A general small-angle X-ray scattering-based screening protocol validated for protein-RNA interactions

Po-chia Chen,* †Pawel Masiewicz, †Vladimir Rybin, † Dmitri Svergun, ‡ and Janosch Hennig* †

† Structural and Computational Biology Unit, EMBL Heidelberg, Meyerhofstrasse 1, 69126
Heidelberg, Germany
‡ EMBL Hamburg, DESY, Building 25A, Notkestraße 85, 22607 Hamburg, Germany

E-mail: pchen@embl.de; janosch.hennig@embl.de

Abstract

We present a screening protocol utilizing small-angle X-ray scattering (SAXS) to obtain structural information of biomolecular interactions independent of prior knowledge, so as to complement affinity-based screening and provide leads for further exploration. This protocol categorizes ligand titrations by computing pairwise agreement between curves, and separately estimates affinities by quantifying complex formation as a departure from the linear sum properties of solution SAXS.

The protocol is validated by sparse sequence search around the native poly-uridine RNA motifs of the two-RRM domain Sex-lethal protein (Sxl). The screening of 35 RNA motifs between 4 to 10 nucleotides reveals a strong variation of resulting complexes, revealed to be preference-switching between 1:1 and 2:2 binding stoichiometries upon addition of structural modelling. Validation of select sequences in isothermal calorimetry and NMR titration retrieves domain-specific roles and function of a guanine anchor. These findings reinforce the suitability of SAXS as a complement in lead identification.

Small angle X-ray scattering (SAXS) intensities from dilute macromolecular solutions contain global structural information averaged over orientations and types of components in the sample, after which, in the absence of interactions can be expressed as a linear combination:

\[ I(q)_{tot} = \sum_i v_i I_i(q), \] (1)

where \( v_i \) and \( I_i(q) \) are the volume fraction and intensity of the \( i \)-th component, respectively, \( q = 4\pi \sin \theta / \lambda \) is a function of scattering angle \( 2\theta \), and \( \lambda \) is the photon wavelength. This linear decomposition property of SAXS enables its usage as a structure-based screening method, where molecular interactions can be detected via the associated change in scattering intensities. Departure from the sum of component signals points to binding, oligomerization, or ligand-induced conformational changes. SAXS is far more sensitive to changes in the global structure than to changes due to binding of small ligands alone, and this distinguishes SAXS-based screens from affinity-based techniques such as isothermal titration calorimetry (ITC). The combination of structural and energetic approaches can thus distinguish a binding event from downstream impacts on macromolecular structure and dynamics. This possibility is tantalizing for ligand identification as either agonists or antagonists, but also for interactions between proteins and nucleic acids.

The identification of a high affinity RNA
Figure 1: Illustration of the SAXScreen workflow on a single screening against 13 RNAs at the European Synchrotron Radiation Facility (ESRF) beamline BM29. Raw scattering intensities in (a) of the protein and RNA mixtures plus their respective buffers can be used directly to compute $\chi_{lin}$. To compute standard parameters such as $V_C$ and $R_g$, buffer subtraction is performed using calculated protein–RNA addition ratios (b), and then smoothed using AUTOGNOM of ATSAS2.8 to reduce noise and remove low and high-$q$ artefacts (c). These three parameters measuring individual aspects of the SAXS curve can be formulated as titration curves (d), which can be fitted simultaneously to derive $K_{D eff}^\alpha$ (e). Variations in $K_{D eff}^\alpha$ reflect the effect of binding on different parameters. An alternative method to visualise variations is to compute the pair-wise reduced-$\chi$ between buffer-subtracted $I(q)$ that share the same titration ratio (f).

motif with a well-defined binding configuration is a prerequisite for structure determination of resulting protein-RNA complexes via X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. However, RNA-binding proteins (RBPs) often exhibit both specific and non-specific binding depending on the oligonucleotide sequence. This complicates the identification of cognate RNA sequences for RBPs. Despite recent methodological advances, RNA sequence identification also suffers a number of limitations: i) maximal sequence length of nine nucleotides, ii) expensive RNA library generation and/or sequencing steps, and iii) limited transferability of in vivo motifs to in vitro conditions due to potential cooperativity with other RBPs in the live cell. Utilising synchrotron beamlines for screening mitigates the costs of expensive RNA synthesis, where measurement automation further permit a throughput and reduced sample requirement that is more efficient than alternative structural methods such as NMR, and comparable to the recently published switchSENSE. While SANS-based ligand identification has been previously reported, we sought to create a generalizable analysis workflow that can be adapted to the wide range of systems measurable by SANS. The resulting workflow for SANS-based screening (SAXScreen) ranks putative interactions based on intensities alone, but can be further augmented by field-specific knowledge and software. This method is applied to describe the affinity profile and binding mechanisms of the two RRM-domain protein Sex-lethal (Sxl) based on its known interaction with G-substituted poly-U rich oligonucleotides, thus illustrating SAXScreen’s role as a hypothesis generator to identify cognate RNA motifs and their potential binding mechanisms.

An illustrated example of the SAXScreen workflow is shown in Figure 1, containing results of a short synchrotron session. In biological SAXS, the raw scattering intensity $I(q)$ contains contributions from sample molecules $I_s$ as well as their buffer components $I_b$. When...
interactions between a large receptor $R$ and a small ligand $L$ obey the simple equilibrium relation $R + L \rightleftharpoons RL$, mixtures between $R$ and $L$ can be decomposed into five components: the two unbound molecules, their respective buffers, plus the bound complex $RL$. Since all components except $I_{RL}$ can be independently derived by measurement in isolation (Fig. 1a), $I_{RL}$ can be formulated as a perturbation from the linear sum (eq. 1) and quantified via weighted $\chi$-minimization:

$$\chi^2_{lin} = \frac{1}{N_q} \sum_q \left( \frac{(I_{max} - (aI_R + bI_L + cI_R + dI_L))}{2} \right)^2 \left( \frac{(\delta I_R)^2 + (\delta I_L)^2 + (\delta I_R)^2 + (\delta I_L)^2}{2} \right)$$

(2)

The stepwise formation of $RL$ in a titration can therefore be modelled by changes in $\chi_{lin}$, until saturation to an unknown maximum value $\chi^\text{max}_{lin}$. When well-sampled with observed ligand saturation, this dosage response curve can be fitted to yield the dissociation constant $K_D$.

In order to reduce RNA synthesis costs, we assume tentatively that all ligands share a comparable binding mechanism and thus the same $\chi^\text{max}_{lin}$. This assumption captures ligand-specific influences upon $\chi_{lin}$ and other deviations from two-state binding within a modified affinity $K_D^{\text{eff}}$ (see Methods), which expresses the relative efficacy of a ligand to the most effective member of a screen under non-saturation conditions. This allows all dosage response curves to be fitted simultaneously, which produces a final ligand ranking based on $K_D^{\text{eff}}$. Aside from $\chi_{lin}$, specific characteristics of biomolecular SAXS such as particle volume $PV$, radius of gyration $R_g$ and volume of correlation $V_C$ can also be fitted. Here, automated processing of buffer-subtracted curves is utilised to reproducibly derive titration curves (Fig. 1b-e), resulting in comparable qualitative $K_D^{\text{eff}}$ rankings. In addition, SAXS curves at equivalent titration points can be directly compared via the pairwise reduced-$\chi$ (Fig. 1f), which can be further averaged across all points to compute a net deviation between two titrations for clustering (see below). These matrices are analogous to the structural similarity map adopted by Hura et al. 17.

To validate the SAXS-Screen procedure, we investigated the RNA interactions of a Sxl mutant containing a 10-residue GS-linker extension (Sxl_{10GS}) across multiple synchrotron sessions. This linker mutation was added to enhance the magnitude of SAXS changes upon RNA-binding, noting that equivalent results have been obtained for wildtype Sxl (Figure S1). The full data of three sessions covering 35 titrations are presented in Figures S1 and S2, while we summarize the main Sxl_{10GS} findings in Figure 2.

Using the $\chi_{lin}$ metric, the closely related U_{10} and UGU_{8} exhibit widely disparate $K_D^{\text{eff}}$, contrasting with the small 2- and 10-fold differences detected via ITC and NMR (Table S1 and Figure S4). This arises due to a difference in preferred binding stoichiometry: when ensemble modelling are applied to SAXS curves at 1:1 protein-RNA ratios (Fig. S5), U_{10} replicates are consistent with 1:1 complexes while two of three UGU_{8} curves are instead consistent with 2:2 complexes. All optimised ensembles are dominated by RRMs in close-contact, which is not dependent upon the presence nor sequence identity of added RNA (Fig. S6). Notably, both of the above findings of 2:2-stoichiometry and transient inter-RRM interactions have been previously reported in tandem RRM proteins.

A broader screen further confirms that Sxl_{10GS}’s stoichiometric preference is RNA sequence dependent. The clustering of titration responses (Fig. 2c) identifies four major clusters, three of which correspond to the preferred stoichiometry seen in ensemble-modelling and one corresponding to short sequences that are unable to induce significant SAXS perturbations. Here, the additional active cluster corresponds to sequences such as U_{5}C_{5} that are consistent with both 1:1 and 2:2-stoichiometry, suggesting that both species are present in solution. Reading in closer detail, the categorizations and measured $K_D^{\text{eff}}$ together reveal the impact of the RRM2 G-anchor motif and influence of G-positioning: The location of guanine at the 2nd, 3rd and 8th positions specifically encourage 2:2-binding, while the 9th position is exclusively 1:1. Meanwhile, C_{5}U_{5} is 2:2-binding, but U_{5}C_{5}/UGU_{3}C_{5} are mixed. Given that during crystallization RRM2 orients upon the 5’ and RRM1 upon the 3’, these results...
suggest that RRM2 significantly outcompetes RRM1 in poly-U interactions, thus dominating the sequence selection.

To validate our assumptions based on SAXSscreen, we performed NMR titrations for three pairs of sequences $U_{10}$ vs. UGU$_8$, U$_7$GU$_2$ vs. U$_5$GU, and U$_5$C$_5$ vs. C$_5$U$_5$ (Fig. S8 and S9). The four poly-U sequences exhibit similar $^1$H-$^{15}$N chemical shift perturbations (CSPs) across both RRM1 and RRM2, indicating that the local RNA contacts are retained regardless of stoichiometry. Whereas, the presence of poly-C lead to selective and significant reduction of CSPs relative to poly-U sequences (Fig. S10): U$_5$C$_5$ CSPs indicate preferential reduction of canonical RRM1 contacts, while C$_5$U$_5$ CSPs indicate elimination of inter-RRM contacts. These reinforce the SAXS results, and further suggest that the 2:2 stoichiometry complexes containing poly-C are supported by non-specific charged interactions of RRM1 rather than specific interactions at the canonical sites. We thus summarize the binding model and site-selectivity in Fig. 2d and e).

The role performed by SAXSscreen in the study of Sxl$_{10GS}$ involves exploration of RNA sequence space and classification of binding modes and affinities for further investigation. It successfully confirms optimal sequence length and composition, including G-anchor positioning and anti-C selectivity. The use of $\chi_{lin}$ as the primary metric here shows that structure-specific knowledge – although useful, – is not required to derive valuable information, and highlights possible adaptation to other applications such as whole-cell SAXS. We caution however that $\chi_{lin}$ can suffer from spurious overfitting, e.g. in cases where binding-associated pertur-
bations coincide with buffer differences or parasitic scattering and aggregation have not been eliminated. The former is illustrated in Fig. S11 where $A_{10}$ and $UGU_4$ titration curves are artificially suppressed in $\chi_{in}$ but not $V_C$. We note that structural quantities can themselves suffer from lack of sensitivity in cases where the overall binding process do not result in a net change, e.g. an internal domain reorientation may result in insignificant $R_g$ changes. Therefore cross-validation across several metrics is recommended, which we exemplified by presenting $R_g$, $V_C$, and $\chi_{in}$ together. Such redundant measurements assist in identification of spurious artefacts such as sample contaminations and aggregation.

Several directions are pertinent for ongoing development of SAXScreen. Although practical expenses and Sxl’s sequence-dependent oligomerisation hinder us from measuring RNA–excess conditions and thus evaluate absolute affinities, we can estimate the theoretical sensitivity of this $\approx 20$ kDa protein to span $0.1 \sim 1000 \mu$M at $1 \text{mg ml}^{-1}$ concentrations. The detection limits for discovering minor populations in SAXS is dependent upon the extent of conformational change and the signal-to-noise ratio of beamline setups, while the applicable affinity range is dependent upon available molar concentrations as dictated by the mass concentrations amenable to scattering measurements. In this context, our protocol represents a balance between throughput and ability to discriminate binding processes by measuring eight titration points below ligand saturation. More measurements per titration are likely required to reproducibly predict binding affinities, as evidenced by sporadic deviations between replicates in our screen (c.f. Fig. S1 and S2). Recommendations and additional considerations will be discussed in the supplementary information.

The optimised protocol at the ESRF and PETRA-III synchrotron beamlines allows for $10^3$ measurements per day, corresponding to $10^2$ titrations depending on the desired precision of affinity estimation. An increase of 1-2 orders of magnitude is expected in the near future with improvements to beamline setups aimed at reducing downtime between measurements. We also note that similar automated setups are available at synchrotron sites around the world, and can be utilised for screening. This puts SAXScreen close to the throughput of biophysics-based screens and can also be used as a source of complementary information, where simultaneous affinity and structural information is required to further refine the data from higher throughput techniques or from in vivo sources. In particular, SXS-based screening is expected to be particularly useful for biochemical processes associated with measurable structural changes.

### Experimental Procedures

We describe below an essential summary of experimental and analysis protocols, and refer readers to the full description in the Supporting Information.

#### Sample preparation

The preparation of wildtype Sxl and Sxl$_{10GS}$ follows previously published protocols, noting that the mutant plasmid contains an additional 10-residue stretch (GGSGSGGGGS) after position 204, between two glycines in the interdomain-linker ARPGGESIK. RNA oligonucleotides were ordered from commercial suppliers Microsynth and IBA to specifications of highest purity, separately for each synchrotron session, and resuspended in milli-Q water to the required concentration for SAXS and NMR experiments.

#### SAXS measurements

Preliminary and production screening sessions were conducted at Hamburg P12 and ESRF BM29 beamlines, respectively, using the automated sample changer setup to simplify measurement on 96-well plates. The plate geometry was leveraged to conduct 6 or 8-point titrations spanning apo-protein to 1:1 mixtures, at 50 $\mu$M protein concentration and 40 $\mu$l final volume. Exposure times was left near recommended val-
ues specified by the site, noting that no radiation damage was observed throughout screening. The frame-averaged scattering intensities were taken as the starting point for automated analysis in SAXScreen below, discarding data outside $0.3 \sim q \sim 3.0 \text{nm}^{-1}$ to universally exclude aggregation, parasitic scattering, and high-noise regions.

**Ensemble-based modelling**

The Ensemble-Optimization Method (EOM) software was used to derive structural models consistent with experimental scattering. The crystallographic 1:1-Sxl complex as assumed to represent canonical binding, with the attachment of RNA to RRM domains varied to represent different encounter complexes. Both 1:1 and 2:2 stoichiometry models were proposed, and initial ensembles of $10^4$ and $10^6$ (respectively) was generated for fitting against SAXS curves at 1:1 protein–RNA ratios.

**SAXScreen analysis**

Buffer subtraction of raw intensity curves was carried out manually to take into account different input ratios of protein buffer versus milli-Q water used to resuspend RNA, using the averaged buffer-water difference signal across the beamtime session as the correction curve.

**Heirarchical clustering**

The pairwise reduced-$\chi$ between two SAXS curves is defined as:

$$\chi_{ab} = \frac{1}{N_q - 1} \sum_{i} \frac{(I(q_i)_a - F \ast I(q_i)_b - C)^2}{\delta I(q_i)_a^2 + \delta I(q_i)_b^2},$$

after minimisation with respect to two free parameters $F$ and $C$ to mitigate concentration and subtraction differences. The instance metric required for heirarchical clustering of titrations is produced by calculating $\chi_{ab}$ for all pairs of SAXS curves at identical protein–RNA ratios, and then defining distance $d_{ab}$ between two titrations as the symmetrized average $\chi_{ab}$ across all protein–RNA ratios as follows:

$$d_{ab} = \max \left( \frac{\chi_{ab} + \chi_{ba}}{2} - 1, 0.0 \right),$$

Analogous clustering procedures can be carried out using the structural similarity map, which removes the bias of $\chi_{ab}$ towards high-intensity regions with lower signal-to-noise ratios.

**Computation of $I(q)$ features**

The ATSAS package version 2.7.1 has been utilised to derived smoothed scattering curves for computing of $I(0)$, $R_g$, $V_C$, and unscaled particle volume $PV$. $\chi_{lin}$ is derived directly from the raw scattering intensities using the measured RNA curve, the averaged apo-protein curve, and the averaged buffer and water curves as the four component curves in Eq. 2. This value is again minimized over free parameters $a$, $b$, $c$, and $d$, which are allowed to be negative so as to conduct buffer subtraction. Uncertainties are computed by 100-trial repetition with added gaussian noise using $\delta I(q)$.

**ITC measurements**

The energetics of RNA oligonucleotides binding to Sxl_{10GS} was studied using a MicroCal iTC200 System (Microcal). All proteins were dialyzed overnight against ITC buffer comprising 10 mM sodium phosphate, 50 mM sodium chloride, and 1 mM DTT at pH 6.5. Titrations were performed at 25°C by stepwise addition of 200-500 uM RNA solution to the 10-20uM protein solution in the cell. Data were corrected for the heat of dilution and analysed using the MicroCal OriginTM software package.

**NMR titration experiments**

Replicate samples are prepared under the same conditions as SAXS except that $^{15}$N-labelled Sxl is used to measure RNA-binding via chemical shift perturbations (CSP). The NMR buffer comprises 10 mM sodium phosphate, 50 mM sodium chloride, and 10 mM DTT at pH 6.5.
1H–15N HSQC spectra have been recorded on a Bruker Avance III with a magnetic field strength corresponding to a proton Larmor frequency of 700.303 MHz, starting at pure Sxl10GS with stepwise addition of corresponding RNA. See supplementary information for a detailed titration protocol.

Plotting software

Visualisation of crystal structures was done in VMD-1.9.2\(^2\) and rendered with Tachyon. All HSQC spectra presented here have been plotted using python modules nmrglue and matplotlib. Most graphs have been visualised in xmgrace, while \(\chi\)-matrices have been visualised in python or gnuplot-5.

Code availability and maintenance

The SAXScreen workflow codebase used for this work has been made available on GitHub under the MIT license:

https://github.com/zharmad/SAXScreen

Data availability

The titration data over SAXS, ITC, and NMR are available from the authors upon request. Four representative SAXS data have been deposited in SASBDB with accession codes 775 to 778, containing apo protein curves for Sxl, Sxl10GS, 1:1 mixture of Sxl10GS:G[2]U8GU, and 1:1 mixture of Sxl10GS:G[3]UGU8.

Acknowledgement

We thank the ESRF Grenoble and DESY Hamburg synchrotrons for providing beamtime allotments. We also acknowledge the assistance of Lukas Rohland in sample preparation for a SAXS measurement. P.C. was supported by the EIPOD postdoctoral programme co-funded by the European Molecular Biology Laboratory (EMBL) and the Marie Curie Actions Cofund grant MSCA-COFUND-FP #664726. J.H. gratefully acknowledges the EMBL and the German Research Council (Deutsche Forschungsgemeinschaft, DFG) for support via an Emmy-Noether Fellowship and the Priority Programme SPP1935.

Supporting Information Available

Detailed methods, and all supporting tables and figures are found in the Supporting Information.

References


For Table of Contents Only