Hybrid Applications of Solution Scattering to Aid Structural Biology

13

Alexander V. Grishaev

Abstract

Biomolecular applications of solution X-ray and neutron scattering (SAXS and SANS, respectively) started in late 1960s – early 1970s but were relatively limited in their ability to provide a detailed structural picture and lagged behind what became the two primary methods of experimental structural biology – X-ray crystallography and NMR. However, improvements in both data analysis and instrumentation led to an explosive growth in the number of studies that used small-angle scattering (SAS) for investigation of macromolecular structure, often in combination with other biophysical techniques. Such hybrid applications are nowadays quickly becoming a norm whenever scattering data are used for two reasons. First, it is generally accepted that SAS data on their own cannot lead to a uniquely defined high-resolution structural model, creating a need for supplementing them with information from complementary techniques. Second, solution scattering data are frequently applied in situations when a method such NMR or X-ray crystallography cannot provide a satisfactory structural picture, which makes these additional restraints highly desirable. Maturation of the hybrid bio-SAS approaches brings to light new questions including completeness of the conformational space sampling, model validation, and data compatibility.

Keywords

Solution X-ray and neutron scattering • SAXS • SANS • Hybrid modeling • Protein and RNA structure determination

13.1 Introduction

Scattering of X-rays or neutrons by an isotropic solution containing the macromolecule of interest produces a one-dimensional intensity signal

A.V. Grishaev (✉)
National Institute of Standards and Technology, The Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850, USA
e-mail: agrishaev@ibbr.umd.edu

B. Chaudhuri et al. (eds.), *Biological Small Angle Scattering: Techniques, Strategies and Tips*, Advances in Experimental Medicine and Biology 1009, DOI 10.1007/978-981-10-6038-0_13

dependent on the scattering angle and is commonly expressed as a function of the scattering vector $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident radiation (X-ray photons or neutrons). A difference signal between such one-dimensional scattering intensity profiles recorded for the sample containing the bio-macromolecule of interest and the one containing an identically matching buffer can then be interpreted solely in the terms of the macromolecular structures present in solution, their surrounding surface solvent layer and, possibly, their inter-particle correlations. Solution scattering data are most often recorded at angles much smaller than the ones sampled in the macromolecular crystallography leading to a notion that these data can only provide structural information with a low nominal resolution. Considering a dramatic difference between a smooth one-dimensional solution scattering curve and an indexable three-dimensional diffraction set containing tens of thousands of reflections common in macromolecular crystallography, it seems natural to treat solution scattering as a technique with an intrinsically low information content. This notion is supported by a fairly small number of degrees of freedom associated with a solution scattering data set. With data recorded within the q-range from q_{min} to q_{max} for a macromolecule with the maximum dimension d_{max}, this number is estimated from the Shannon-Nyquist sampling theorem as $N = \pi(q_{max}\text{-}q_{min})*d_{max}$ (Svergun and Koch 2003), and in practice rarely exceeds ca. 20, indicating that only a limited number of model parameters could be obtained from such data.

The question is then, given this apparent low resolution and low information content, why are these data useful at all in modern structural biology which aims for sub-Å precision of atomic positions? The answer to this question lies in the ability of SAXS data to offset the formal resolution and information content limitations noted above with a superior signal/noise attainable when using modern synchrotron sources and single photon counting detectors. Contrary to the seemingly featureless and noisy scattering profiles that were common up till ~20 years ago unless very high protein concentrations were used (Durschlag 1975; Fedorov and Denesyuk 1978), modern synchrotron beam lines are capable of producing solution scattering data with very low visible noise and wide angular range, largely free of inter-particle correlations effects while handling relatively dilute samples. These changes occurred as SAXS, as a field, was able to capitalize on the same dramatic improvements in the photon flux and low-noise detector technologies that propelled the growth of the bio-molecular X-ray crystallography in the past two decades and are now responsible for the explosive growth in applications of cryo-electron microscopy.

These improvements in the instrumentation were occurring simultaneously with three crucial developments in data analysis capitalizing on the higher quality of SAS data, specifically: (i) generation of distance probability distribution functions *via* regularized Fourier transforms of the scattering data (Svergun 1992; Brunner-Popela and Glatter 1997), (ii) development of *ab initio* low resolution electron density reconstruction approaches (Chacon et al. 1998; Svergun 1999), and (iii) formulation of approaches for accurate calculation of the scattering data from the macromolecular atomic coordinates taking into account the surface solvent layer, groundwork for which was laid in mid-1990s (Svergun et al. 1995, 1998). Taken together, these developments now create a possibility of using solution scattering data to effectively discriminate between closely related candidate structural models with a representative example shown in Fig. 13.1. This ability to select between the candidate structural models taking advantage of accurate methodologies for connecting scattering data to the macromolecular atomic coordinates is central to the hybrid approaches that use SAXS/SANS in combination with the restraints from other experimental techniques (Putnam et al. 2007).

13.2 SAS Data: Advantages and Drawbacks

The main practical advantage of solution scattering data results from their ability to reflect the

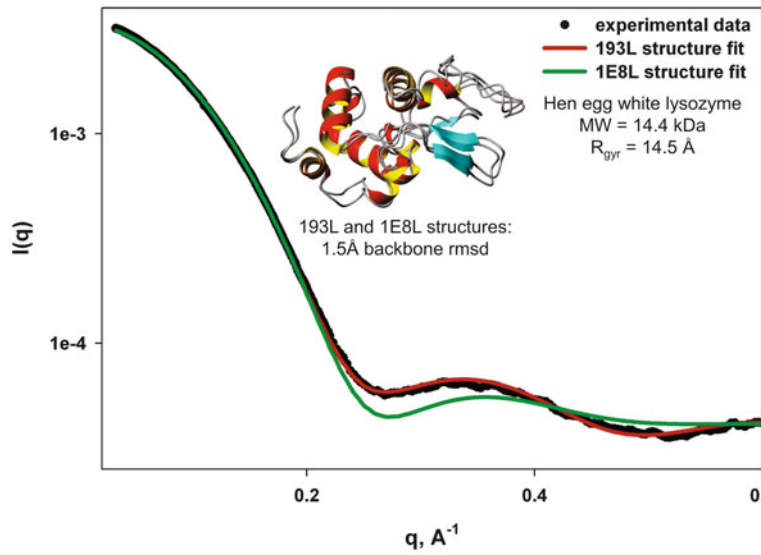

Fig. 13.1 Solution X-ray scattering data can be very effective at distinguishing between closely related structural models describing the macromolecule

molecular geometries for the complete set of conformations present in solution, covering such important and challenging cases as intrinsically disordered proteins (IDPs) (Mittag et al. 2010), detergent micelle-solubilized membrane proteins (Lipfert et al. 2007), or macromolecules with dynamic regions, including flexibly linked multi-domain proteins (Bernado et al. 2007), or amyloid fibrils (Lu et al. 2003), as illustrated in Fig. 13.2. SAXS is applied with a similar ease to proteins, RNA, or DNA, with the latter two benefiting from an increased signal precision due to the presence of phosphorus atoms and a lower fraction of hydrogen atoms relative to the proteins (Zuo and Tiede 2004).

Solution neutron scattering brings an additional possibility of performing contrast variation measurements, particularly informative for multi-subunit system which can be prepared with selective protonation/deuteration of the individual components, or protein/oligonucleotide complexes (Whitten and Trewhella 2009). Such contrast-variation, or contrast-matched scattering data collected when varying the H_2O/D_2O ratios in the buffer, can be distinctly different from SAXS data in H_2O, increasing the overall information content of the solution scattering data set.

SAS data can be collected with ease within a wide range of experimental conditions, including salt concentration up to 1–2 molar, temperature from the freezing point to ca. 90 °C, or with solute concentrations as low as 0.02–0.05 mg/mL for SAXS and 0.5–1.0 mg/mL for SANS (Grishaev 2012).

Information content of the scattering data is limited relative to the crystal diffraction data set, consistent with their one-dimensional nature. SAS intensity profile reports on the entire scattering particle, thus providing very little site-specific information unlike techniques such as macromolecular crystallography or NMR. However, this is one of the reasons for the growing popularity of the hybrid techniques that include SAXS or SANS: complementary biophysical techniques capable of providing precise site-specific restraints can be very effective at compensating the lack of such information in solution scattering.

A practical limitation of solution scattering is the requirement for a highly monodisperse and pure sample needed for the data analysis in terms of the structural model (Jacques and Trewhella 2010). From our own experience, the presence of ~5% dimer fraction relative to the monomeric species of interest produces a noticeable impact

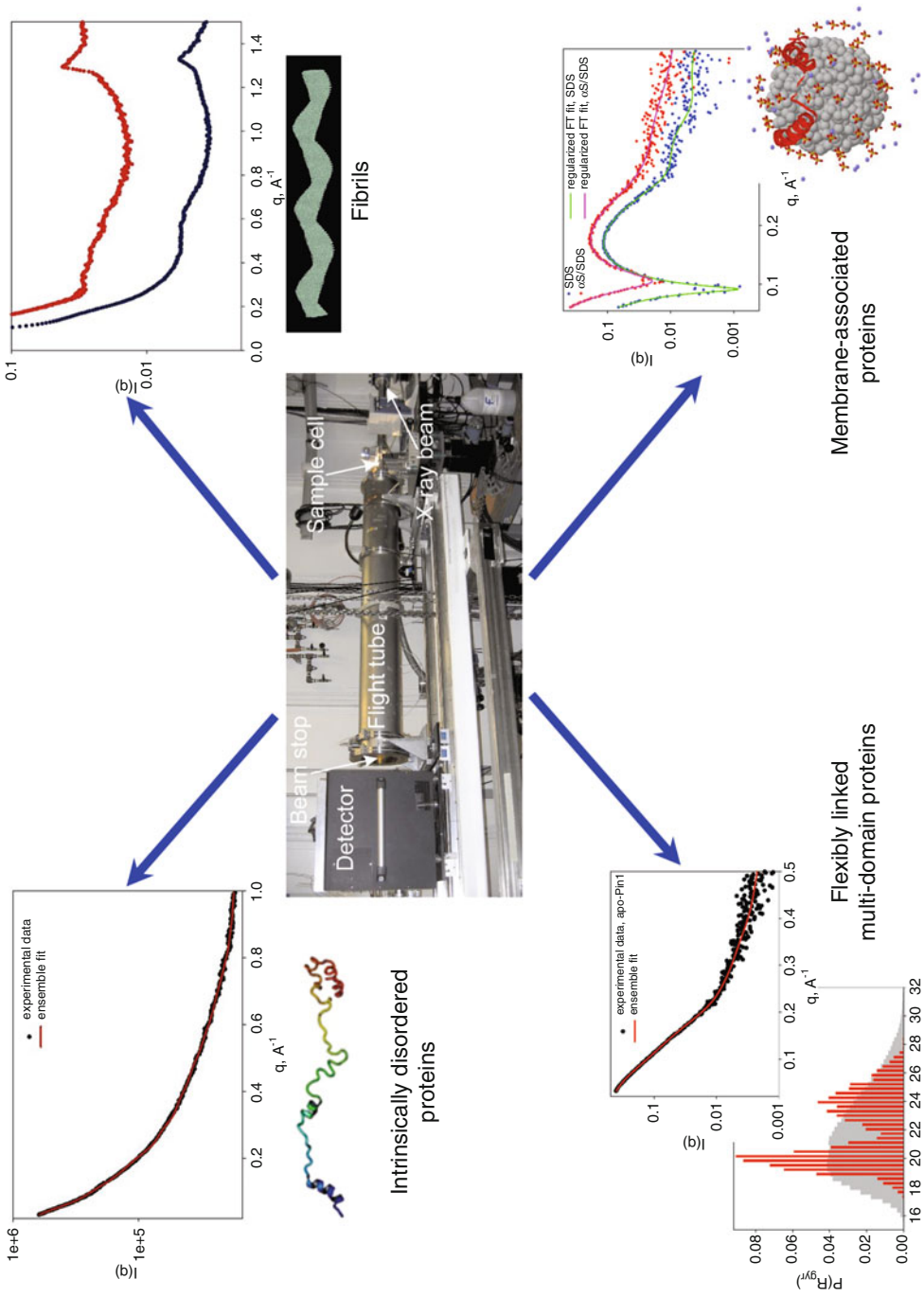

Fig. 13.2 Solution scattering can be applied to study a diverse set of challenging systems in structural biology, frequently in combination with other biophysical techniques

on the SAS data within a wide angular range, and in the case of the aggregates ~10 fold larger relative to the mass of the molecule of interest, this threshold can become as low as ~0.1%, reminiscent of the situation encountered with light scattering data analysis. Even though the main effect of such impurities comes at low scattering angles, it cannot always be removed by simply discarding these data and becomes particularly detrimental with highly elongated macromolecular geometries such as those observed for the for the IDPs (Johansen et al. 2011).

13.3 SAS Data: Complementary Sources of Information

Owing to the limited information content of the solution scattering data noted above, they are now increasingly more often combined with the restraints from other techniques in structural studies. This was not the case in the early days of bio-SAS applications – 15 years ago a study involving SAXS or SANS rarely incorporated the results of complementary experimental techniques, with the end result commonly being formulated in terms of the low-resolution electron density determined solely based on the scattering intensities. This reflected both fascination with the newly found ability to derive realistic three dimensional shapes solely from 1D SAXS data – clearly, a milestone development, as well as the general lack of computational methods with which these data could be integrated with precise site-specific experimental restraints. Needless to say, it did not take long for such techniques to be formulated, with the early applications concentrating on the rigid-body refinement against SAXS data when structures of the individual subunits of the overall scattering particle were known (Petoukhov and Svergun 2005), or incorporation of the fixed relative orientational restraints from NMR (Sunnerhagen et al. 1996; Mattinen 2002).

13.3.1 Hybrid Structure Determination Using SAS Data and Partial Structural Information from X-Ray Crystallography

SAS data can be readily combined with the partial structural information from a complementary high-resolution technique such as X-ray crystallography, taking advantage of a number of computational methodologies developed for accurate comparison of the experimental scattering data with those predicted from the candidate all-atom models (Svergun et al. 1995; Park et al. 2009; Schneidman-Duhovny 2010; Grishaev et al. 2010; Koefinger and Hummer 2013; Chen and Hub 2014). The differences between these approaches are subtle but numerous, including: the number and nature of the fitting parameters, fit of the buffer-subtracted scattering signal or the pair of the sample and buffer scattering profiles, specifications for the exact positioning of the surface solvent layer relative to the macromolecule, and finally, the use of implicit or explicit models for the displaced and surface solvent. Implicit models are those that assume the waters displaced by the macromolecule to exactly coincide with the macromolecular atomic coordinates and use a simplified "shell" representation for the surface solvent layer. In contrast, explicit solvent models rely on the results of the molecular dynamics simulations to describe the structure of the displaced and surface solvent. It would be fair to state that all of the methods listed above were more extensively tested and optimized for proteins compared to oligonucleotides.

Several scenarios for combining SAS data with prior structural knowledge are possible. In one, the system is composed of a number of subunits with known structures whose relative arrangements are derived from fitting the candidate all-atoms models to the SAS data. The need for using SAS data in this case stems either from difficulties in finding conditions for obtaining

well-diffracting crystals for the entire assembly (Comoletti et al. 2007), or from the impact of the crystal packing forces on the overall geometry for a highly non-globular particle (Heidorn and Trewhella 1988). A second scenario is when the crystal structure is either missing interpretable electron density for the fraction of its sequence due to dynamics, or when the full-length construct with flexible parts cannot be crystallized, necessitating the use of a truncated variant (Hickman et al. 2014). In these two cases either the relative arrangement of the particle subunits, or the coordinates for the missing fraction of the macromolecule have been positioned to agree with the experimental SAS data. Finally, the scattering data can become a useful constraint for determining the geometry of the assemblies such as protein/protein complexes when the structures of the individual partners are known (Schneiderman-Duhovny et al. 2010; Pons et al. 2010; Karaka and Bonvin 2013). In all of these cases, constraints based as chain connectivity, clash avoidance, and gyration radius by reference to the experimental SAS data, can all be very effective at decreasing the number of the feasible candidate models and are best applied prior to the full SAS data fit. A complication possible with all of these scenarios occurs when the experimentally determined structural models are not available for some, or all of the system subunits, leading to the use of homology-based structures with worse than ~2 Å coordinate accuracy (Comoletti et al. 2007). The consequence of relying on such lower quality subunit models is that the residual structural inaccuracies embedded in them can propagate into the structural inaccuracy of the best-fitted model for the overall assembly.

In all of the above cases, both combination of SAS information with the partial structural data from X-crystallography, and optimization of the model geometry to best-fit the scattering intensities are relatively straightforward to perform. Far more challenging is establishing that the best-fitting model obtained starting the fit from a set of initial coordinates is accurate, representing the actual set of conformations present in solution, or even unique. Possible ways to address this issue include (i) supplementing SAXS data with complementary data from other experimental techniques for model validation and refinement, and (ii) performing an exhaustive sampling of a large number of starting/candidate geometries with preset rotational and translational steps. This problem is further complicated for flexible or disordered macromolecules, for which data over-fitting becomes much more likely due to the increase in the number of model parameters. In general, proof of a non-degenerate nature of the structural model found to best-fit SAS data can be difficult to establish due to the potentially astronomical number of possible distinct candidate geometries that need to be sampled, reaching ca. 10^{11} even for multi-domain proteins linked by short (less than 10 amino acid) stretches of residues (Grishaev et al. 2012). This issue becomes more pronounced for protein/protein complexes where the chain connectivity constraints cannot be applied (Schneidman-Duhovny et al. 2011). Such cases will benefit from additional relative geometry constraints, which may include mutagenesis data reporting on complex breakage, or other prior information allowing approximate determination of the locations of the subunit/subunit interaction sites. It is worth keeping in mind that the structural model best-fitted to the SAS data using rigidly held high-resolution structures of the individual subunits cannot be expected to have the same structural accuracy as its constituents even though it will appear to resemble a high-resolution structure. For instance, relative domain orientations resulting from the model optimized against SAS data cannot be assumed to be uniquely and correctly identifiable even for the non-globular subunits, in the absence of validation *via* additional experimental restraints.

An additional challenge while performing SAS data driven macromolecular structure determination is that there is generally no guarantee of a perfect correlation between the structural accuracy of the candidate model and the goodness of its SAS data fit (Grishaev et al. 2011). The exact appearance of such correlation plot, which requires the knowledge of the correct structure can be impacted by a multitude of factors

including the particular method used to predict SAXS/SANS data from the atomic coordinates, limited experimental signal/noise or resolution range of the scattering data, possible inaccuracies for the parts of the structural model that are held fixed, improper representation of the multiple conformations present in solution, or errors in the interpretation for SAS data due to unrecognized contributions of aggregation, sample/buffer mismatch, or sample purity or composition issues. Therefore, when performing structural model selection against solution scattering data, it is advisable to use several methodologies for linking SAS data to the atomic coordinates deriving the set of consensus models, tightly control systematic errors in the data outside of the photon counting statistics, and use a proper representation for the conformational heterogeneity when necessary.

It should be clear from the above comments that the issue of validation of the structural model restrained by SAS experimental data is of utmost importance. Such validation should rely on complementary experimental restraints weakly correlated with the fitted scattering data set. A possible way to introduce this complementarity is by recording contrast variation solution scattering data with SANS (Comoletti et al. 2007) or SAXS (Grishaev et al. 2012), allowing either a reduction in the degeneracy in the pool of the best-fitting structural models, or validation in case of a single best-fitting solution.

It is also worth pointing out that the ultimate success of using SANS data with mixed $^{1}H/^{2}H$ labeling of the complex constituents at the contrast matching conditions (~42% D_2O for protonated proteins) in practice can be affected by both the dissociation constant of the complex and the relative masses of its protonated and deuterated components. In one case (Comoletti et al. 2007), accurate determination of the centers-of-mass separation between the deuterated components with protonated subunits contrast-matched was possible for a 2:2 protein/protein complex with low-nM affinity where the deuterated units comprised ca. 25% of the overall particle. In another (Schwieters et al. 2010), such determination was not successful for a nearly identical SANS data collection on a similarly sized 2:2 protein complex with ~20 μM binding affinity requiring a sixfold molar excess of the smaller deuterated component needed in order to obtain 100% binding occupancy, and deuterated components comprising only ca. 13% of the complete particle.

13.3.2 Hybrid NMR/SAS Macromolecular Structure Determination

Nuclear magnetic resonance (NMR) in solution had been long recognized to be a useful complementary source of structural information when combined with the SAXS data. The symbiosis of the two techniques is driven by the fact that they tend to offset each other's deficiencies such as the decrease in the density of attainable restraints with the increase in the size of the macromolecule, or difficulties in describing the molecular structure of flexible constructs, in the case of NMR; or the lack of site-specific information and model degeneracy in the case of SAS. Solution scattering data should ideally be acquired on the same sample used for NMR data collection, minimizing chances of any inconsistencies between the experimental conditions for the two measurements.

From a practical perspective, macromolecular samples used in solution NMR rarely require significant modifications for application of SAXS (Grishaev 2012). $^{2}H/^{13}C/^{15}N$ isotopic labeling has no effect on the X-ray scattering data while the presence of ^{2}H does contribute to the signal measured by SANS. Therefore, the least expensive (in practice, ^{15}N–labeleld) material frequently ends up being used for solution scattering measurements. Slight changes in the SAXS buffer composition relative to typical NMR conditions may include increase in the salt concentration to suppress inter-particle repulsion, replacement of the commonly used phosphate in the NMR buffers by agents containing lighter elements (TRIS, HEPES, etc), and addition of the free radical scavengers such as DTT or TCEP to the dialysis buffers in

order to suppress protein radiation damage. 5–10% D_2O typically included in the NMR samples for frequency lock does not need to be present for SAXS data collections, most commonly carried out in pure H_2O. X-ray scattering measurements are typically done at the same (or lower) concentrations as those used in solution NMR and should be performed at the temperature exactly matching NMR data collection for seamless combination of the restraints from the two techniques. Monodispersity requirements for SAXS are more stringent compared to NMR where the aggregated populations simply becomes invisible, and preliminary characterization by techniques such as analytical ultracentrifugation or light scattering is a must. A growing number of bio-SAXS beam lines at synchrotrons now offer in-line size-exclusion chromatography setups immediately preceding SAXS measurement which greatly enhance the quality of the collected scattering data at the expense of the decrease in sample concentration and, therefore, the signal/noise of the recorded data. While planning SAXS measurement in the context of an NMR study it is always a good idea to predict theoretical scattering curves beforehand when structural models are available. This step will help to both select the appropriate experimental angular range, and estimate the resolving power of the SAXS by reference to the uncertainty of the structural models resulting from NMR-only structure determination.

It is now generally accepted that out of all types of currently accessible NMR restraints, residual dipolar couplings (RDCs) observed *via* weak alignment of the macromolecules induced by the strongly aligned liquid crystalline media (Bax 2003), are by far the most useful type of data when combined with the solution scattering intensity profiles. This complementarity is easily rationalized keeping in mind both relative insensitivity of the solution scattering data to domain rotations around their centers of mass, and independence of the orientational restraints from NMR from the translations of the particle subunits. In our own early work on combining SAXS and NMR data in a hybrid structure determination we observed no improvement of the structural accuracy when the NMR restraints were composed entirely from the short-range inter-proton distances and torsion angles; such improvement occurred only when RDCs were included in the NMR data set (Grishaev et al. 2005).

Hybrid structure determination using RDC and SAXS data can proceed under two scenarios – rigid body structure optimization involving solely translational and orientational degrees of freedom that specify relative domain positioning, or a fully flexible refinement in which all of the internal degrees of freedom are active. The choice between using the two depends on the density of the available site-specific NMR restraints and the coordinate accuracy for the individual domains, with lower restraint density and higher structural accuracy generally favoring rigid body techniques. As a rule of thumb, fully flexible refinement is warranted when the structures of the individual domains can be determined solely from NMR data with the coordinate accuracy better than ca. 1.5 Å, or backbone N-H RDC cross-validation Q-factors (Bax 2003) better than ca. 0.4. On the other hand, rigid-body refinement techniques are preferable if the available structures of the individual subunits can be fitted to the experimental backbone N-H RDC data with Q-factors better than ca. 0.3, likely associated with higher resolution (better than ca. 2.4 Å) crystal structures.

A joint rigid-body fit of RDC and SAXS data relies on accurate determination of the molecular alignment tensors for the individual domains, which in practice requires measurement of at least ca. 35 backbone RDCs for each domain. In cases when the molecular alignment tensors determined by a singular value decomposition (SVD) fits (Losonczi et al. 1999) for the individual domains are strongly correlated (generalized scalar product between the corresponding alignment tensors larger than ca. 0.9, corresponding to the tensor orientation difference not exceeding ca. 20°), their relative positions can be assumed to be rigid, allowing single-model representation during the structure refinement. In cases of small numbers or large certainties of the experimentally attainable RDCs an approximate criterion

for a single conformation representation would be for the Q-factor of the joint SVD fit of the RDCs for the two domains allowed to reorient not exceeding those fitted separately to the individual domains by more than ~0.05. Otherwise, an ensemble representation for the complete macromolecule would be required, possibly coupled with a scheme for prediction of the conformation-dependent molecular alignment tensor (Zwezkstetter et al. 2004; Marsh et al. 2008; Venditti et al. 2015). When performing both single-model and ensemble-averaged rigid body refinement against RDCs it should be kept in mind that even though the relative orientation of the two domains can be fitted from the experimental NMR data with a precision of ca 3–5°, it comes with a fourfold degeneracy corresponding to 180° rotations around the three axes of the alignment tensor. Even though in theory this degeneracy could be resolved by collecting similarly precise RDCs from a different alignment medium with a weakly correlated alignment tensor, in practice this situation is relatively rare for proteins, and even less common for oligonucleotides. Therefore, all four distinct orientations often need to be sampled in a joint RDC/SAXS data fit. When present, domain connectivity constraints with short linkers not exceeding ca. 10 residues often reduce this fourfold degeneracy by a factor of 2, with the burden of distinguishing between the remaining ones placed solely on the SAXS data. In cases of the protein/protein complexes, unless the locations of the interacting sites on each of the subunits are determined independently, all four possible conformations have to be distinguished based on the SAXS data fits alone, with several successful examples reported in the literature (Parsons et al. 2008; Zuo et al. 2008). It can shown that use of the RDCs as orientational restraints corresponds to an approximately 100-fold reduction in the number the possible candidate structural models when rigid-body refinement is employed (Grishaev et al. 2012).

With a sufficient density of site-specific NMR restraints, fully flexible model refinement against the combined NMR and SAXS data sets becomes possible, often performed either when the complete structure cannot be separated into the individual domains, or when the structural accuracy of the individual domains is insufficient for the application of the rigid body methods according to the criteria listed above. In practice sufficiently high density can correspond to as few as ca. 1 backbone torsion angle restraint and 1–2 distance restraints per residue, supplemented by backbone RDCs in two independent alignment media, or RDCs combined with the backbone anisotropic chemical shifts in a single alignment medium. A number of such joint structure determinations were performed starting from ca. 2005, with many of the early applications centered on validation – establishing proof that a joint fully flexible NMR/SAXS structure refinement leads to a clear improvement in the structural accuracy, whether by reference to a previously determined crystal structure (Grishaev et al. 2005, 2007), or *via* cross-validation with respect to the experimental RDC data (Grishaev et al. 2008). One of the conclusions that emerged from this work is the need for a set of locally rigid restraints from both NMR experimental data and database or homology constraints that also allow global flexibility of the macromolecule when performing refinement against the solution scattering data. The number of such applications increases at a steady pace, now including flexibly linked proteins and oligonucleotides, as well as IDPs.

13.4 Hybrid Applications of Solution Scattering: Computational Modeling Tools

The software that is capable of using solution scattering data for determination of macromolecular structure can be grouped into three broadly defined classes. The first of these contains standalone packages dealing primarily with SAXS/SANS data that allow very limited input from complementary techniques and rely on simple molecular simulation engines favoring sampling efficiency over force field accuracy, exemplified by EOM (Bernado et al. 2007) or SASREF (Petoukhov and Svergun 2005). The second class of programs adds capability of fitting against SAXS/SANS data to a previously

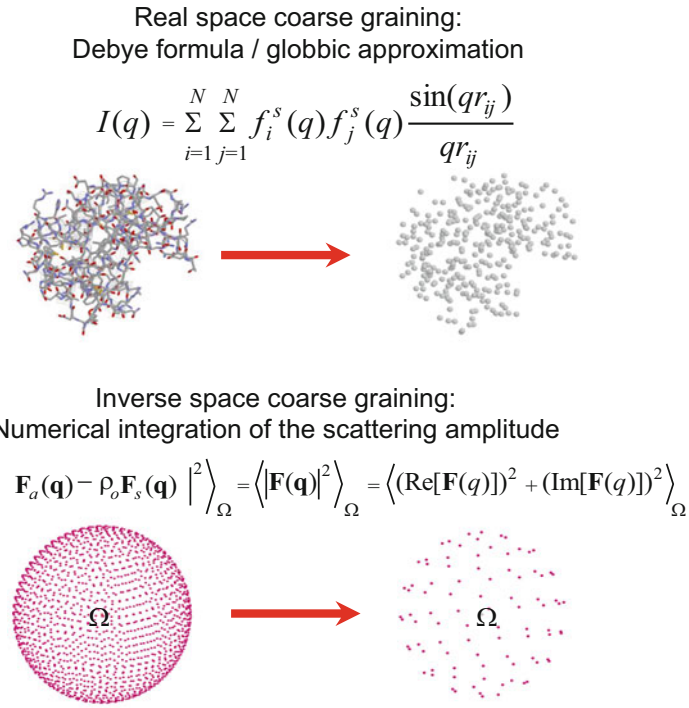

Fig. 13.3 Refinement against solution scattering data in molecular simulation requires application of coarse graining, either in the direct or inverse space, in order to speed up calculation of these data from the macromolecular coordinates

developed package including either a sophisticated molecular simulation core with advanced force fields, or a structure optimization engine with integrative modeling, with examples such as SASSIE (Curtis et al. 2012), IMP (Russel et al. 2012), flexible-Meccano (Ozenne et al. 2012), and RASREC-Rosetta (Rossi et al. 2015). The software belonging to this class would have a somewhat limited ability of handling diverse types of experimental data from complementarity techniques but will be typically capable of including subsets of such data across different techniques in an integrative fashion. Finally, the third class of software adds an ability to refine against SAXS or SANS data to a package whose primary task is macromolecular structure determination against experimental crystal diffraction or NMR data driven by a somewhat simplified molecular simulation engine, with examples including CNS (Brunger et al. 1998) and Xplor-NIH (Schwieters et al. 2003). Therefore, the three classes correspond to emphases on either the scattering data, or the advanced simulation engines and force fields, or effective handling of the complementary experimental restraints from crystallography or NMR. The choice of the class of software would likely depend on the amount of complementary site-specific experimental restraints, with higher restraint density best coupled to the third class of programs. The decision between using the first two classes of software in practice depends on the user's familiarity with the more advanced simulation setup of the second class, or a preference for the greater simplicity of operation generally exhibited by the first class.

When using any software for refining against SAS data, it is important to keep in mind the mode of scattering data calculation, and to be aware of its limitations as the codes coupled to the molecular simulation engines often emphasize the speed of calculation over accuracy, and will likely fall behind dedicated scattering data simulations tools for a single-model analysis listed in Sect. 13.2.2, in terms of the fidelity of data prediction and fit quality. This is an

unavoidable consequence of the generally high computational costs associated with the calculation of the predicted scattering data, and the associated molecular forces along the molecular dynamics trajectory. For example, all of the structure refinement packages that include SAS data use the faster but not necessarily more accurate implicit model for the displaced and surface solvent. Calculation speedup is achieved either by coarse-grading the representation of the macromolecule in the real space *via* globbic approximation to the Debye formula (Grishaev et al. 2005), or by coarse-graining in the inverse space *via* approximations to the spherical averaging of the complex scattering amplitude (Grishaev et al. 2008), illustrated in Fig. 13.3. Recently formulated algorithm for a fast accurate hierarchical approximation to the spherical harmonic expansion of the Debye formula (Berlin et al. 2014) is of particular interest, yet to be implemented in a structure refinement package. In many cases the approximation noted above are a reasonable price to pay for the ability to optimize the structure against SAS data within a molecular dynamics simulation, and to perform joint refinement with restraints from complementary techniques.

13.5 Summary and Conclusions

The past decade has brought a rapid expansion in the number of studies that use solution X-ray and neutron scattering data for derivation of macromolecular structures. Even though the initial methodology developments in bio-SAS focused on the benefits of applying SAXS or SANS data in isolation, an increasing fraction of structural studies involving SAS nowadays use these data in combination with experimental restraints from a growing set of complementary techniques that include crystallography, NMR, FRET or electron microscopy. This fact reflects both universal appreciation of the value of combining data across multiple techniques, and a rapidly expanding repertoire of computational methodologies that allow such combination. Hybrid applications involving SAS are expected to become a norm as structural biology shifts towards studies of challenging architectures that cannot be a analyzed with a single technique, including flexible or dynamic macromolecules, weak macromolecular interactions, or membrane-associated assemblies.

References

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

Berlin K, Gumerov NA, Fushman D, Duraiswami R (2014) Hierarchical O(N) computation of small-angle scattering profiles and their associated derivatives. J Appl Crystallogr 47:755–761

Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle scattering. J Amer Chem Soc 129:5656–5664

Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system (CNS), a new software suite for macromolecular structure determination. Acta Crystallogr D54:905–921

Brunner-Popela J, Glatter O (1997) Small-angle scattering of interacting particles. I. Basic principles of a global evaluation method. J Appl Crystallogr 30:431–442

Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. Biophys J 74:2760–2775

Chen PC, Hub JS (2014) Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. Biophys J 107:435–447

Comoletti D, Grishaev A, Whitten A, Tsigelny I, Taylor P, Trewhella J (2007) Synaptic arrangement of the neuroligin/beta-neurexin complex revealed by X-ray and neutron scattering. Structure 15:693–705

Curtis JE, Raghunandan S, Nanda H, Krueger S (2012) SASSIE: a program to study disordered biological molecules and macromolecular ensembles using experimental scattering restraints. Comput Phys Commun 183:382–389

Durchschlag H (1975) X-ray small-angle studies of pyruvate-dehydrogenase core complex from Escherichia-coli K-12.2. Biophys Struct Mechanism 1:153–188

Fedorov BA, Denesyuk AI (1978) Sperm whale myoglobin structure in solution differs from its structure in crystal by a shift of hairpin GH. FEBS Lett 88:114–117

Grishaev A (2012) Sample preparation, data collection and preliminary data analysis in biomolecular solution X-ray scattering. Curr Protocols Protein Sci 17 (14):1–18

Grishaev A, Wu J, Trewhella Bax A (2005) Refinement of multi-domain protein structures by combination of solution small angle X-ray scattering and NMR data. J Amer Chem Soc 127:16621–16628

Grishaev A, Ying J, Canny M, Pardi A, Bax A (2008) Solution structure of tRNA-Val from joint refinement against dual alignment residual dipolar couplings and SAXS data. J Biomol NMR 42:99–109

Grishaev A, Guo L, Irving T, Bax A (2010) Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. J Amer Chem Soc 132:15484–15486

Grishaev A, Anthis N, Clore GM (2012) Contrast-matched small angle X-ray scattering from a heavy atom-labeled protein in structure determination: application to a lead-substituted calmodulin-peptide complex. J Amer Chem Soc 134:14686–14689

Heidorn DB, Trewhella J (1988) Comparison of the crystal and solution structures of calmodulin and troponin C. Biochemistry 27:909–905

Hickman A, Ewis H, Li X, Knapp J, Laver T, Doss AL, Tolun G, Steven A, Grishaev A, Bax A, Atkinson P, Craig N, Dyda F (2014) Structural basis of hAT transposon end recognition by Hermes, an octameric transposase from Musca domestica. Cell 158:353–367

Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology – expanding the frontier while avoiding the pitfalls. Protein Sci 19:642–657

Johansen D, Trewhella J, Goldenberg DP (2011) Fractal dimensions of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage lambda N protein. Protein Sci 20:1955–1970

Karaka E, Bonvin AMJJ (2013) On usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. Acta Cryst D69:683–694

Koefinger J, Hummer G (2013) Atomic-resolution structural information from scattering experiments on macromolecules in solution. Phys Rev E 87:052712

Lipfert J, Columbus L, Chu VB, Doniach S (2007) Analysis of small-angle X-ray scattering data of protein–detergent complexes by singular value decomposition. J Appl Crystallogr 40:s235–s239

Losonczi JA, Andrec M, Fischer MWF, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 138:334–342

Lu K, Jacob J, Thiyagarajan P, Conticello VP, Lynn DG (2003) Exploiting amyloid fibril lamination for nanotube self-assembly. J Amer Chem Soc 125:6391–6393

Marsh JA, Baker JMR, Tollinger M, Forman-Kay JD (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment. J Amer Chem Soc 130:7804–7805

Mattinen ML, Paakkonen K, Ikonen T, Craven J, Drakenberg T, Serimaa R, Waltho J, Annila A (2002) Quaternary structure built from subunits combining NMR and small-angle X-ray scattering data. Biophys J 83:1177–1183

Mittag T, Marsh JA, Grishaev A, Orlicky S, Lin H, Sicheri F, Kay LE, Tyers M, Forman-Kay JD (2010) Structure/function implications in an intrinsically disordered protein and its dynamic complex. Structure 18:494–506

Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Celine BP, Charavay C, Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470

Park S, Bardhan JP, Roux B, Makowski L (2009) Simulated X-ray scattering of protein solution using explicit solvent models. J Chem Phys 130:134114

Parsons L, Grishaev A, Bax A (2008) The periplasmic domain of TolR forms a large hydrophobic groove and may contact TolQ through its C-terminus: NMR solution structure and its comparison to SAXS data. Biochemistry 47:3131–3142

Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89:1237–1250

Pons C, D'Abramo M, Svergun DI, Orozco M, Bernado P, Fernandez-Recio J (2010) Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. J Mol Biol 403:217–230

Putnam CD, Hammel M, Hura GM, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285

Rossi P, Shi L, Liu GH, Barbieri CM, Lee HW, Grant TD, Luft JR, Xiao R, Actorn TB, Snell EH, Montelione GT, Baker D, Lange OF, Sgourakis NG (2015) A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta. Protein Struct Func Bioinform 83:309–317

Russel D, Lasker K, Webb B, Velazques-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 10: e1001244

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

Schneidman-Duhovny D, Hammel M, Sali A (2011) Macromolecular docking restrained by a small-angle X-ray scattering profile. J Struct Biol 173:461–471

Schwieters CD, Kuszewski J, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:66–74

Schwieters C, Suh JY, Grishaev A, Takayama Y, Clore GM (2010) Solution conformation of the 128 kDa enzyme I dimer from Escherichia Coli and its 146kDa complex with HPr using residual dipolar couplings and small angle x-ray scattering. J Amer Chem Soc 132:13026–13045

Sunnerhagen M, Olah GA, Stenflo J, Forsen S, Drakenberg T, Trewhella J (1996) The relative

orientation of Gla and EGF domains in coagulation factor X is altered by Ca2+ binding to the first EGF domain. A combined NMR small angle X-ray scattering study. Biochemistry 35:11547–11559

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25:495–503

Svergun DI (1999) Restoring low resolution structure of biological macromolecules in solution using simulated annealing. Biophys J 76:2879–2886

Svergun DI, Koch MHJ (2003) Small-angle studies of biological macromolecules in solution. Rep Progr Phys 66:1735–1782

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

Svergun DI, Richard S, Koch MHJ, Sayers Z, Kuprin S, Zaccai G (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc Natl Acad Sci U S A 95:2267–2272

Vendittii V, Schwieters CD, Grishaev A, Clore GM (2015) Dynamic equilibrium between closed and partially closed states of the bacterial enzyme I unveiled by solution NMR and X-ray scattering. Proc Natl Acad Sci U S A 37:11565–11570

Whitten AE, Trewhella J (2009) Small-angle scattering and neutron contrast variation for studying bio-molecular complexes. Micro and nano technologies in bioanalysis: methods and protocols book series. Methods Mol Biol 544:307–323

Zuo X, Tiede DM (2004) Resolving conflicting crystallographic and NMR models for DNA with solution X-ray diffraction. J Amer Chem Soc 127:16–17

Zuo X, Wang J, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX (2008) Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. J Amer Chem Soc 130:3292–3293

Zweckstetter M, Hummer G, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86:3444–3460