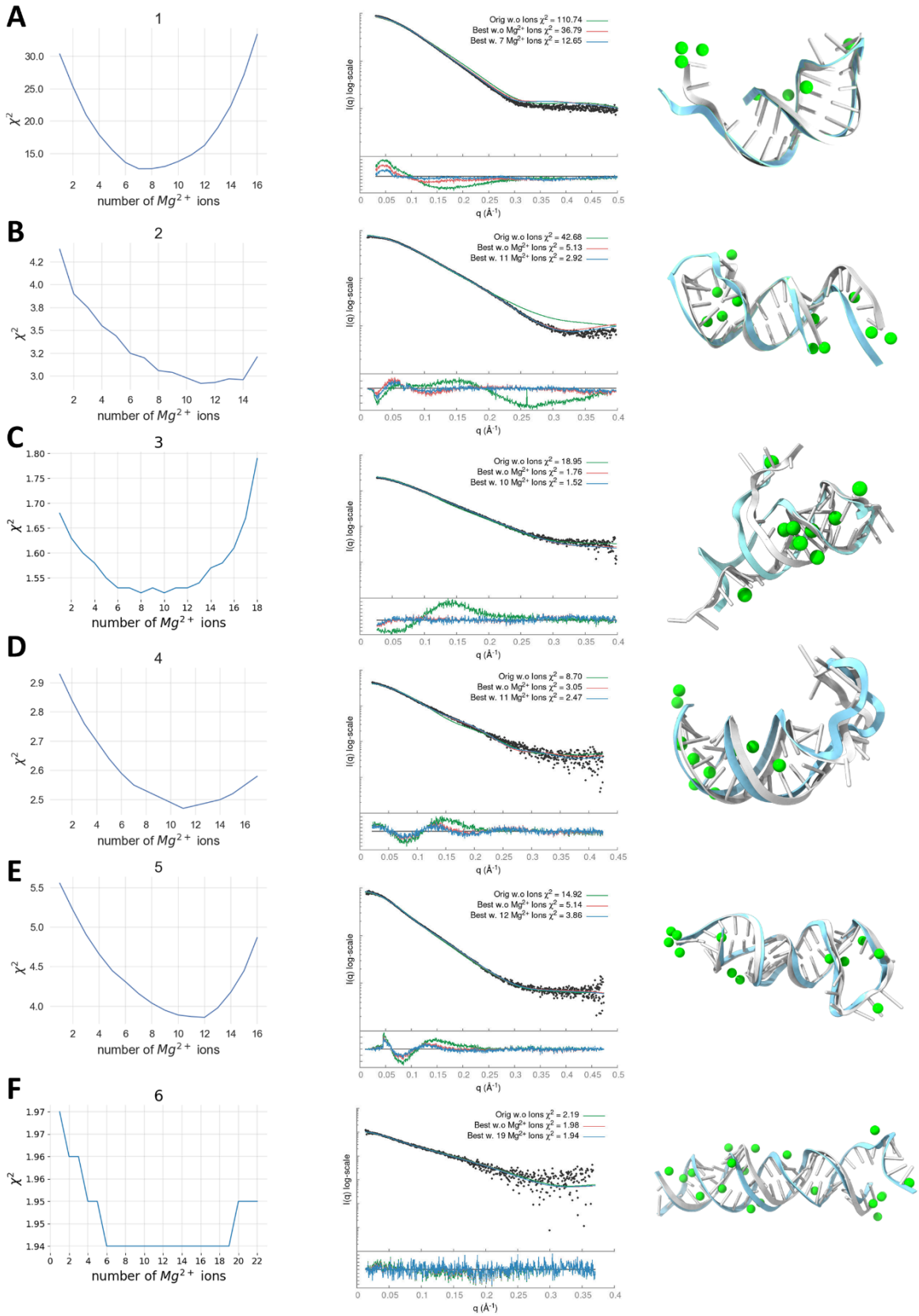
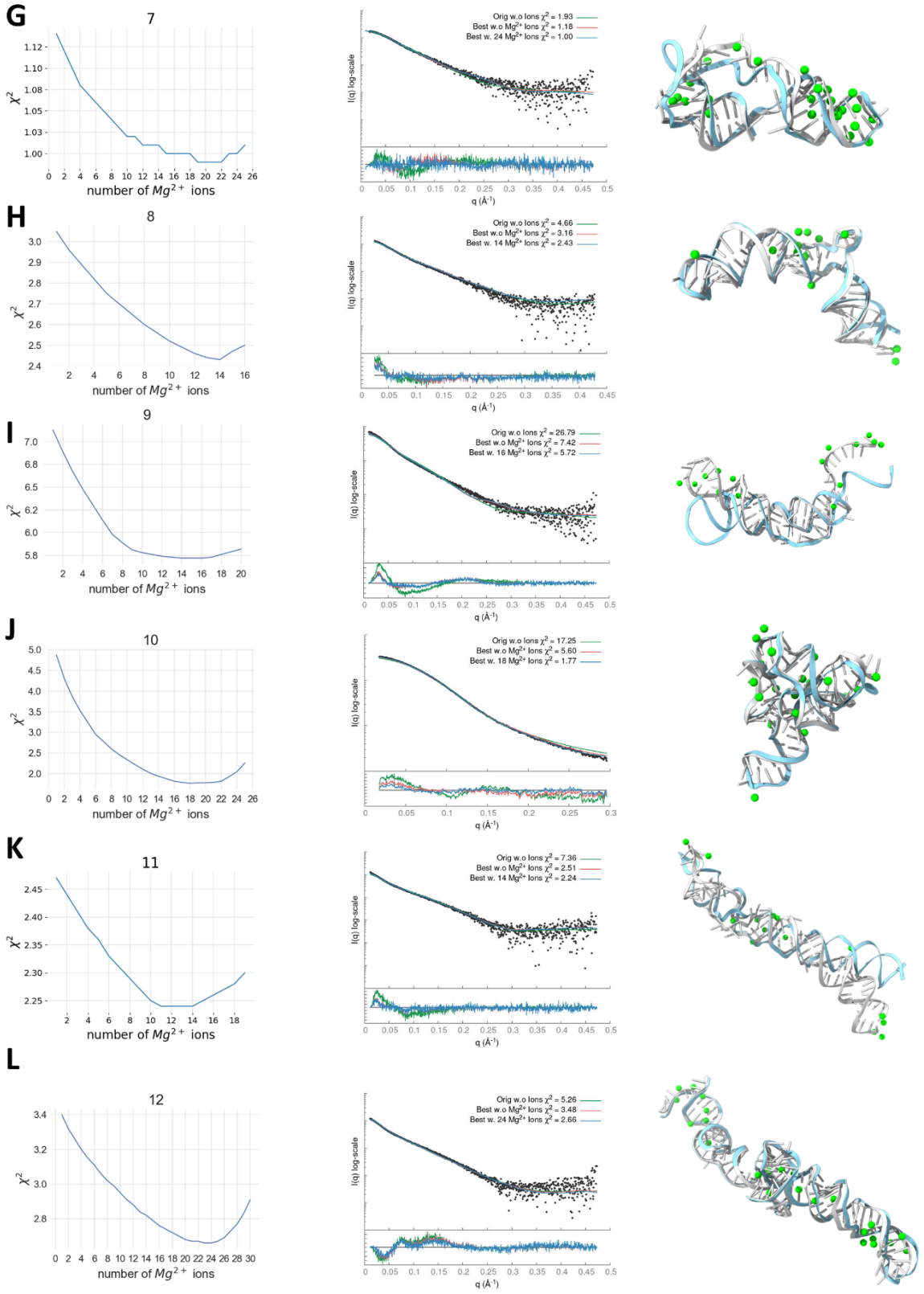


Figure 3. Impact of ionic strength on the hydration layer, Mg^{2+} placement, and RNA conformation
A-B. P(r) functions of two RNAs (#3- RNA stem-loop, #13- P4P6) calculated from SAXS curves (see Figure S4) measured at three different salt concentrations, 50 mM (blue), 150 mM (green), and 300 mM (pink). **C.** Experimental SAXS curves collected at 300mM salt concentration for RNA #3 (stem-loop, black) and #13 (P4P6, gray). SAXS fit of the best-fit structure with Mg^{2+} (red) and two-state model without Mg^{2+} (cyan) reveal similar goodness of fit. The c_1 and c_2 parameters were adjusted in both fitting approaches (see Table 1). Bottom panel: SAXS fit residual with its χ^2 values. **D.** Best-fit conformer with Mg^{2+} and 2-state model shown for #3- RNA stem-loop, #13- P4P6 with corresponding weights. The χ^2 and c_2 values for all salt concentrations are in Table 1.


A
1
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 110.74
Best w.o Mg^{2+} Ions χ^2 = 36.79
Best w. 7 Mg^{2+} Ions χ^2 = 12.65
B
2
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 42.68
Best w.o Mg^{2+} Ions χ^2 = 5.13
Best w. 11 Mg^{2+} Ions χ^2 = 2.92
C
3
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 18.95
Best w.o Mg^{2+} Ions χ^2 = 1.76
Best w. 10 Mg^{2+} Ions χ^2 = 1.52
D
4
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 8.70
Best w.o Mg^{2+} Ions χ^2 = 3.05
Best w. 11 Mg^{2+} Ions χ^2 = 2.47
E
5
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 14.92
Best w.o Mg^{2+} Ions χ^2 = 5.14
Best w. 12 Mg^{2+} Ions χ^2 = 3.86
F
6
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 2.19
Best w.o Mg^{2+} Ions χ^2 = 1.98
Best w. 19 Mg^{2+} Ions χ^2 = 1.94
I(q) log-scale
q (Å^{-1})


G
7
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 1.93
Best w.o Mg^{2+} Ions χ^2 = 1.18
Best w. 24 Mg^{2+} Ions χ^2 = 1.00
H
8
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 4.66
Best w.o Mg^{2+} Ions χ^2 = 3.16
Best w. 14 Mg^{2+} Ions χ^2 = 2.43
I
9
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 26.79
Best w.o Mg^{2+} Ions χ^2 = 7.42
Best w. 16 Mg^{2+} Ions χ^2 = 5.72
J
10
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 17.25
Best w.o Mg^{2+} Ions χ^2 = 5.60
Best w. 18 Mg^{2+} Ions χ^2 = 1.77
K
11
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 7.36
Best w.o Mg^{2+} Ions χ^2 = 2.51
Best w. 14 Mg^{2+} Ions χ^2 = 2.24
L
12
number of Mg^{2+} ions
Orig w.o Ions χ^2 = 5.26
Best w.o Mg^{2+} Ions χ^2 = 3.48
Best w. 24 Mg^{2+} Ions χ^2 = 2.66
I(q) log-scale
q (Å^{-1})

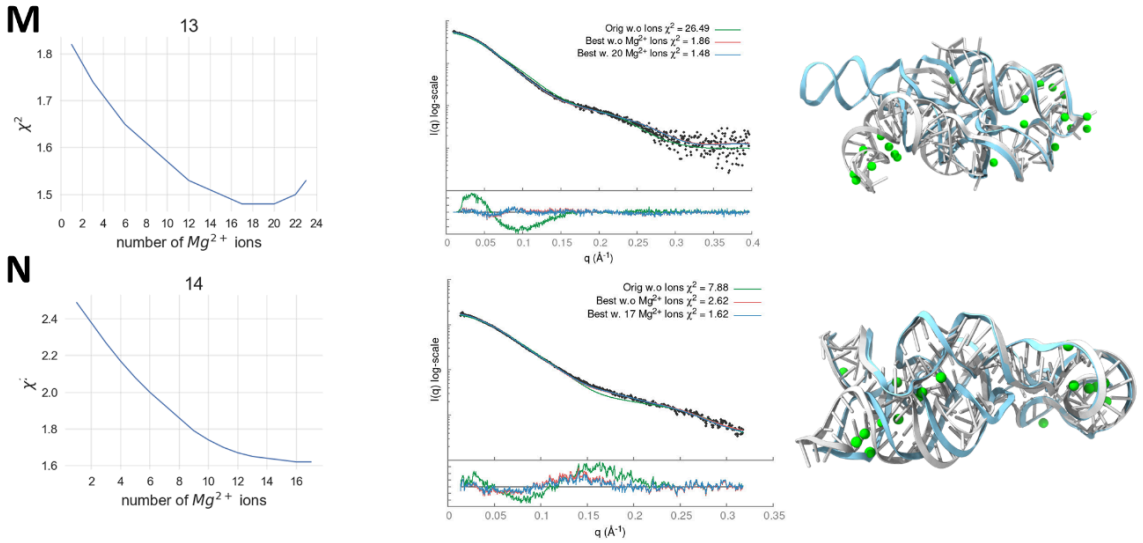


Figure 4. SCOPER results for 14 benchmark cases (labeled A to N). The decrease in χ^2 per Mg^{2+} ion added (left column plots), SAXS profile fit of original starting structure (green), best structure without Mg^{2+} ions (red), and best structure with added Mg^{2+} ions (blue) vs. experimental profile (black dots) (middle column plots), initial RNA structure (cyan) and best fitting RNA structure (white) with predicted Mg^{2+} ions (green) (right column).

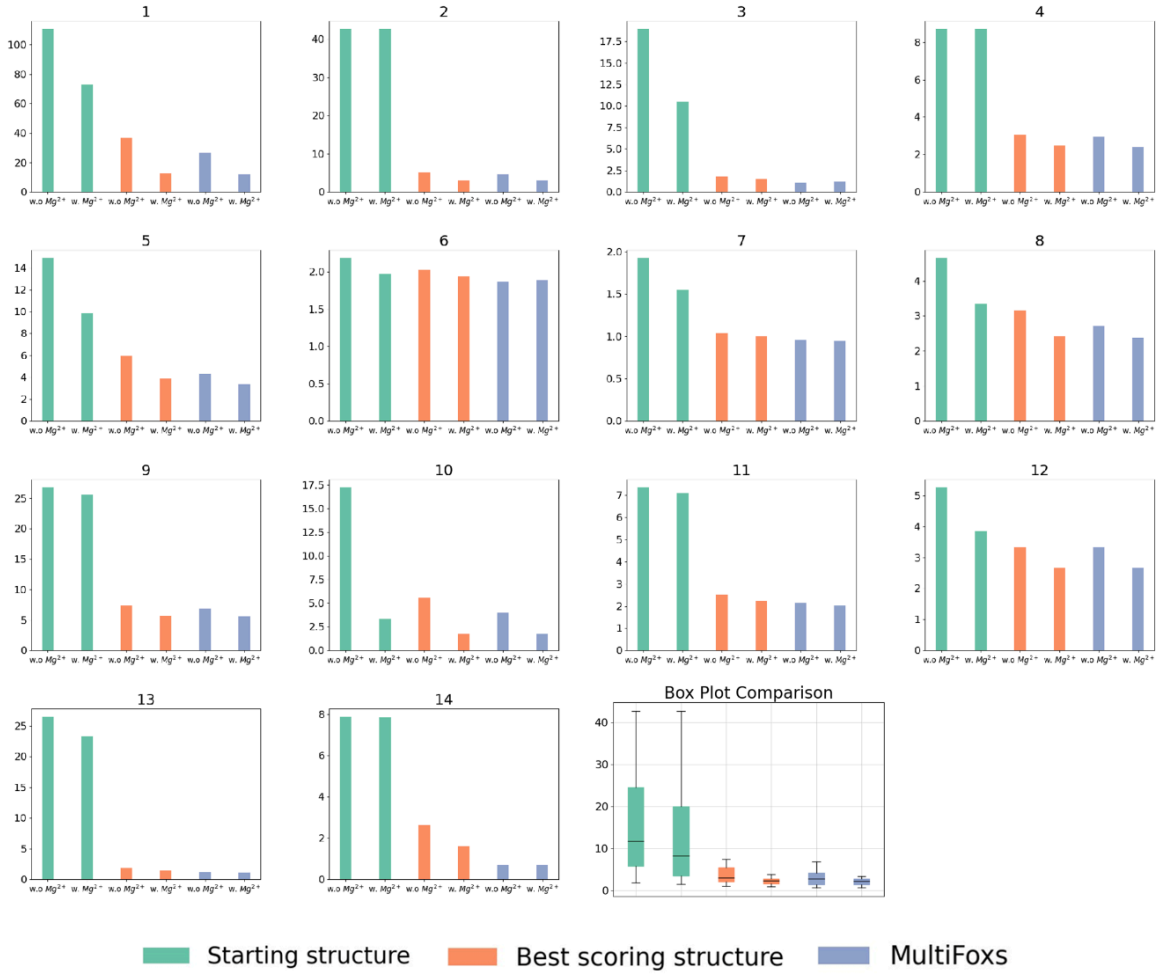


Figure 5. Bar plots of χ^2 scores for the 14 benchmark cases: starting structures (green), best scoring conformations (orange), and multi-state models (purple), all with and without the addition of Mg^{2+} ions. A summary of the scores for all benchmark cases is provided in the last boxplot.

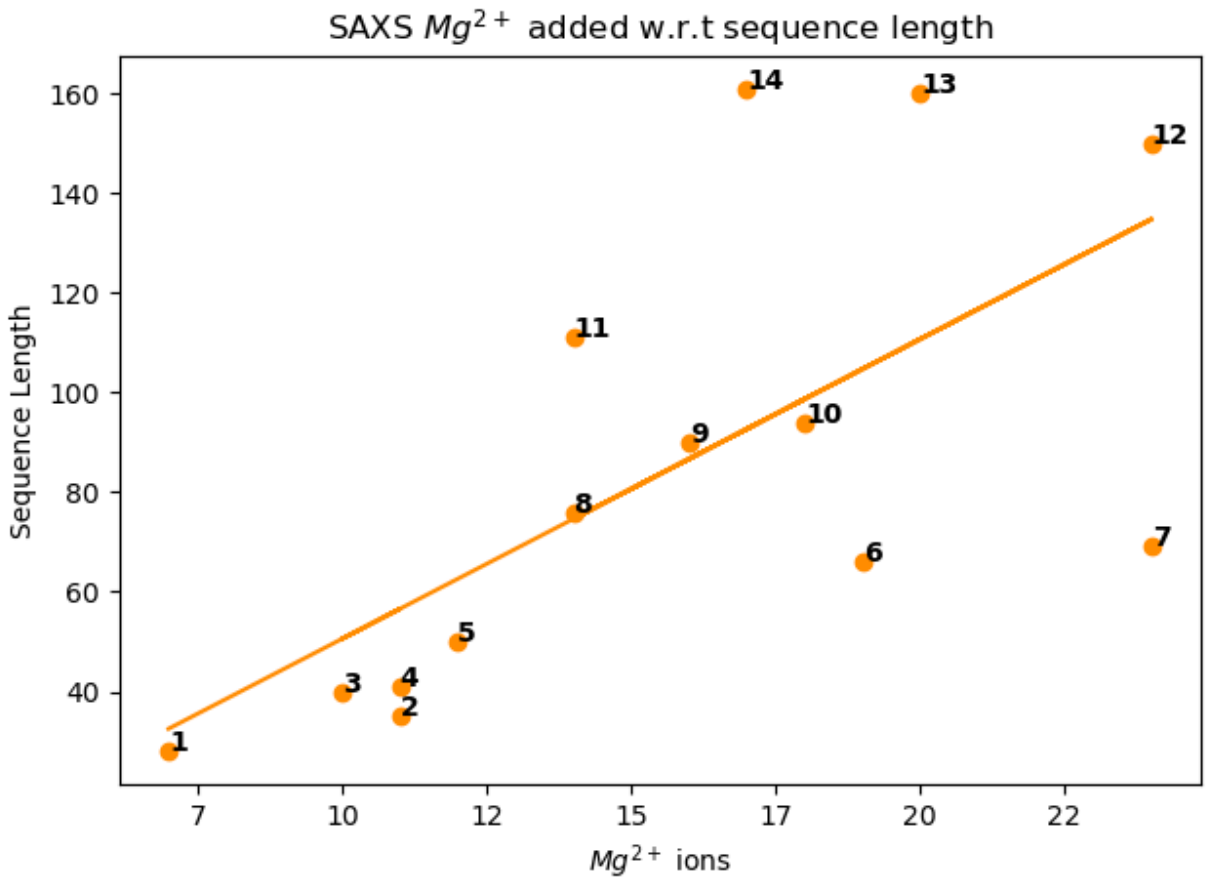


Figure 6. Number of Mg^{2+} ions (x-axis) vs. the sequence length (y-axis). The correlation coefficient is 0.668

Figure 7. An example of the SCOPER job results page.