Published in final edited form as:

Methods. 2017 January 15; 113: 46–55. doi:10.1016/j.ymeth.2016.10.008.

Analysis of RNA structure using small-angle X-ray scattering

William A. Cantara, Erik D. Olson, and Karin Musier-Forsyth*

Department of Chemistry and Biochemistry, Center for Retrovirus Research, and Center for RNA Biology, The Ohio State University, Columbus OH 43210

Abstract

In addition to their role in correctly attaching specific amino acids to cognate tRNAs, aminoacyl-tRNA synthetases (aaRS) have been found to possess many alternative functions and often bind to and act on other nucleic acids. In contrast to the well-defined 3D structure of tRNA, the structures of many of the other RNAs recognized by aaRSs have not been solved. Despite advances in the use of X-ray crystallography (XRC), nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (cryo-EM) for structural characterization of biomolecules, significant challenges to solving RNA structures still exist. Recently, small-angle X-ray scattering (SAXS) has been increasingly employed to characterize the 3D structures of RNAs and RNA-protein complexes. SAXS is capable of providing low-resolution tertiary structure information under physiological conditions and with less intensive sample preparation and data analysis requirements than XRC, NMR and cryo-EM. In this article, we describe best practices involved in the process of RNA and RNA-protein sample preparation, SAXS data collection, data analysis, and structural model building.

Keywords

SAXS; RNA structure; aminoacyl-tRNA synthetase; size-exclusion chromatography; tRNA-like structures; native PAGE

1. Introduction

Structural features defining tRNA and amino acid specificity of aminoacyl-tRNA synthetases (aaRS) have been studied for decades [1]. More recent work has demonstrated that many aaRS have alternative functions beyond their canonical role in protein synthesis [2–4]. One of the earliest noncanonical functions of an aaRS that was discovered was the role of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18) in group I intron splicing [5]. It was later revealed that CYT-18 and other aaRS that interact with group I introns do so by binding to a tRNA-like structure in the intron core [6,7]. AaRS also recognize elements within the untranslated regions (UTR) of viral RNAs, which mimic the

*To whom correspondence should be addressed. Tel: +1 614 292 2021; Fax: +1 614 688 5402; musier@chemistry.ohio-state.edu (K. Musier-Forsyth).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

L-shaped conformation typical of tRNAs [8]; a subset of these elements are even aminoacylated by the synthetase [9–15]. These tRNA-like structures (TLS) are crucial for efficient viral replication, playing roles in regulating translation vs. transcription of the viral genome, 3' end maintenance and genome packaging in plant viruses [16,17], and localization of the reverse transcription primer in HIV-1 [8,18–20]. Similar elements have been identified at the 3' ends of MALAT1 and MEN β long noncoding RNAs (lncRNA). Both elements are processed by the canonical tRNA-processing enzymes RNaseP, RNaseZ and CCA-adding enzyme, and are predicted to adopt a tRNA-like secondary structure [21–23]. The MEN β tRNA-like small RNA is targeted for degradation, but MALAT1-associated small cytoplasmic RNA (mascRNA) is retained at higher levels through differential regulation by CCA-adding enzyme [24]. While aminoacylation of mascRNA has not been detected and its cellular function remains unknown, it is well-conserved in vertebrates [22], suggesting evolutionary pressure to preserve this structure. We speculate that if its overall global structure mimics an L-shaped tRNA fold, it may play a role in the non-canonical function of an aaRS or other tRNA-binding protein.

Structural characterization of tRNA-like elements have relied heavily on small-angle X-ray scattering (SAXS) experiments [25–27] due in part to difficulties associated with high-resolution structure determination of large RNAs by X-ray crystallography (XRC), nuclear magnetic resonance spectroscopy (NMR) and cryo-electron microscopy (cryo-EM). Crystallography of RNA is complicated by several factors including poor packing due to flexibility or electrostatic repulsion of the polyanionic backbone [28]. These difficulties result in a general trend in which larger RNAs tend to yield poorer resolution [27]. Similarly, obstacles must be overcome in order to solve structures of large RNAs by NMR, including spectral overlap from poor chemical shift dispersion due to chemical structure similarity among the four canonical nucleotides. In addition, large non-globular RNAs generally have slow tumbling times that result in signal loss [27]. Unlike NMR, which is generally confined to relatively small RNAs (50 nt or ~17 kDa), high-resolution cryo-EM is currently limited to molecules larger than 100 kDa [29]. XRC, NMR and cryo-EM all suffer from the requirement of expensive instrumentation and operators with highly-specialized expertise.

SAXS has seen a recent rise in popularity due to these complications with crystallographic, NMR and cryo-EM studies of RNA and RNA-protein complexes, and specific advantages inherent to the methodology [30]. SAXS is a solution technique that can be performed at much lower concentrations than either NMR or XRC, and does not require the sometimes arduous task of crystallization, which can require screening of thousands of conditions. SAXS can also be performed in biologically relevant solvent conditions with applicability in a broad range of both mono- and divalent salt conditions, and has no inherent size limitation. However, in our experience SAXS analysis of RNAs less than ~300 nucleotides (nt) is more reliable. Additionally, SAXS requires a relatively small amount of sample at lower concentrations than NMR or XRC. Most beamlines suggest 10- to 30- μ L samples at concentrations of 0.5 – 5.0 mg/mL. Due to the higher electron density of RNA, the lower part of the concentration range is sufficient for most practical applications. Moreover, available technologies at synchrotron sources, such as more sensitive detectors and automated liquid-handling robotics, are constantly improving to allow for analysis at much lower concentrations and volumes. As for the other techniques, specialized instrumentation

is required, but unlike NMR and cryo-EM, data collection services provided by synchrotron facilities are available at no cost. Also, data analysis is relatively straightforward. Despite these advantages, care must be taken during sample preparation and analysis to ensure the reliability of the resulting data. This review will focus on the use of SAXS for studying RNAs and RNA-protein complexes including a discussion of methods to identify potential problems in the sample or data, and ways they can be alleviated.

2. Theory and applications

2.1 Principles for construct design and sample preparation

Beyond focusing on a functional RNA of biological interest, there are several things to consider in selecting a particular RNA sequence for SAXS analysis. Investigating elements of a much larger RNA, such as a tRNA-like structure within the 6.3 kb genomic RNA of Turnip yellow mosaic virus [31], necessitates truncation of the RNA down to a much smaller functional construct. It is important to take available secondary structural information into account when designing RNA constructs. For example, it is beneficial to have a stable helix formed at the 5'/3' terminus of the RNA, as this can reduce conformational heterogeneity and improve stability of the RNA. If this is not possible, then additional non-native 5' G and 3' C residues can be added to stabilize a terminal helix. There are a number of available experimental techniques to probe the secondary structures of RNAs. Some common techniques currently employed include: selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), dimethyl sulfate (DMS), hydroxyl radical, and RNase digestion probing, as well as others (reviewed in [32]). A basic knowledge of secondary structure is a practical necessity for the generation of SAXS-based all-atom models of your RNA (see section 2.6.2). If this information is lacking, then it is worth considering first undertaking a secondary structure determination study before carrying out SAXS analysis.

While experimental data is preferred, additional methods exist that predict RNA secondary structure. The most popular is Mfold [33], a thermodynamic folding algorithm based on nearest neighbor analysis of RNA base pairs. Mfold can be useful in conjunction with experimental secondary structure data as a means to assess alternative conformations that the RNA may adopt. Co-variation analysis is another method capable of predicting RNA secondary structure by analyzing multiple sequence alignments of your RNA. When sets of nucleotides covary in a manner that maintains a base pair, predictions can be made about these nucleotides interacting in the secondary structure [34,35]. Sequence alignments can also be useful in determining regions of RNA to focus on in the absence of other information. For example, RNA elements that are poorly conserved may be dispensable for biological function and structural study, and this information can be used when designing constructs. In removing functionally dispensable RNA segments, efforts should be taken to ensure the global structure is not disturbed. For example, if an RNA domain protruding from a helix can be removed, then it should be replaced by a stable tetraloop to ensure that the helix structure is maintained [8,36]. Removal of seemingly non-functional sequences should be carried out with great caution and only if absolutely necessary since such deletions may lead to confusion in understanding structure/function relationships. Thus, all constructs should be biochemically validated for functionality and attempts should be made to analyze

the most genuine version of the RNA possible. In general, studies of core functional elements in the context of a larger, native RNA are possible, highlighting a potential advantage of SAXS over other structural techniques.

2.2 Methods of template preparation and transcription

2.2.1 PCR amplification—PCR amplification is one approach to generating template DNA for *in vitro* transcription. If available, a plasmid encoding the RNA sequence is prepared, and then forward and reverse primers are designed so that only the desired construct sequence is amplified. The need for a plasmid containing the sequence can be circumvented for short transcripts (< 100 nt), however, as DNA oligonucleotides can be purchased that serve as the template sequence, and primers then used to amplify from that duplex. The forward primer should additionally contain the T7 RNA polymerase promoter in front of the 5' template sequence. If the 5' nucleotide of the RNA sequence is not a guanosine, then additional G residues (2–3) should be added after the T7 promoter to facilitate efficient transcription activity [37,38]. It should be noted that these additional residues will be transcribed as part of the RNA product. While not required, it is beneficial to obtain reverse primers where the final 2 5' nucleotides have been modified with C2'-methoxy groups, as this has been shown to reduce non-templated nucleotide (N+1) addition at the 3' end of the transcript by T7 RNA polymerase [39].

A benefit of the PCR amplification strategy for generating template DNA is that it takes less time to implement than other approaches that will be discussed. None of the time intensive steps associated with cloning (i.e., growth of bacterial cultures, digestions, sequencing, etc.) are required. Only small amounts of plasmid and primers are required to amplify large amounts of transcription-ready template. A drawback to this approach is that large amounts of DNA polymerase are needed for the PCR used to generate template, and the quantity of template produced will not be as great as that of plasmids made in *E. coli*. Also, whereas plasmid-based approaches have a very low error rate and allow for sequencing, PCR polymerases may rarely introduce mutations into the template. High-fidelity enzymes and buffers should be used to minimize errors during PCR. Furthermore, more complex additions to the RNA transcript, such as 5' - or 3'-end ribozymes discussed below, are not trivial to include.

2.2.2 Cloning/linearized plasmid—A second approach to generating template for *in vitro* transcription is through the use of linearized plasmids. Here, the T7 promoter, any additional 5'-G residues, and a 3'-end restriction enzyme recognition site are cloned into the plasmid so that the linearized plasmid can serve as the template. Using FokI endonuclease is a good choice for the 3' terminus restriction enzyme due to its ability to cut a specific number of nucleotides upstream of its recognition motif, allowing for the generation of a completely native 3' template end.

Benefits of this approach include the ability to generate large amounts of transcription template in *E. coli*, and to include more extensive modifications to the RNA. For example, in cases where precise 5' and 3' termini are required, self-cleaving ribozymes can be incorporated at either end of the transcript, which upon cleavage will generate homogenous

termini. This approach avoids any necessary addition of G residues to the 5' end or any aberrant N+1 addition, as they occur external to the eventually excised ribozymes. The most commonly used 5' and 3' end-cleaving ribozymes are the hammerhead ribozyme and hepatitis delta virus ribozyme, respectively [40]. A drawback to the use of linearized plasmids as templates is that it is more time consuming to perform the multiple cloning steps necessary to generate them. A method of potentially overcoming the need for ribozymes to produce native, homogenous 5' and 3' ends would be to use an alternative polymerase. While not yet commercially available, the polymerase from the marine cyanophage Syn5 may overcome issues associated with T7 polymerase. It has recently been shown to have high processivity, high salt tolerance, significantly reduced N+1 addition and tolerance for A or C as the initiation nucleotide [41].

2.2.3 Transcription and purification—T7 RNA polymerase is the most commonly used enzyme for *in vitro* transcription [38]. Purified T7 is available for purchase (Thermo Fisher Scientific), and can also be easily purified from *E. coli* [42]. T7 is a robust polymerase capable of generating RNA transcripts hundreds of nucleotides in length, with its only requirement to initiate efficient transcription being the T7 promoter sequence followed by preferably 2–3 G residues [37,38]. While T7 is generally a processive polymerase, abortive transcription can occur [38], and so further purification of the generated RNAs is required. Denaturing (urea) polyacrylamide gel electrophoresis (PAGE) is the most common purification approach, and is capable of isolating the RNA of interest from abortive transcripts or cleaved ribozyme fragments. After RNA isolation, it is advisable to analyze the sample on an analytical scale denaturing PAGE to ensure that a single, homogenous RNA species has been achieved. Any significant length heterogeneity in the SAXS-analyzed RNA population will lead to convoluted data and create significant difficulties in successful structural interpretation.

2.3 RNA folding and sample preparation

2.3.1 Assessing RNA homogeneity—In addition to length homogeneity, RNA structure determination requires conformational and oligomeric homogeneity for successful SAXS analysis. Therefore, the RNA conformational and oligomeric state should be assessed prior to SAXS data acquisition. The two main tools employed are native PAGE and size exclusion chromatography (SEC). Native gels lack RNA denaturing agents (such as urea) and may contain Mg^{2+} to maintain native RNA folds [43]. Native-PAGE analysis is useful as an initial assessment of RNA conformational homogeneity and can be used as a means to screen different RNA folding conditions (see section 2.3.2). Benefits of native-PAGE are its simplicity and high throughput nature. However, while overall homogeneity can easily be assessed, without well-established controls such as known monomeric RNA mutants in the case of an RNA that can dimerize, or RNA standards such as a well-folded L-shaped tRNA, it can be difficult to definitively assign the identity of RNA species observed on the gel. Another drawback is that the gel environment during electrophoresis is less physiologically relevant and conformations or oligomeric states may be disrupted.

SEC is a lower throughput method than native-PAGE but possesses distinct advantages. Like native-PAGE, SEC is capable of distinguishing RNA oligomeric and conformational states,

but unlike native-PAGE the sample can be run in buffers closely resembling physiological conditions. SEC is also capable of natively purifying large quantities of RNA, an important step prior to SAXS submission. The choice of SEC column depends on the size of the RNA and whether there is a need to resolve similarly eluting conformers. Ideally, columns should provide optimal resolution near the retention time of the chosen construct. In some cases, optimal choice of column can allow for resolution of slightly misfolded species [44]. As for native PAGE, it is difficult to definitively identify RNA species using SEC alone. However, SEC can be coupled to a multi-angle laser light scattering (MALS) instrument to overcome this limitation [45]. MALS is capable of measuring the molecular weight of a biomolecule, thus, SEC-MALS is able to definitively assign the oligomeric state of the RNA eluting in each peak of the SEC chromatogram [46]. Multiple RNA species are often not completely resolved and MALS can be useful in identifying if the convoluted peaks are the result of different RNA conformers or due to different oligomeric states. Additionally, peaks can be fit to Gaussian approximations so that population estimates can be assigned to each peak/species present and better ensure that only one species is being selected.

It should be noted that samples may contain more than a single species, resulting from either mass or conformational differences, even if a single peak is observed in the chromatogram. For example, an asymmetric peak can be indicative of polydispersity. In such cases, MALS may be helpful in determining whether this is due to species with different masses or conformations [47]. For conformational differences, the resulting SAXS data may still be used for ensemble modeling, but mass differences will require more careful purification or optimization of preparation protocols. A recently-described use of evolving factor analysis in preparation of SAXS samples by SEC may be useful for differentiating complex elution profiles [48].

2.3.2 Optimizing RNA folding—Using the tools described above, it is important to next optimize RNA folding and buffer conditions to achieve the most homogenous sample possible prior to SAXS analysis. RNA folding is typically achieved by heating a purified RNA to high temperatures (80–90 °C) and then slowly cooling it concomitant with the addition of Mg^{2+} . However, different cooling approaches can be employed and optimization may be required to obtain a single, well-folded species as assessed by native PAGE. For example, step-wise cooling to 60 °C and then addition of Mg^{2+} to a final concentration of 1 mM followed by incubation at 37 °C and storage on ice for at least 30 minutes is a protocol we generally follow for folding tRNA [8]. Alternatively, thermocycler instruments can be programed to yield a slow, controlled temperature decrease. The length of time the sample spends at 37°C or on ice, the velocity of the temperature ramp on the thermocycler, and the concentration of Mg^{2+} added are all aspects of RNA folding that can be optimized and assessed to determine a protocol that yields the most homogenous sample. While Mg^{2+} ions are typically used to reconstitute native RNA structure, K^+ ions are also capable of this [49], and so the replacement of Mg^{2+} with K^+ ions or their use in combination is another factor that can be optimized. Mutation of the RNA may also be necessary to yield a homogenous product. Point mutations can be introduced into known multimerization sites to stabilize a single oligomeric form of the RNA [8]. If multiple oligomers are observed despite the absence of known multimerization motifs, the RNA can be inspected for palindromic

sequences (http://www.biophp.org/minitools/find_palindromes/demo.php) and mutations introduced at these sites. A benefit of SAXS is that the measurement can be carried out in buffers closely resembling physiological conditions. A typical SAXS sample buffer we have used contains 150 mM NaCl, 1 mM MgCl₂, 50 mM HEPES, pH 7.4, and 3% glycerol (to reduce radiation damage to the sample from the X-ray beam).

2.3.3 Preparing RNA complexes—RNA-RNA or RNA-protein complexes require additional optimization steps to reduce sample polydispersity. An important starting point is to optimize binding conditions, such as buffer conditions, temperature and time of incubation, using a binding assay of choice (e.g., electrophoretic mobility shift assay [50,51] or fluorescence anisotropy [52,53]). When preparing bimolecular complexes it is useful to add the smaller of the two components in excess so that during purification by SEC, the larger component will be saturated, making it easier to resolve the complex peak. After purification, it is best to keep the sample cold, but not frozen, to reduce the possibility of the sample dissociating into different components or forming oligomers. It is also good practice to test the complex over time to determine how long the sample can be stored at 4 °C before it shows signs of equilibration to an undesired state. SAXS analysis should only be performed on samples during this window of time.

2.3.4 Production-scale preparation of samples for SAXS analysis—As mentioned previously, all RNA samples or RNA-protein complexes should be subjected to native SEC purification prior to submission for SAXS analysis. This serves the dual purpose of ensuring sample homogeneity, as well as allowing for the sample to be exchanged into the buffer of choice. Since both the biomolecules and the buffer will scatter X-rays, it is critical that a SAXS profile of the exact buffer match is also collected so that it can be subtracted from the sample-containing SAXS data. Imperfectly matched buffers are a common cause of poor SAXS data quality; therefore, the exact buffer preparation that is used for SEC purification must be analyzed to perform the buffer subtraction. Fractions from the peak corresponding to the RNA of interest are collected and may be run on native gels to determine the purest fractions to pool and concentrate. The choice of Amicon centrifugal concentrator unit molecular weight cutoff (MWCO) size should reflect the molecular weight of the RNA or complex being purified. However, given that RNAs tend to be more extended and less globular on average than proteins of similar size, it is advisable to select a MWCO size that is at least 5–10 kDa smaller than the RNA molecular mass. Following concentration, native PAGE analysis may be carried out to ensure that no aggregation has occurred. RNA samples should be reduced to a final concentration between 1–3 mg/mL in a volume of 80–100 µL. A sufficient final concentration/volume is typically achieved by starting with ~500 µg of total RNA. Once a 1–3 mg/mL final concentration has been achieved, the sample should be serially diluted, yielding 3 final samples with concentrations ranging from 0.25–3 mg/mL. The minimum volume required will depend on the beamline's flow cell size.

2.4 SAXS data acquisition

While benchtop SAXS setups are available, they lack the intensity of high-energy synchrotron beamlines, rendering them less commonly used for biomolecules, and they will

not be discussed herein. There are several synchrotron beamline sources in the United States that are equipped to collect SAXS data on biological samples (Table 1). Beamline sources produce higher quality SAXS data due to their high-energy radiation producing more intense scattering profiles. Every beamline is different and prior to collecting data, it is advisable to research their requirements and contact the staff with any questions.

Most beamlines require users to apply for beamtime and visit in person to collect SAXS data. In this case, beamline support staff are available to help users familiarize themselves with the equipment, data collection, and raw scattering data processing. Most beamlines possess liquid-handling robotics to transfer the samples from the plate to the sample cell. Due to smaller beam profiles resulting in very high radiation doses to a small volume, some beamlines employ a flow cell where the solution is controllably flowed into the beam path. Oscillation of the solution is recommended during X-ray exposure at such beamlines to reduce radiation damage. Many beamlines also employ high frame-rate detectors such that data can be collected in the range of 100 milliseconds to 1 second, further reducing the probability of radiation damage to the sample. Cameras are typically directed at the sample so that the solution can be monitored during data acquisition, as the presence of bubbles or visible precipitates can explain noisy SAXS data. A benefit of collecting SAXS data in person is that you have complete control over your experiment and can troubleshoot any problems that arise in real time.

The SIBYLS beamline at the Advanced Light Source provides “mail-in” services where samples are analyzed by beamline staff. In this approach, all samples are prepared in your lab, aliquoted into 96-well plates (Axygen Scientific), and shipped to the beamline. The mail-in approach has several benefits: travel is not required, and data are collected, averaged, and buffer subtracted by beamline staff. Comments on the quality of the SAXS data (if aggregation was present, if the data appear noisy, etc.) are provided, giving the user an indication of potentially poor quality data.

An alternative approach is SEC-SAXS where, rather than performing SEC purification and concentration of the samples in the days prior to SAXS analysis, the SEC is linked to the flow cell so that RNA species are analyzed directly as they come off of the column. This technique is advantageous for samples where multiple species can be separated by SEC but quickly re-equilibrate after isolation; SEC-SAXS allows for characterization of each species separately, and in a single run. For some RNAs and RNA-protein complexes, this may be a practical necessity in order to collect SAXS data on distinct oligomeric states or conformations. Downsides are that it may take longer to analyze a single sample depending on the length of the SEC run and more material is required due to inherent dilution of the sample during SEC purification. The amount of dilution will depend on the selected column as well as the fraction of the sample that corresponds to the peak of interest. For example, if the SEC results in a 3-fold dilution and the peak of interest constitutes one-half of the total peak area of the chromatogram, a 50 μ L injected sample would need to be at 6 mg/mL for a total of 300 μ g to result in a final concentration of 1 mg/mL for analysis. Users are encouraged to investigate these parameters in their own labs to improve efficiency at the beamline. SEC-SAXS also negates the need for a concentration series since different

concentrations will be analyzed at different points along the eluted peak, which may alleviate some time considerations.

2.5 Analysis of SAXS data

2.5.1 Determination of sample quality—For applications in which the sample is not analyzed via SEC-SAXS, a finite time will elapse between sample preparation and analysis. Despite samples being kept cold, constitutional or conformational heterogeneity may exist in the samples. Collecting SAXS data on a homogenous sample is important to ensure the most reliable data analysis. For this reason, it is recommended that an aliquot of the sample be analyzed by native PAGE at the same time that SAXS data is being collected. Multiple or diffuse bands are indications of poor monodispersity, and data resulting from such samples will be unreliable for structure determination.

Although attention to sample preparation and examination of sample homogeneity at the time of data collection are important factors for reducing the likelihood of poor data quality, all SAXS data should be carefully scrutinized prior to analysis. A 2-dimensional plot of the intensity of scattered X-rays (I) as a function of the modulus of the scattering vector or momentum transfer (q , which is equal to $4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle and λ is the wavelength of the incident X-rays), can indicate potential issues arising from sample aggregation, interparticle interactions/repulsions or radiation damage. Specifically, the low- q^2 region of a Guinier plot ($\ln I$ vs q^2) is linear in a well-behaved system, whereas an upward curve that increases with increasing concentration or exposure time as q^2 approaches zero is a telltale indicator of aggregation or radiation damage (Fig. 1A). A similar upward curve may be caused by beamstop noise, but this will be independent of concentration or exposure time. Additionally, interparticle repulsion is suggested by a dropoff in the intensity of the raw SAXS curve ($\log I$ vs q) as q approaches zero (Fig. 1B). Interparticle affinity, for example multimerization or minor aggregation, will manifest as slight differences in the SAXS curves from different sample concentrations after normalizing their scale (Fig. 1C). These differences will also be apparent in concentration-dependent trends in the calculated radius of gyration (R_g) and extrapolated intensity at zero angle (I_0) values (described in section 2.5.3).

In addition to issues stemming from sample preparation, the X-rays can generate free radicals resulting in radiation damage to the samples [54]. Ideally, users should collect data at different exposure times, where radiation damage is indicated by differences in the R_g and I_0 values as a function of exposure time. As mentioned in section 2.3.2, radical scavengers such as glycerol can be used to alleviate issues with radiation damage. Data wherein signs of aggregation, interparticle interactions/repulsions or radiation damage are observed may result in incorrect conclusions and, therefore, should not be analyzed. It is recommended that sample preparation methods are re-optimized before recollecting SAXS data.

Kratky analysis of the data ($q^2 \cdot I(q)$ vs q) can also reveal information regarding the “foldedness” of the molecule [55]. A bell-shaped curve in the Kratky plot suggests a well-folded, globular and compact conformation (Fig 2) [56]. Conversely, a Kratky plot that either forms a plateau or increases at high- q suggests a more elongated or, alternatively, a partially-folded molecule. It should be noted that while tRNAs usually result in Kratky plots

that are typical of compact, well-folded molecules, tRNA-like structures may not [25]. Similarly, aaRS-RNA complexes may not adhere to a bell-shaped Kratky plot. In this case, the data can still be analyzed and will likely result in a more elongated final model. A newer form of Kratky analysis, the dimensionless Kratky plot ($(qR_g)^2 \cdot I(q)/I(0)$ vs qR_g), results from normalization and is becoming a standard in the SAXS community [57]. Using this analysis, a spherical particle will be characterized by a peak at 1.1, and a completely unfolded molecule will have a plateau at 2. This is particularly useful for identifying aggregation or radiation damage because dimensionless Kratky plots for different concentrations and different exposures should be superimposable.

2.5.2 Truncating and merging curves—A typical SAXS experiment will involve collecting multiple datasets for a single sample. Data is generally collected for four different exposure times for each of three concentrations. As a first step, data showing signs of radiation damage, usually only seen at the longer exposure times, should be removed from analysis. The remaining curves should be combined into a single dataset for further analysis. A good rule of thumb is that the lower concentration samples and the shorter exposure times are more likely to suffer from low signal/noise in the high- q region of the curve, whereas higher concentration samples and longer exposure times are more likely to have larger interference or detector saturation in the low- q region near the beamstop. To remove these artifacts from the final dataset, curves should be truncated and aligned in the intermediate- q region prior to merging the data (Fig 3).

2.5.3 Determination of physical parameters—After initial visual inspection of SAXS curves, several important parameters can be determined directly from the data. The radius of gyration (R_g), a measure of non-sphericity that describes the mass-weighted distribution of atoms around the center of gravity [58], and the forward scattering intensity (I_0) can be estimated using a Guinier approximation, where at low angles (small q values), a linear fit of the low- q region of a $\ln[I(q)]$ vs. q^2 plot allows direct R_g and I_0 calculation [59]. Importantly, particles with the same volume but different R_g values will have different shapes [58]. Therefore, a change in R_g value for a specific molecule in response to different conditions must indicate either a conformational change (same volume, different shape) or a change in oligomeric state (different volume, different shape). I_0 , which is extrapolated from the Guinier approximation, is proportional to the molecular weight of the particle [60]. A concentration series of a well-behaved, monodisperse sample will yield the same values for I_0 normalized for concentration ($I_{0,C}$). An upward trend in $I_{0,C}$ with increasing concentration suggests an increase in molecular weight, or oligomerization.

Another informative calculation involves an indirect Fourier transform of the scattering plot to generate a pair-distance distribution function (PDDF), which is a resolution-limited, real-space transform of the SAXS intensities that results in a histogram of all inter-electron distances within a particle [61–63]. Best practices require that the PDDF should be reported to the resolution of the SAXS data that supported the transform. The PDDF can be thought of as being analogous to the Patterson function in XRC, except that it is a radial function for SAXS because the particles are in solution rather than being organized into a well-defined lattice. The PDDF is calculated using an indirect Fourier transform [61] because, for all

practical purposes, it can be assumed that the PDDF will have boundary conditions where there are no pairwise distances of zero Å and the function should converge to zero at the maximum particle dimension ($D_{\max}$). Calculation of the PDDF is automated in many analysis programs [62,63]; however, careful inspection of the resulting plot is important and may require manual adjustment. The PDDF should not be negative at any point, as this is a clear indication of a poorly calculated function. In addition, the values should smoothly approach zero near the $D_{\max}$. Calculation of the PDDF also results in a real-space R_g determination. The real-space R_g , which is calculated from the entire dataset that supported the PDDF transform, and the Guinier R_g , determined from a very small fraction of the SAXS data, should be within 10%.

In addition to calculated values of R_g , I_0 and $D_{\max}$, the recently defined V_c (the ratio of I_0 to the total scattered intensity) and Q_R (the ratio of the square of V_c to R_g) can be used to identify conformational changes and to more accurately determine particle mass [64]. It should be noted that no single parameter should be taken out of context. A combination of many different parameters, which may be specific to a particular experiment, should be considered and will help to constrain further analysis or to better determine the validity of structural models. Also, many of these parameters can be calculated using different mathematical approaches as a way of determining the quality of the data [58].

2.6 Structural modeling

2.6.1 Calculation of *ab initio* models—Two independent methods of generating SAXS-derived structural models can be applied. First, low-resolution *ab initio* models (or molecular envelopes) of the overall molecular shape can be generated directly from the PDDF (Fig. 4A). Programs have been developed for generating *ab initio* models, the most popular of which are DAMMIN [65] and a faster version, DAMMIF [66]. These programs utilize dummy atoms (also known as beads) with an initial state of being tightly packed into a sphere with a diameter equal to the $D_{\max}$ (see section 2.5.3). These programs use a simulated annealing process in which the identity of each bead is randomly changed from a particle atom to a solvent atom, solvent atom to a particle atom, or not changed at all during each step. After each step, the model is evaluated for its fit to the experimental data. The algorithm is set up such that the data will nearly always tend toward a better fit to the data with penalties for looseness or connectivity. A similar program, MONSA, is capable of deconvoluting and generating bead models for different parts of a multicomponent sample [65,67]. This can be very helpful for RNA-protein complexes [68–70]. Also, if a divide-and-conquer experiment is designed where different domains of an RNA are analyzed individually and then together in the intact structure [27,71,72], MONSA may be capable of determining which domain corresponds to a specific region of density in the larger RNA.

A key issue with generating *ab initio* models is uniqueness of the solution. Specifically, for each PDDF, there may be many different model shapes that accurately represent the SAXS data. Therefore, it is important to generate many different models (at least 20–30) and evaluate them for consistency. For this step, we typically use the default settings in the “Jagged” mode of DAMMIN [65] because it is a relatively fast calculation, but includes more dummy atoms than faster modes. In general, a normalized spacial discrepancy (NSD, a

measure of inconsistency in *ab initio* models) of greater than 1.0 should be regarded as not converging to a unique shape due to systematic differences in the models generated [73]. This may be the result of poor quality SAXS data and can be alleviated in many cases by carefully re-optimizing the sample preparation as described in section 2.3.

Another cause of high NSD values may be an incorrect $D_{\max}$ value. In many cases, the $D_{\max}$ may be underestimated, resulting in distorted *ab initio* models [74]. It is generally not advised to simply use an unusually high $D_{\max}$ because it will alter the calculated PDDF and may cause artifacts. A better approach is to start with the originally calculated $D_{\max}$ and calculate additional PDDFs using a gradual increase in $D_{\max}$ values (typically by 1–2 Å at a time). From each of these, a small number (5–8) of *ab initio* models can be calculated, and the $D_{\max}$ at which further increases in $D_{\max}$ do not result in larger models is considered optimal. It is notable that the optimal $D_{\max}$ value will also tend to have a better fit to the experimental data (lower χ^2 value) and lower NSD.

After generating a set of molecular envelopes with an average pairwise NSD value of less than 1.0, the models are averaged and filtered. Each model contributing to the average will fit well to the experimental SAXS data; however, the filtered model may not fit well because it is simply a depiction of the beads that are most represented within the set of models. Therefore, to get a filtered envelope that more closely recapitulates the scattering data, a second round of dummy atom modeling may be applied. In this case, the averaged envelope from the first round represents the locations of all beads that are represented in the original model set, and is therefore a good starting point for the second round. Also, with a smaller initial volume, more beads can be added without significantly increasing the time required for the calculation. The number of beads that are added will depend on the available computational resources and the size of the molecule being analyzed. As a general rule, a good starting point is to use a number of beads equal to one to two times the amount present in the sphere used for the first round. After the second round, the models should converge to a single, most likely shape for the molecule of interest.

2.6.2 Generation of all-atom models—In addition to calculation of *ab initio* models, all-atom modeling of RNA molecules based on SAXS data is a complementary, independent approach (Fig. 4B). A description of all-atom modeling techniques has recently been published [75] and will only be briefly described here. The general overall scheme in this approach includes generation of all-atom models followed by evaluation of their fit to the experimental data. Due to the hierarchical nature of RNA folding and the canonical helical conformation of double-stranded regions, the secondary structure that is used will directly determine the possible tertiary folds that are modeled. A common method of 3-dimensional (3D) modeling involves generating a large number of possible tertiary models and filtering them based on their predicted fit to experimental data [8,76,77]. This can be accomplished using one of a number of *de novo* modeling programs, each with slightly different mechanics [78,79] and, thus, different limitations. For example, programs such as MC-SYM [80] and FARNA [81] have the capability of generating a large number of models using multiple secondary structures in a short period of time, but they both depend on base-pairing geometries that are represented in existing structural databases. Other methods employ a physics-based approach where *de novo* models are generated through coarse-grained or all-

atom molecular dynamics (MD) simulations. Such methods do not require information from structure databases, but require significantly more computational resources to obtain a large library of models [82]. An alternative approach is to generate and screen fewer models and perform short (5–10 ns) all-atom MD simulations on only select models that fit reasonably well to the SAXS data, allowing for a large number of slightly different conformational states to be extracted from the MD trajectory and screened against the SAXS data [8]. An advantage of this approach is that it can be performed in an iterative manner until good agreement with the SAXS data is achieved. Models selected from *de novo* predictions that fit well to the SAXS data can then be further refined using conformational sampling approaches such as simulated-annealing [8,83] or “natural” hierarchical moves [75].

A critical consideration for all-atom modeling of RNAs is the validity of defined secondary structures. Typical methods of determining RNA secondary structure, as discussed in section 2.1, involve the use of chemical or enzymatic probes that will preferentially modify or cleave the RNA at bases that are either more likely to be base-paired or unpaired [32]. These experiments may result in many different secondary structures that fit the data and are within a small range of free energy values. This possibility underscores the importance of using multiple possible secondary structures [8] as inputs for 3D modeling including possibilities for pseudoknots. It is unlikely that significantly different secondary structures will have the ability to adopt conformations that recapitulate the SAXS data, thus allowing for the most likely secondary structure to be validated in the process of generating all-atom models.

Another consideration for all-atom modeling of RNA is the ability to use complementary experimental data. For example, SAXS has been used successfully in conjunction with NMR to more accurately derive models of RNA structure [74,84–88]. Indeed, other restraints on modeling such as RNA chemical probing and Förster resonance energy transfer or other structural data would be helpful.

2.7 Validation of models

Due to the low resolution nature of the technique and the possibility for non-uniqueness in modeling, validation of RNA structures determined by SAXS is important. RNA structures are composed of helices connected by bulges, junctions or other non-helical motifs, which allows for specific manipulation of the RNA structure to produce predictable changes in the SAXS data and subsequent *ab initio* models. For example helices can be truncated or extended by insertion of base pairs. If possible, this method should be applied to multiple helices to further validate the modeling.

An alternative method for validation of all-atom models is introduction of isosteric, electron-rich nucleotide analogs into the RNA structure that will create predictable differences in the SAXS curve (W. Cantara, E. Arnold and K. Musier-Forsyth, unpublished). This is similar to isomorphous replacement in X-ray crystallography where introduction of heavy atoms can be used to determine diffraction phases. As a demonstration of this, SAXS data was obtained for the primer binding site (PBS) domain of the HIV-1 5' UTR annealed to a complementary unmodified 18-nt DNA oligonucleotide [8], as well as to a modified DNA wherein four thymines were replaced with 5-bromo-2'-deoxyuridine (BrdU). The resulting SAXS data matched the predictions based on theoretical back-calculated SAXS curves from the all-

atom model in each case (Fig. 4). In general, when more base analogs are present, larger differences between the two sets of data are expected.

Finally, as with other experimental methodologies, reproducibility should always be tested. Independent sample preparations and analyses should be performed, complete with a concentration series, full data analysis and modeling of molecular envelopes. Samples should yield reproducible SAXS data and quantifiable parameters. Generally, R_g values should be within a few angstroms of each other and scaled intensity data and PDDFs should be superimposable.

3. Conclusion and outlook

Due to distinct advantages over more high-resolution methods, SAXS has seen a recent rise in popularity. However, it is important to understand the caveats and limitations associated with SAXS and to ensure that proper optimizations and controls are carefully planned and performed. Sample preparation is critical for successful SAXS analysis, and the sample should be tested for homogeneity at every step. Similarly, the resulting data should be scrutinized using the methods described, as analysis of poor-quality samples may result in incorrect conclusions. To alleviate these issues, methods of validation should be employed whenever possible. If proper precautions are taken into account, SAXS has the unique ability to act as a stand-alone method of determining low-resolution molecular envelopes and generating reliable all-atom models, as well as acting as a tool for screening constructs or providing additional data for high-resolution methods.

Acknowledgments

We acknowledge the support of NIH grants P50 GM103368 (HIVE Center), RO1 GM065056, and GM113887. E.D.O. was supported by NIH T32 GM008512 and F31 AI120868. W.A.C. was supported by a Pelotonia Postdoctoral Fellowship from the Ohio State University.

References

1. Giegé R. The early history of tRNA recognition by aminoacyl-tRNA synthetases. *J Biosci.* 2006; 31(4):477–88. [PubMed: 17206068]
2. Guo M, Schimmel P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol.* 2013; 9(3):145–53. [PubMed: 23416400]
3. Pang YL, Poruri K, Martinis SA. tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip Rev RNA.* 2014; 5(4):461–80. [PubMed: 24706556]
4. Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol Med.* 2013; 5(3): 332–43. [PubMed: 23427196]
5. Akins RA, Lambowitz AM. A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. *Cell.* 1987; 50(3): 331–45. [PubMed: 3607872]
6. Myers CA, Kuhla B, Cusack S, Lambowitz AM. tRNA-like recognition of group I introns by a tyrosyl-tRNA synthetase. *Proc Natl Acad Sci U S A.* 2002; 99(5):2630–5. [PubMed: 11854463]
7. Caprara MG, Lehnert V, Lambowitz AM, Westhof E. A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell.* 1996; 87(6):1135–45. [PubMed: 8978617]

8. Jones CP, Cantara WA, Olson ED, Musier-Forsyth K. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proc Natl Acad Sci U S A*. 2014; 111(9):3395–400. [PubMed: 24550473]
9. Agranovsky AA, Karasev AV, Novikov VK, Lunina NA, Loginov S, Tyulkina LG. Poa semilient virus, a hordeivirus having no internal polydisperse poly(A) in the 3' non-coding region of the RNA genome. *J Gen Virol*. 1992; 73(Pt 8):2085–92. [PubMed: 1645144]
10. Beachy RN, Zaitlin M, Bruening G, Israel HW. A genetic map for the cowpea strain on TMV. *Virology*. 1976; 73(2):498–507. [PubMed: 960575]
11. Dreher TW, Goodwin JB. Transfer RNA mimicry among tymoviral genomic RNAs ranges from highly efficient to vestigial. *Nucleic Acids Res*. 1998; 26(19):4356–64. [PubMed: 9742235]
12. Felden B, Florentz C, McPherson A, Giegé R. A histidine accepting tRNA-like fold at the 3'-end of satellite tobacco mosaic virus RNA. *Nucleic Acids Res*. 1994; 22(15):2882–6. [PubMed: 8065897]
13. Goodwin JB, Dreher TW. Transfer RNA mimicry in a new group of positive-strand RNA plant viruses, the furoviruses: differential aminoacylation between the RNA components of one genome. *Virology*. 1998; 246(1):170–8. [PubMed: 9657004]
14. Koenig R, Barends S, Gulyaev AP, Lesemann DE, Vetten HJ, Loss S, Pleij CW. Nemesis ring necrosis virus: a new tymovirus with a genomic RNA having a histidylatable tobamovirus-like 3' end. *J Gen Virol*. 2005; 86(Pt 6):1827–33. [PubMed: 15914862]
15. Mans RM, Pleij CW, Bosch L. tRNA-like structures. Structure, function and evolutionary significance. *Eur J Biochem*. 1991; 201(2):303–24. [PubMed: 1935928]
16. Dreher TW. Role of tRNA-like structures in controlling plant virus replication. *Virus Res*. 2009; 139(2):217–29. [PubMed: 18638511]
17. Fechter P, Rudinger-Thirion J, Florentz C, Giegé R. Novel features in the tRNA-like world of plant viral RNAs. *Cell Mol Life Sci*. 2001; 58(11):1547–61. [PubMed: 11706983]
18. Cen S, Khorchid A, Javanbakht H, Gabor J, Stello T, Shiba K, Musier-Forsyth K, Kleiman L. Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. *J Virol*. 2001; 75(11):5043–8. [PubMed: 11333884]
19. Kleiman L, Jones CP, Musier-Forsyth K. Formation of the tRNA^{Lys} packaging complex in HIV-1. *FEBS Lett*. 2010; 584(2):359–65. [PubMed: 19914238]
20. Jones CP, Saadatmand J, Kleiman L, Musier-Forsyth K. Molecular mimicry of human tRNA^{Lys} anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. *RNA*. 2013; 19(2): 219–29. [PubMed: 23264568]
21. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res*. 2009; 19(3):347–59. [PubMed: 19106332]
22. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*. 2008; 135(5):919–32. [PubMed: 19041754]
23. Wilusz JE, Spector DL. An unexpected ending: noncanonical 3' end processing mechanisms. *RNA*. 2010; 16(2):259–66. [PubMed: 20007330]
24. Wilusz JE, Whipple JM, Phizicky EM, Sharp PA. tRNAs marked with CCACCA are targeted for degradation. *Science*. 2011; 334(6057):817–21. [PubMed: 22076379]
25. Hammond JA, Rambo RP, Filbin ME, Kieft JS. Comparison and functional implications of the 3D architectures of viral tRNA-like structures. *RNA*. 2009; 15(2):294–307. [PubMed: 19144910]
26. Hammond JA, Rambo RP, Kieft JS. Multi-domain packing in the aminoacylatable 3' end of a plant viral RNA. *J Mol Biol*. 2010; 399(3):450–63. [PubMed: 20398674]
27. Cantara WA, Olson ED, Musier-Forsyth K. Progress and outlook in structural biology of large viral RNAs. *Virus Res*. 2014; 193:24–38. [PubMed: 24956407]
28. Holbrook SR, Kim SH. RNA crystallography. *Biopolymers*. 1997; 44(1):3–21. [PubMed: 9097731]
29. Cheng Y. Single-particle cryo-EM at crystallographic resolution. *Cell*. 2015; 161(3):450–7. [PubMed: 25910205]
30. Graewert MA, Svergun DI. Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr Opin Struct Biol*. 2013; 23(5):748–54. [PubMed: 23835228]

31. Dreher TW. Turnip yellow mosaic virus: transfer RNA mimicry, chloroplasts and a C-rich genome. *Mol Plant Pathol.* 2004; 5(5):367–75. [PubMed: 20565613]
32. Weeks KM. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol.* 2010; 20(3):295–304. [PubMed: 20447823]
33. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31(13):3406–15. [PubMed: 12824337]
34. Sankoff D. Simultaneous solution of the RNA folding, alignment and protosequence problems. *Siam Journal on Applied Mathematics.* 1985; 45(5):810–825.
35. Kiryu H, Tabei Y, Kin T, Asai K. Murlet: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics.* 2007; 23(13):1588–98. [PubMed: 17459961]
36. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF. Structure of the HIV-1 RNA packaging signal. *Science.* 2015; 348(6237):917–21. [PubMed: 25999508]
37. Imburgio D, Rong M, Ma K, McAllister WT. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry.* 2000; 39(34):10419–30. [PubMed: 10956032]
38. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* 1987; 15(21):8783–98. [PubMed: 3684574]
39. Kao C, Zheng M, Rudisser S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA.* 1999; 5(9):1268–72. [PubMed: 10496227]
40. Avis JM, Conn GL, Walker SC. Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. *Methods Mol Biol.* 2012; 941:83–98. [PubMed: 23065555]
41. Zhu B, Tabor S, Richardson CC. Syn5 RNA polymerase synthesizes precise run-off RNA products. *Nucleic Acids Res.* 2014; 42(5):e33. [PubMed: 24285303]
42. Rio DC. Expression and purification of active recombinant T7 RNA polymerase from *E. coli*. *Cold Spring Harb Protoc.* 2013; 2013(11)
43. Tran T, Liu Y, Marchant J, Monti S, Seu M, Zaki J, Yang AL, Bohn J, Ramakrishnan V, Singh R, Hernandez M, Vega A, Summers MF. Conserved determinants of lentiviral genome dimerization. *Retrovirology.* 2015; 12:83. [PubMed: 26420212]
44. Kazantsev AV, Rambo RP, Karimpour S, Santalucia J Jr, Tainer JA, Pace NR. Solution structure of RNase P RNA. *RNA.* 2011; 17(6):1159–71. [PubMed: 21531920]
45. Sahin E, Roberts CJ. Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. *Methods Mol Biol.* 2012; 899:403–23. [PubMed: 22735967]
46. Foltá-Stogniew E. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods Mol Biol.* 2006; 328:97–112. [PubMed: 16785643]
47. Reyes FE, Schwartz CR, Tainer JA, Rambo RP. Methods for using new conceptual tools and parameters to assess RNA structure by small-angle X-ray scattering. *Methods Enzymol.* 2014; 549:235–63. [PubMed: 25432752]
48. Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. *J Am Chem Soc.* 2016; 138(20):6506–16. [PubMed: 27145334]
49. Draper DE. A guide to ions and RNA structure. *RNA.* 2004; 10(3):335–43. [PubMed: 14970378]
50. Gagnon KT, Maxwell ES. Electrophoretic mobility shift assay for characterizing RNA-protein interaction. *Methods Mol Biol.* 2011; 703:275–91. [PubMed: 21125497]
51. Ryder SP, Recht MI, Williamson JR. Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol Biol.* 2008; 488:99–115. [PubMed: 18982286]

52. Heyduk T, Ma Y, Tang H, Ebright RH. Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol.* 1996; 274:492–503. [PubMed: 8902827]
53. Jameson DM, Sawyer WH. Fluorescence anisotropy applied to biomolecular interactions. *Methods Enzymol.* 1995; 246:283–300. [PubMed: 7752928]
54. Kuwamoto S, Akiyama S, Fujisawa T. Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *J Synchrotron Radiat.* 2004; 11(Pt 6):462–8. [PubMed: 15496733]
55. Doniach S. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev.* 2001; 101(6):1763–78. [PubMed: 11709998]
56. Glatter, O., Kratky, O. Small-angle X-ray scattering. Academic Press; 1982.
57. Durand D, Vives C, Cannella D, Perez J, Pebay-Peyroula E, Vachette P, Fieschi F. NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol.* 2010; 169(1):45–53. [PubMed: 19723583]
58. Jacques DA, Trehwella J. Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci.* 2010; 19(4):642–57. [PubMed: 20120026]
59. Guinier, A. La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. Univ. de Paris; Paris: 1939. 3 p.l., 80, 2 p., 1 l
60. Mylonas E, Svergun DI. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J Appl Crystallogr.* 2007; 40:S245–S249.
61. Glatter O. Data evaluation in small-angle scattering - calculation of radial electron-density distribution by means of indirect fourier transformation. *Acta Physica Austriaca.* 1977; 47(1–2): 83–102.
62. Liu H, Zwart PH. Determining pair distance distribution function from SAXS data using parametric functionals. *J Struct Biol.* 2012; 180(1):226–34. [PubMed: 22659403]
63. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst.* 1992; 25:495–503.
64. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 2013; 496(7446):477–81. [PubMed: 23619693]
65. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 1999; 76(6):2879–86. [PubMed: 10354416]
66. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst.* 2009; 42:342–346. [PubMed: 27630371]
67. Rambo RP. Resolving individual components in protein-RNA complexes using small-angle X-ray scattering experiments. *Methods Enzymol.* 2015; 558:363–90. [PubMed: 26068747]
68. Dzananovic E, Patel TR, Chojnowski G, Boniecki MJ, Deo S, McEleney K, Harding SE, Bujnicki JM, McKenna SA. Solution conformation of adenovirus virus associated RNA-I and its interaction with PKR. *J Struct Biol.* 2014; 185(1):48–57. [PubMed: 24291322]
69. Dzananovic E, Patel TR, Deo S, McEleney K, Stetefeld J, McKenna SA. Recognition of viral RNA stem-loops by the tandem double-stranded RNA binding domains of PKR. *RNA.* 2013; 19(3):333–44. [PubMed: 23329698]
70. Svergun DI, Nierhaus KH. A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome. *J Biol Chem.* 2000; 275(19):14432–9. [PubMed: 10799526]
71. Fang X, Stagno JR, Bhandari YR, Zuo X, Wang YX. Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr Opin Struct Biol.* 2015; 30:147–60. [PubMed: 25765781]
72. Lipfert J, Ouellet J, Norman DG, Doniach S, Lilley DM. The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure.* 2008; 16(9):1357–67. [PubMed: 18786398]
73. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr.* 2001; 34:33–41.
74. Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX. A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J Mol Biol.* 2009; 393(3):717–34. [PubMed: 19666030]

75. Bhandari YR, Jiang W, Stahlberg EA, Stagno JR, Wang YX. Modeling RNA topological structures using small angle X-ray scattering. *Methods*. 2016; 103:18–24. [PubMed: 27090001]
76. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*. 2009; 15(2):189–99. [PubMed: 19144906]
77. Yang S, Parisien M, Major F, Roux B. RNA structure determination using SAXS data. *J Phys Chem B*. 2010; 114(31):10039–48. [PubMed: 20684627]
78. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijenyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszyńska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18(4):610–25. [PubMed: 22361291]
79. Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cheng C, Chojnowski G, Chou FC, Cordero P, Cruz JA, Ferre-D'Amare AR, Das R, Ding F, Dokholyan NV, Dunin-Horkawicz S, Kladwang W, Krokhotin A, Lach G, Magnus M, Major F, Mann TH, Masquida B, Matelska D, Meyer M, Peselis A, Popenda M, Purzycka KJ, Serganov A, Stasiewicz J, Szachniuk M, Tandon A, Tian S, Wang J, Xiao Y, Xu X, Zhang J, Zhao P, Zok T, Westhof E. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*. 2015; 21(6):1066–84. [PubMed: 25883046]
80. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*. 2008; 452(7183):51–5. [PubMed: 18322526]
81. Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci U S A*. 2007; 104(37):14664–9. [PubMed: 17726102]
82. Cao S, Chen SJ. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B*. 2011; 115(14):4216–26. [PubMed: 21413701]
83. Zuo X, Wang J, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX. Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc*. 2008; 130(11):3292–3. [PubMed: 18302388]
84. Burke JE, Sashital DG, Zuo X, Wang YX, Butcher SE. Structure of the yeast U2/U6 snRNA complex. *RNA*. 2012; 18(4):673–83. [PubMed: 22328579]
85. Fang X, Wang J, O'Carroll IP, Mitchell M, Zuo X, Wang Y, Yu P, Liu Y, Rausch JW, Dyba MA, Kjems J, Schwieters CD, Seifert S, Winans RE, Watts NR, Stahl SJ, Wingfield PT, Byrd RA, Le Grice SF, Rein A, Wang YX. An unusual topological structure of the HIV-1 Rev response element. *Cell*. 2013; 155(3):594–605. [PubMed: 24243017]
86. Grishaev A, Ying J, Canny MD, Pardi A, Bax A. Solution structure of tRNA^{Val} from refinement of homology model against residual dipolar coupling and SAXS data. *J Biomol NMR*. 2008; 42(2):99–109. [PubMed: 18787959]
87. Wang YX, Zuo X, Wang J, Yu P, Butcher SE. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. *Methods*. 2010; 52(2):180–91. [PubMed: 20554045]
88. Zuo X, Wang J, Yu P, Eyler D, Xu H, Starich MR, Tiede DM, Simon AE, Kasprzak W, Schwieters CD, Shapiro BA, Wang YX. Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of turnip crinkle virus. *Proc Natl Acad Sci U S A*. 2010; 107(4):1385–90. [PubMed: 20080629]

HIGHLIGHTS

- 1.** SAXS is ideal for RNAs that are intractable for crystallography, NMR and cryo-EM
- 2.** Non-canonical synthetase functions include recognition of tRNA-like elements
- 3.** Studying tRNA-like elements by SAXS requires rigorous sample preparation
- 4.** Preparation of RNA and RNA-protein complexes for SAXS studies
- 5.** Principles for accurate analysis of SAXS data is presented

Figure 1. Common problems with SAXS data

Many issues related to polydispersity can be identified by initial visual inspection of the SAXS data such as (A) aggregation or radiation damage, which manifests as a gradual upward slope in the Guinier plot that increases with sample concentration or exposure time, (B) interparticle repulsion, identified by a gradual downward slope in the low- q region of the SAXS data, and (C) production of multimers or low-molecular weight aggregates with increasing concentration.

Figure 2. Identification of foldedness or globularity in Kratky plots

Compact globular macromolecules will exhibit a bell-shaped Kratky plot with non-globular molecules yielding a range of Kratky plot shapes where the high- q region does not regress back toward the X-axis.

Figure 3. Merging of SAXS data

To alleviate issues with detector saturation or large error in high-concentration or long-exposure data (black line) and poor signal to noise in low-concentration or short-exposure data (red line), data can be truncated (truncations denoted by barbells) and merged to yield a single curve with these complicating features absent (bottom panel).

Figure 4. Generating SAXS-derived models

(A) For generating *ab initio* models, the SAXS data is transformed to generate a PDDF that is a histogram of all inter-electron distances within a molecule (e^- distances) which is used to generate molecular envelopes. This method is not able to generate all-atom models (top right). (B) All-atom modeling starts with a secondary structure from which many *de novo* models are generated. These models are then screened for the ones that best fit the SAXS data. The model can then be refined using SAXS data as a restraint to obtain models that better fit the data.

Figure 5. Isosteric thymidine analog, BrdU, allows validation of all-atom model
 Back calculated SAXS curve for an all-atom model of the primer binding site domain (PBS) in the HIV-1 5' UTR annealed to a complementary 18-nt DNA oligonucleotide designed to mimic the 3' end of the tRNA^{Lys3} reverse transcription primer (blue line) closely matches experimental SAXS data (blue dots). Replacement of four thymidine residues in the 18-nt DNA with BrdU produces predictable changes in the back-calculated curve (red line) that also very closely matches the SAXS data of that complex (red dots).

Table 1

SAXS-enabled beamlines in the United States.

Beamline Location	Mail-in	SEC-SAXS	Robotic Sample Injection
SIBYLS Beamline 12.3.1 Advanced Light Source Lawrence Berkeley National Laboratory http://sibyls.als.lbl.gov	YES	YES	YES
BioCAT Beamline 18-ID Advanced Photon Source Argonne National Laboratory http://www.bio.aps.anl.gov	NO	YES	YES
Beamline BL4-2 Stanford Synchrotron Radiation Light Source SLAC National Accelerator Laboratory https://www-ssrl.slac.stanford.edu/~saxs/	NO	YES	YES
Beamline G1 Cornell High Energy Synchrotron Source Cornell University http://www.macchess.cornell.edu/MacCHESS/biosaxs.html	NO	YES	YES
LiX Beamline 16-ID National Synchrotron Light Source II Brookhaven National Laboratory https://www.bnl.gov/ps/beamlines/beamline.php?b=LiX	NO	YES	YES

A comprehensive list of all light sources worldwide along with their capabilities can be found at: <http://www.lightsources.org/light-source-facility-information>