

Small Angle X-Ray Scattering as a Complementary Tool for High-Throughput Structural Studies

Thomas D. Grant,¹ Joseph R. Luft,^{1,2} Jennifer R. Wolfley,¹ Hiro Tsuruta,³ Anne Martel,³ Gaetano T. Montelione,^{4,5} Edward H. Snell^{1,2}

¹ Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203

² SUNY Buffalo Department of Structural and Computational Biology, 700 Ellicott St., Buffalo, NY 14203

³ Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025

⁴ Northeast Structural Genomics Consortium, Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854

⁵ Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 088546

Received 11 January 2011; revised 11 March 2011; accepted 14 March 2011

Published online 1 April 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21630

ABSTRACT:

Structural crystallography and nuclear magnetic resonance (NMR) spectroscopy are the predominant techniques for understanding the biological world on a molecular level. Crystallography is constrained by the ability to form a crystal that diffracts well and NMR is constrained to smaller proteins. Although powerful techniques, they leave many soluble, purified structurally uncharacterized protein samples. Small angle X-ray scattering (SAXS) is a solution technique that provides data on the size and multiple conformations of a sample, and can be used to reconstruct a low-resolution molecular envelope of a macromolecule. In this study, SAXS has been used in a high-throughput manner on a subset of 28 proteins, where structural information is available from crystallographic and/or NMR techniques. These crystallographic and NMR structures were used to validate the accuracy of molecular envelopes reconstructed from SAXS data on a statistical level, to

compare and highlight complementary structural information that SAXS provides, and to leverage biological information derived by crystallographers and spectroscopists from their structures. All the *ab initio* molecular envelopes calculated from the SAXS data agree well with the available structural information. SAXS is a powerful albeit low-resolution technique that can provide additional structural information in a high-throughput and complementary manner to improve the functional interpretation of high-resolution structures. © 2011 Wiley Periodicals, Inc. *Biopolymers* 95: 517–530, 2011.

Keywords: SAXS; high-throughput; structural biology; NMR; crystallography

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of the preprint by emailing the *Biopolymers* editorial office at biopolymers@wiley.com

INTRODUCTION

Structural biology aims to understand life on an atomic scale by using structural information to discern a molecule’s functional attributes. To date more than 60,000 macromolecular structures have been deposited in the Protein Data Bank (PDB).¹ Of these,

Correspondence to: E. H. Snell; e-mail: esnell@hwi.buffalo.edu

* Present address: Institut de Biologie Structurale J.P. Ebel, 41 rue Jules Horowitz, F-38027 Grenoble Cedex 1, France.

© 2011 Wiley Periodicals, Inc.

~86% were determined with single crystal X-ray diffraction methods. The importance of this method is reflected in the number of Nobel prizes that have been associated with it including the determination of the structure of DNA,² the structure of vitamin B12,³ the structure of the photosynthetic reaction center,⁴ the enzymatic mechanism underlying the synthesis of adenosine triphosphate,⁵ the structure of potassium channels,⁶ the molecular basis of eukaryotic transcription,⁷ and most recently the ribosome.⁸ Unfortunately, most proteins do not readily produce diffraction-quality crystals. Where failure as well as success has been rigorously tracked only 34% of expressed and purified targets provide a crystal and out of that only 12% result in a structure deposited in the PDB.⁹ Crystallographic structures require crystals, whereas crystallization remains fundamentally a hit-or-miss proposition.

The Hauptman-Woodward Medical Research Institute provides a high-throughput crystallization screening service to the structural genomics and biological crystallography community. Macromolecular samples are screened against 1536 chemically diverse cocktails¹⁰ using the microbatch-under-oil technique.¹¹ This service has been in operation for more than 10 years and to date has screened 12,500 proteins for more than 1000 laboratories worldwide. The screening laboratory has worked in close collaboration with the Northeast Structural Genomics (NESG) consortium, screening their samples for crystallization leads. In this effort, ~50% of soluble proteins that enter the screening laboratory provide promising crystallization lead conditions; ~45% of these have been successfully optimized by NESG resulting in a PDB deposition. Although this success rate is relatively good in the structural genomics field (providing evidence of good initial sample preparation and crystallization methods), this means that 78% of the soluble, purified proteins do not result in crystallographic structures. Nuclear magnetic resonance (NMR) techniques can provide structural information for samples recalcitrant to crystallization. In the NESG case, ~44% of the structural depositions result from NMR methods. For the Protein Structure Initiative (PSI) as a whole, only ~14% of the soluble purified targets make it to a PDB deposition, 21% of which were determined by NMR. To put this into perspective, there are greater than 30,000 soluble, purified samples from the US PSI that failed to provide structures; this number is almost half of the current structural information in the PDB. Even low-resolution structural information from these samples would significantly enhance the understanding of the biological world.

Small angle X-ray scattering (SAXS) is a technique that can provide low-resolution structural information, a molecular envelope from a solution of the sample;¹² a crystal is not

required. In this study, SAXS has been used in a high-throughput manner as a technique that is complementary to crystallography and NMR. The remainder of solutions from samples provided by NESG for high-throughput crystallization screening have been used for SAXS analysis. To date, this has been carried out for more than 500 samples. In this article, the information obtained from SAXS on a subset of 28 samples where either crystallographic, NMR, or a combination of both of these structures are available is described. We demonstrate how information from SAXS can complement and enhance high-resolution structural information. On the basis of these observations, it is proposed that SAXS should be adopted as a routine complementary analysis technique in structural biology that can be used to resolve and improve the interpretation of biological function from structural information.

MATERIALS AND METHODS

Samples

In this study, complementary SAXS data were collected from 28 different protein samples, where structural information is available through crystallography, NMR, or a combination of both techniques. The protein samples are NESG targets that represent large protein domain families, biomedical themes, and targets nominated by the biomedical community. The NESG biomedical themes focus on eukaryotic proteins, particularly human proteins involved in cancer biology, protein-protein interaction networks, specific biochemical pathways, and proteins implicated in other human diseases. The protocols for selection, cloning, expression, purification, and crystallization of each sample are described elsewhere.^{13–15} After purification, each sample is concentrated to between 5 and 10 mg/ml using an Amicon (Millipore, Billerica, MA) centrifugal filtration unit with a 5 kDa molecular weight cutoff membrane. SDS-PAGE and mass spectrometry analysis are used to confirm purity and molecular weight, respectively. Analytical gel filtration with static light scattering detection is used to screen aggregation and determine the oligomeric state of each sample.

These 28 protein samples are summarized in Table I and can be divided into four sets. The first set encompasses 13 proteins, where a crystallographic structure is available. These range in molecular weight from 9.5 kDa to 48.5 kDa. The second set consists of two proteins, where two constructs were studied for each protein target. Two crystallographic structures were available from different constructs for the first and a single crystallographic structure for the second. The third set consists of nine proteins for which there is an NMR structure; the fourth set includes two protein targets, where both NMR and crystallographic structures are available.

There are a high percentage of crystallographic structures that have residues missing in the coordinates deposited in the PDB, compared with the total number of residues in the protein sequence. Indeed, it is estimated that conformational flexibility results in unstructured regions of 40 amino acids or more long in 50% of eukaryotic proteins.⁴⁵ Although some efforts were made in construct design to eliminate large disordered N- and C-terminal segments,⁴²

Table I Samples Used for the SAXS Analysis are Divided into Four Sets. The First Set 1–13 Contains 13 Proteins, Each Having Crystallographic Structures. The Second Set 14–17 Contains 2 Proteins with Two Different Constructs of the First Having Two Crystallographic Structures and the Second a Single Structure. The Third Set 18–26 Contains 9 Proteins, Each Having an NMR Structure. The Fourth Set 27–28 Contains Two Proteins Where Both NMR and Crystallographic Structures are Available. The Sample Name, ID, PDB Identifier, Reference, the Oligomeric State in Solution Characterized on Preparation by Light Scattering and Gel Filtration, Initial Concentration (mg/ml), Molecular Weight (Da) and Number of Residues are Listed. The Oligomeric State in Solution is Defined in the Table as M (Monomer), D (Dimer), Tri (Trimer), T (Tetramer), Hep (Heptamer) or a Combination. While All the Samples Have Structures Deposited in the PDB the Majority are as Yet Unpublished. We are Grateful to the Authors in the References for the Ability to Use This Structural Data at This Early Stage

| # | Name | NESG ID | PDB | Ref | State | Conc | MW | Res |
|--|--|------------------|------------|-----|-------|------|--------|-----|
| Samples where crystallographic structures were available | | | | | | | | |
| 1 | Domain of unknown function | DhR2A | 3HZ7 | 16 | M | 6.9 | 9523 | 87 |
| 2 | Diguanylate cyclase with PAS/PAC sensor | MqR66C | 3H9W | 17 | D | 8.2 | 13,611 | 210 |
| 3 | Nmul_A1745 protein from <i>Nitrosospira multiformis</i> | NmR72 | 3LMF | 18 | T | 6.9 | 14,069 | 484 |
| 4 | Domain of unknown function | DhR85C | 3MJQ | 19 | D | 10.7 | 14,609 | 252 |
| 5 | Sensory box/GGDEF family protein | SoR288B | 3MFX | 20 | D | 9.1 | 14,779 | 258 |
| 6 | MucBP domain of the adhesion protein PEPE_0118 | PtR41A | 3LYY | 21 | M | 9.5 | 14,300 | 131 |
| 7 | Sensory box/GGDEF domain protein | CsR222B | 3LYX | 22 | D | 12.7 | 15,341 | 248 |
| 8 | HIT family hydrolase | VfR176 | 3I24 | 23 | D | 11.0 | 17,089 | 298 |
| 9 | EAL/GGDEF domain protein | McR174C | 3ICL | 24 | M | 5.0 | 18,738 | 171 |
| 10 | Diguanylate cyclase | MqR89A | 3IGN | 25 | M | 7.5 | 20,256 | 177 |
| 11 | Putative NADPH-quinone reductase | PtR24A | 3HA2 | 26 | D | 9.5 | 20,509 | 354 |
| 12 | MmoQ (response regulator) | McR175G | 3LJX | 27 | M | 8.8 | 32,032 | 288 |
| 13 | Putative uncharacterized protein | DhR18 | 3HXL | 28 | M | 9.6 | 48,519 | 446 |
| Samples where multiple constructs and crystallographic structures were available | | | | | | | | |
| 14 | Putative hydrogenase | PfR246A (78–226) | 3LRX | 29 | D | 11.4 | 17,701 | 316 |
| 15 | | PfR246A (83–218) | 3LYU | 30 | D | 8.4 | 16,321 | 284 |
| 16 | Alr3790 protein | NsR437I | 3HIX | 31 | M | 5.3 | 11,760 | 105 |
| 17 | | NsR437H | 3HIX | 31 | M | 6.5 | 15,700 | 141 |
| Samples where NMR structures were available | | | | | | | | |
| 18 | MKL/myocardinlike protein 1 | HR4547E | 2KW9 (NMR) | 32 | D | 10.4 | 8276 | 75 |
| 19 | MKL/myocardinlike protein 1 | HR4547E | 2KVU (NMR) | 33 | D | 10.4 | 8276 | 75 |
| 20 | Putative peptidoglycan bound protein (LPXTG motif) | LmR64B | 2KVZ (NMR) | 34 | M | 5.0 | 9712 | 85 |
| 21 | E3 ubiquitin-protein ligase Praj1 | HR4710B | 2L0B (NMR) | 35 | M/D | 5.6 | 10,297 | 91 |
| 22 | Transcription factor NF-E2 45 kDa subunit | HR4653B | 2KZ5 (NMR) | 36 | M | 10.0 | 10,623 | 91 |
| 23 | YlbL protein | GtR34C | 2KL1 (NMR) | 37 | M | 11.0 | 10,661 | 94 |
| 24 | Cell surface protein | MvR254A | 2L0D (NMR) | 38 | Tri | 5.9 | 12,385 | 114 |
| 25 | Domain of unknown function | MaR143A | 2KZW (NMR) | 39 | M | 6.6 | 16,312 | 145 |
| 26 | N-terminal domain of protein PG_0361 from <i>P. gingivalis</i> | PgR37A | 2KW7 (NMR) | 40 | M | 12.9 | 17,485 | 157 |
| Samples where both crystallographic and NMR structures were available | | | | | | | | |
| 27 | GTP pyrophosphokinase | CtR148A | 2KO1 (NMR) | 41 | D | 8.0 | 10,042 | 176 |
| | | | 3IBW | 42 | T | 8.0 | 10,042 | 176 |
| 28 | Lin0431 protein | LkR112 | 2KPP (NMR) | 43 | M/Hep | 6.3 | 12,747 | 114 |
| | | | 3LD7 | 44 | M | 6.3 | 12,747 | 100 |

in many cases, disordered ends and disordered internal loops are observed in these protein structures. As dynamics and conformational changes are crucial for the function of many macromolecular complexes and enzymes,⁴⁶ even low-resolution information about conformational distributions of these residues is useful.

Crystallization

Each sample (450 μ l at \sim 5–13 mg/ml concentration) was shipped to the Hauptman-Woodward Medical Research Institute's high-throughput screening laboratory on dry ice, thawed on arrival (typically within 1 day of receipt), and set up in 1536 crystallization plates.¹⁰ Each of the 1536 experiments was imaged immediately af-

ter the sample was added, and then in weekly intervals for 6 weeks (as well as a control imaging before the sample was added to the cocktails). For all the NESG samples, each image was manually inspected and classified as "crystal" or "no crystal." These classifications and the images were then communicated to NESG scientists for crystallization optimization and structural data collection.

X-Ray Crystallographic and Solution NMR Structure Determination

The crystallographic and NMR structures used in this study were all solved by NESG staff scientists, and have been deposited in the

PDB.¹ In cases where a article on the structure is not currently available, the Digital Object Identifier (DOI) for the PDB deposition is provided in the reference list accompanied by the authors involved. The crystallographic asymmetric unit is not necessarily the biological oligomer. This oligomer was predicted using a theoretical analysis of binding energy and entropy of dissociation with the Protein Interfaces, Surfaces and Assemblies (PISA) service at the European Bioinformatics Institute.⁴⁷

SAXS Data

SAXS data were collected at beamline 4-2⁴⁸ of the Stanford Synchrotron Radiation Lightsource (SSRL). The SAXS preparations used samples that were refrozen after crystallization screening; all the SAXS samples underwent two freeze/thaw cycles. Typically, a minimum sample volume of 60 μl was used. The sample was diluted with its matching sample buffer to prepare three solutions of known concentrations. At the beamline, an automated sample loader (manuscript in preparation), compatible with the polymer chain reaction (PCR) tubes, was used to collect data on as many as 96 experiments without user intervention, or the need to open the hutch. A wavelength of 1.3 Å was used for eight consecutive two-second exposures collected at each of the three sample concentrations. Each sample was oscillated back and forth in a quartz capillary cell during data collection to minimize radiation damage effects. All these samples had an identical matching buffer. The samples were loaded in eight-well PCR strips such that a buffer blank was recorded followed by three concentrations of each of the two samples and then a final buffer blank with a wash cycle between each. The original concentration was diluted to 2:1, 1:2, and 1:5 ratios of sample and buffer blank. Using the 96-well capacity of the beamline sample loader, a series of 24 proteins were studied in each automated run. Typical time for a single sample concentration series was ~ 15 min with the majority of that time spent on liquid handling, for example, sample loading and washing the fluid apparatus between each concentration. The data were processed and azimuthally integrated with SASTool (manuscript in preparation) and then visually examined with Primus.⁴⁹ Each of the eight exposures was compared for similarity to ensure no radiation damage took place and was averaged using SASTool to increase the signal to noise ratio. The SAXS data for different protein concentrations were assessed with Kratky plots and screened for aggregation using Guinier plots.⁵⁰ Guinier regions and radius of gyration (R_g) estimates were derived by the Guinier approximation $I(q) = I(0)\exp(-q^2 R_g^2/3)$ with $qR_g < 1.3$ using the AutoRg function of Primus where $q = 4\pi \sin \theta/\lambda$. The highest quality estimate as determined by AutoRg was used to select which of the three concentrations would go on to further processing. Zero extrapolated curves were not used because the examination of the concentration series showed no evidence of aggregation or repulsion in the higher concentration, stronger signal data. AutoGNOM⁵¹ was used to compute the pair distribution functions, $P(R)$, for each sample and to determine the maximum particle dimension, D_{max} , and these values were compared with those determined manually by GNOM to ensure consistency. A molecular weight was estimated from the program AutoPOROD of the ATSAS package.⁵² Five ab initio shape reconstructions (molecular envelopes) were generated by DAMMIF⁵³ and averaged with DAMAVER.⁵⁴ The program CRY SOL⁵⁵ was used to calculate the scattering intensity from deposited crystallographic and NMR structures and estimate an R_g and fit the data by minimizing the discrepancy, χ , according to

$$\chi^2(r_0, \delta_\rho) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(q_i) - cI_c(q_i, r_0, \delta_\rho)}{\sigma(q_i)} \right]^2$$

where I_e is the experimental scattering, I_c is the calculated scattering, and σ is the experimental error as determined by SASTool.⁴⁸ Other variables are given elsewhere.⁵⁵ In our case, the experimental errors are underestimated, as the detector is treated as an ideal photon counter; χ values here should therefore be regarded as a relative indicator of goodness of fit. Volume fractions for cases of oligomeric mixtures were estimated using the program OLIGOMER.⁴⁹ In these cases, estimates for χ and the convoluted R_g of the mixture in solution were taken from OLIGOMER.

Comparison of Structural Data

R_g and D_{max} were calculated from the crystallographic and NMR structures using the program CRY SOL.⁵⁵ These values were compared with those derived from the SAXS data with constant subtraction enabled. For visualization purposes, envelopes produced by the SAXS data were automatically overlaid with structures derived from X-ray crystallographic and NMR techniques using the program SUPCOMB, and were followed by manual adjustments using the program PyMOL.

RESULTS

Table II summarizes and compares data calculated from crystallographic and NMR structural information with that measured from the SAXS data. In general, experimentally determined R_g values are consistent with those calculated from the structural information. Usually, the SAXS calculated D_{max} are somewhat larger than those for the crystallographic cases, but smaller than those for the NMR structures. This might be expected because of missing residues in the crystallographic case. On the other hand, these NMR structure ensembles may overestimate the breadth of the true conformational distribution, as the set of 20 conformers deposited in the PDB does not account for the population distribution across the ensemble. The Prod calculated molecular weights are, again for the most part, integer multiples of the measured molecular weight. The SAXS determined oligomer is shown along with the relationship to that seen in the crystallographic structure. The SAXS data are recorded under the conditions diluted from the initial sample preparation and the crystallographic structures are necessarily determined under different biochemical conditions. Details for each group, and in particular deviations from the known crystallographic structure, are described below. The observed data and structural fit to the observed data (continuous line) are shown with the structures and ab initio envelopes calculated for each group in Figures 1–4 and for those where a mixture was observed (Figure 5). In the majority of the cases, the fit to the experimental data is good.

Table II A Summary of Structural (Crystallography and NMR) and SAXS Results. The Sample # Refers to the Identical Number in Table I. The Number of Unresolved Residues in the Structure (Mainly Crystallographic) is Listed Together with the R_g and D_{max} (in Å) Determined from the Available Structure. The R_g and D_{max} from the SAXS Data are Shown Together with the Difference from the Available Structural Information. The Molecular Weight (in Da) Calculated from a Porod Analysis is Listed Along with the Ratio of this Weight with that Derived Initially from Mass Spectrometry in Table I. Finally the SAXS Determined Oligomer, (Monomer, Dimer or Tetramer), the Relationship to the Available Structure and the χ of the Fit are Listed. A Special Case is Described Below for Samples 16 and 17. Further Details are Given in the Text

| # | Residues Observed | # Res Missing | R_g Structure | D_{max} Structure | R_g SAXS | ΔR_g | D_{max} SAXS | Δd_{max} | Porod MW | MW Ratio | SAXS Oligomer | Oligomer Assign. | SAXS Fit (χ) |
|--|-------------------|---------------|-----------------|---------------------|------------|--------------|----------------|------------------|----------|----------|------------------|------------------|---------------------|
| Samples where crystallographic structures were available | | | | | | | | | | | | | |
| 1 | 74 | 13 | 13.7 | 42.0 | 14.9 | 1.2 | 53.2 | 11.2 | 7827 | 0.8 | M | | 4.2 |
| 2 | 198 | 12 | 16.6 | 67.0 | 19.8 | 3.2 | 67.4 | 0.4 | 24,555 | 1.8 | D | sym | 2.6 |
| 3 | 436 | 48 | 22.4 | 62.3 | 23.2 | 0.8 | 75.3 | 13.0 | 50,064 | 3.6 | T | sym | 1.6 |
| 4 | 214 | 38 | 23.3 | 81.2 | 23.6 | 0.3 | 82.7 | 1.5 | 37,348 | 2.6 | D/T ^a | PDB | 2.6 |
| 5 | 224 | 34 | 19.9 | 57.6 | 19.8 | -0.1 | 64.2 | 6.6 | 28,828 | 2.0 | D | PDB | 2.2 |
| 6 | 107 | 24 | 19.6 | 76.3 | 21.5 | 1.9 | 82.0 | 5.7 | 11,085 | 0.8 | M | | 6.1 |
| 7 | 236 | 12 | 21.4 | 64.7 | 22.2 | 0.8 | 76.8 | 12.1 | 31,410 | 2.0 | D | PDB | 3.8 |
| 8 | 286 | 12 | 20.5 | 63.1 | 21.1 | 0.6 | 71.4 | 8.3 | 34,786 | 2.0 | D | PDB | 2.0 |
| 9 | 162 | 9 | 17.6 | 54.0 | 18.7 | 1.1 | 65.5 | 11.5 | 20,468 | 1.1 | M | | 3.7 |
| 10 | 165 | 12 | 17.5 | 58.0 | 18.5 | 1.0 | 65.8 | 7.8 | 19,069 | 0.9 | M | | 4.2 |
| 11 | 336 | 18 | 26.1 | 80.8 | 26.0 | -0.1 | 89.7 | 8.9 | 59,937 | 2.9 | D/T ^a | PDB/sym | 1.4 |
| 12 | 252 | 36 | 21.3 | 61.5 | 22.5 | 1.2 | 81.9 | 20.4 | 37,254 | 1.2 | M | | 2.9 |
| 13 | 416 | 30 | 28.5 | 95.0 | 27.6 | -0.9 | 98.5 | 3.5 | 40,027 | 0.8 | M | | 1.4 |
| Samples where multiple constructs and crystallographic structures were available | | | | | | | | | | | | | |
| 14 | 272 | 44 | 20.8 | 59.6 | 21.1 | 0.3 | 69.2 | 9.6 | 30,670 | 1.9 | D | PDB | 1.9 |
| 15 | 258 | 26 | 21.1 | 61.8 | 22.0 | 0.9 | 79.7 | 17.9 | 32,657 | 2.0 | D | PDB | 1.8 |
| 16 | 93 | 12 | 18.0 | 59.5 | 18.2 | 0.2 | 64.7 | 5.2 | 15,875 | 1.3 | D2 | PDB | 1.7 |
| 17 | 93 | 48 | 20.4 | 75.0 | 20.8 | 0.4 | 73.0 | -2.0 | 15,920 | 1.0 | D1 | PDB | 2.5 |
| Samples where NMR structures were available | | | | | | | | | | | | | |
| 18 | 75 | 0 | 22.5 | 122.4 | 16.8 | -0.9 | 58.4 | -64.0 | 6771 | 0.8 | M | | 4.7 |
| 19 | 75 | 0 | 17.7 | 94.4 | 16.5 | -1.2 | 58.4 | -36.0 | 6771 | 0.8 | M | | 1.4 |
| 20 | 85 | 0 | 19.0 | 80.8 | 18.7 | -0.3 | 68.0 | -12.8 | 9724 | 1.0 | M | | 1.7 |
| 21 | 91 | 0 | 16.4 | 71.0 | 15.9 | -0.5 | 59.6 | -11.4 | 7862 | 0.8 | M | | 1.5 |
| 22 | 91 | 0 | 22.3 | 123.1 | 19.6 | -2.7 | 68.0 | -55.1 | 10,762 | 1.0 | M | | 1.6 |
| 23 | 87 | 7 | 14.3 | 55.8 | 14.5 | 0.2 | 49.7 | -6.1 | 8479 | 0.8 | M | | 1.4 |
| 24 | 114 | 0 | 16.5 | 67.8 | 19.6 | 3.1 | 66.6 | -1.2 | 12,609 | 1.0 | M | | 5.9 |
| 25 | 145 | 0 | 49.0 | 325.5 | 26.6 | -22.4 | 94.7 | -230.8 | 15,386 | 0.9 | M | | 7.4 |
| 26 | 157 | 0 | 19.8 | 67.5 | 17.5 | -2.3 | 60.6 | -6.9 | 15,238 | 0.9 | M | | 2.1 |
| Samples where both crystallographic and NMR structures were available | | | | | | | | | | | | | |
| 27 ^a | 176 | 0 | 18.0 | 66.7 | 19.1 | 1.1 | 68.3 | 1.6 | 22,589 | 2.2 | D | PDB | 2.5 |
| | 158 | 18 | 18.1 | 52.5 | 19.0 | 0.9 | 68.3 | 15.8 | | | | PDB | 2.4 |
| 28 ^a | 114 | 0 | 18.5 | 104.4 | 18.5 | 0.0 | 68.2 | -36.2 | 10,721 | 0.8 | M | | 2.3 |
| | 87 | 13 | 14.8 | 44.1 | 18.4 | 3.6 | 68.2 | 24.1 | | | | | 7.4 |

^a The SAXS data compared with the NMR structure is listed above with the crystallographic comparison as in Table I.

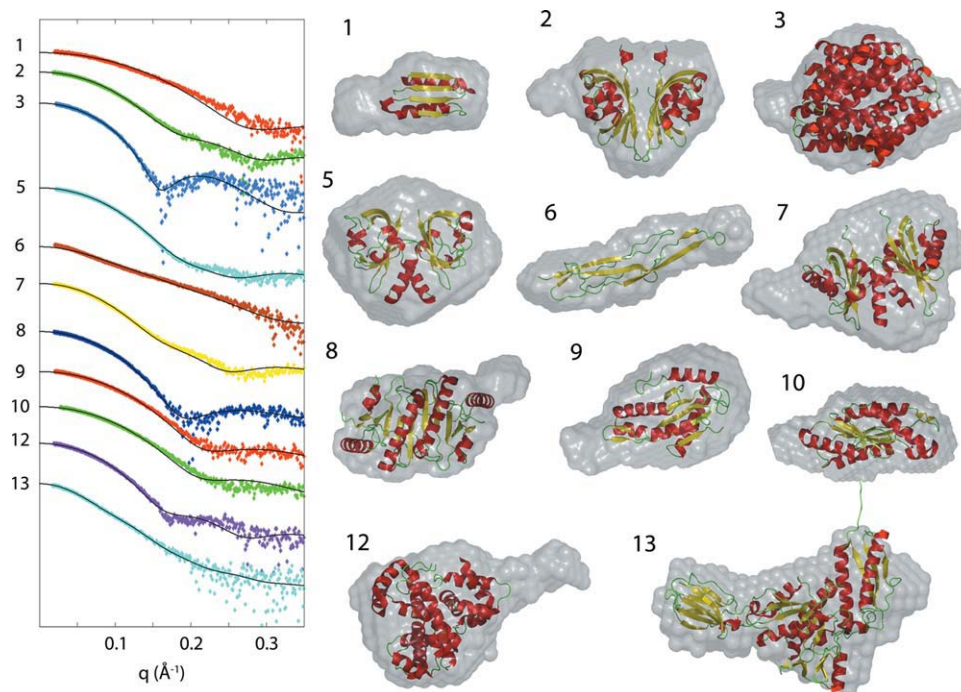


FIGURE 1 The observed SAXS data and structural fit to the observed data (continuous line) for samples with crystallographic structure. The ab initio SAXS-derived envelopes overlaid with crystallographic structure are also shown illustrating the agreement between the techniques. The samples are numbered as in Tables I and II. Samples 4 and 11 contained a mixture of oligomers and are shown below.

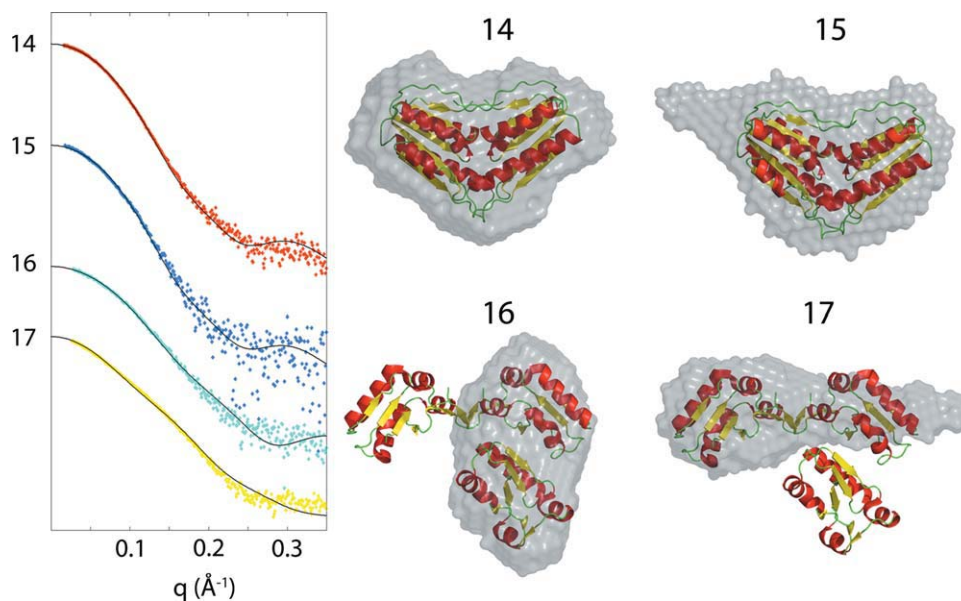


FIGURE 2 The observed SAXS data and structural fit to the observed data (continuous line) for samples with crystallographic structure and multiple constructs. Ab initio SAXS-derived envelopes are overlaid with crystallographic structure. Sample 17 contained a mixture shown below. The figures are shown to the same approximate scale as those in Figure 1 and the remaining structural representations in Figures 3 and 4.

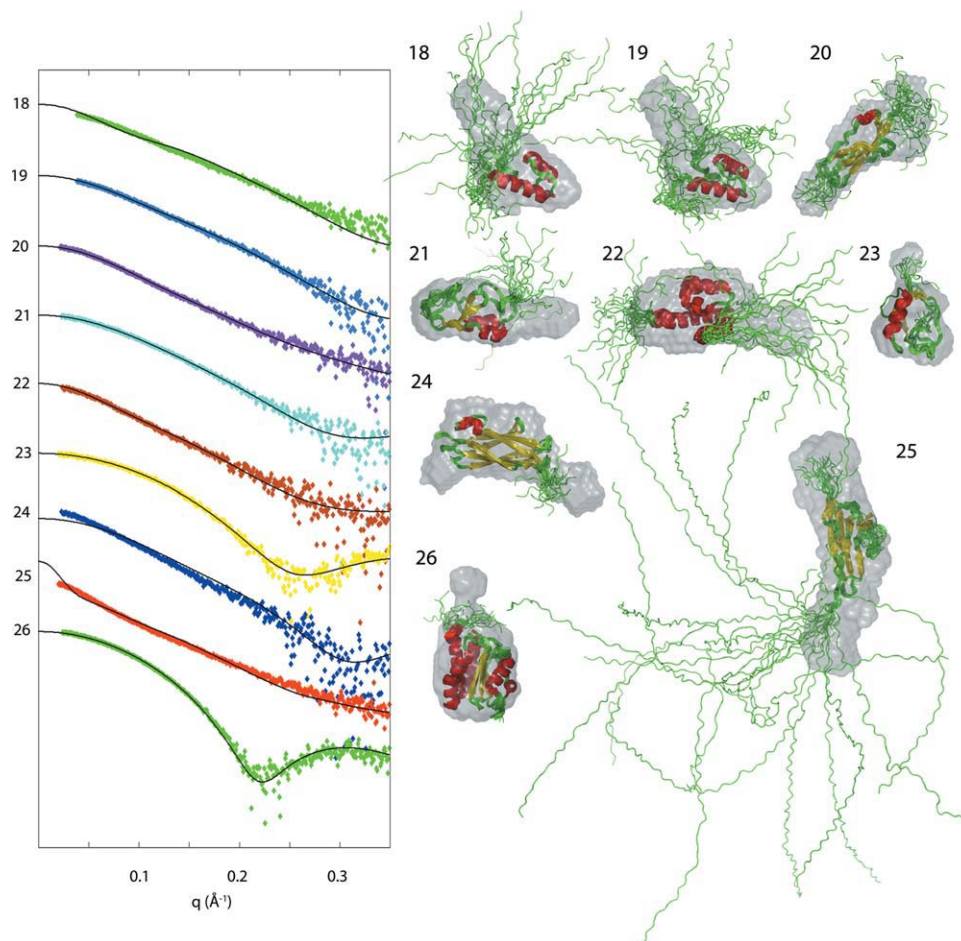


FIGURE 3 The observed SAXS data and structural fit to the observed data (continuous line) for samples with NMR structures. The ab initio SAXS-derived envelopes are overlaid with the NMR structures to show the agreement between the data. The figures are shown to approximate scale as in the other structural illustrations and illustrate multiple conformations determined from the NMR data.

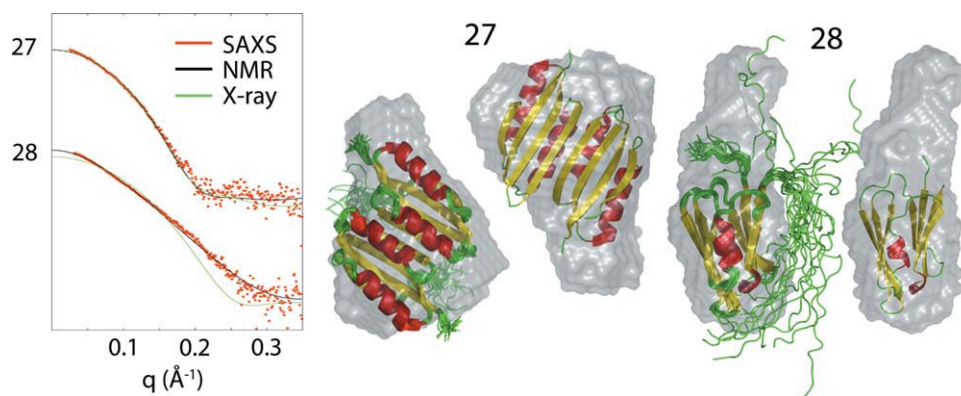


FIGURE 4 Ab initio SAXS-derived envelopes overlaid with NMR and crystallographic structure to show the agreement between the different structural methods. The figures are shown to approximate scale and illustrate multiple conformations determined from the NMR data.

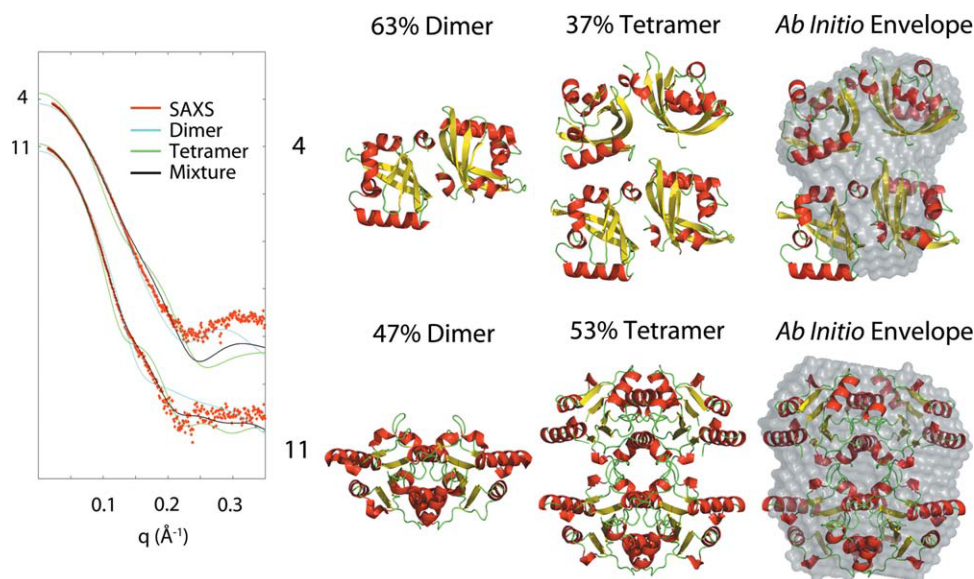


FIGURE 5 Structures of oligomers based on analysis of the SAXS data and known monomer structure. The ab initio SAXS-derived envelopes are shown assuming a monodisperse solution.

Crystallographic and SAXS Comparison

For the 13 samples in the set of crystallographically determined structures, there was relatively good agreement between the R_g of the model and that calculated from the SAXS data with an average deviation of $<1 \text{ Å}$ (Table II). The difference in D_{\max} between the crystallographic and SAXS envelope is greater, having no correlation with the percentages of missing residues in the crystallographic structure. This is not surprising given that missing residues may contribute to the D_{\max} if missing from the longest axis or may have little to no contribution if predominately missing from a shorter axis.

The observed data and structural fit to the observed data (continuous line) with crystallographic structures and ab initio molecular envelopes are shown in Figure 1 with the exception of Samples 4 and 11 where a mixture of oligomers is indicated (see below). Outliers clearly visible by eye occur in Samples 1 and 6; however, there is a good correlation with all the envelopes and the known structure. The known crystallographic structure is represented in a ribbon form for clarity but in reality occupies more space when side chains are taken into account. For many cases, the molecular envelope clearly extends beyond the known structure, extending further than can be explained by side chains on the backbone. These instances are consistently located in areas with residues missing from the crystallographic structure, but could also be attributed to slight undetected aggregation artificially enhancing the calculated D_{\max} . The highest χ values are observed for Samples 6 ($\chi = 6.1$), 1 ($\chi = 4.2$), and 10 ($\chi = 4.2$). The crystallographic structures for these samples are

missing 22%, 18%, and 7% of the residues, respectively, which could contribute to this, although better χ values are observed for Samples 4 and 12, which are both missing 14% of their residues in the crystallographic structures. In Sample 1, a total of 13 residues are unresolved in the crystallographic structure. The molecular envelope reconstruction suggests evidence of these on the left-hand side of the envelope. Sample 2 is a dimer in solution and 12 residues are unresolved in the crystallographic structure. When reconstructing the molecular envelope, it is possible to add known symmetry information to the reconstruction and averaging (as in an oligomer), but in our case, ab initio modeling and averaging without symmetry constraints were used. A similar effect is seen for Sample 7, where 12 residues were missing from the crystallographic structure. The molecular envelope accounts for missing residues in Samples 8, 9, and 12, with 12, 9, and 36 unresolved residues, respectively. In each case, the portion of envelope unexplained by the available crystallographic structure is positioned adjacent to the point where residues become unresolved in the structure. In Sample 5, there are 34 residues missing, but in this case, it is not clear where those residues reside. These samples are structurally diverse yet in all the cases, the molecular envelopes show good agreement (at the resolution of the technique) with the known structures.

Where the molecular weight calculated from the Porod volume indicated an oligomer, different oligomers were compared with the experimental scattering. In Table II, the oligomer assignment is noted as either “PDB,” where the chosen oligomer is present within the asymmetric structure

provided by the PDB, or as “sym” where the oligomer in solution is not present in the PDB, but is chosen based on the crystal symmetry operator. In two cases, the oligomer seen in solution using SAXS was not the oligomer indicated by the asymmetric unit in the PDB, but was a smaller unit present within it. In Case 5, the PDB contains a trimer, whereas the SAXS data not only favored a dimer, but was also able to clearly distinguish which dimer from the two possibilities. In Case 9, the PDB oligomer is a dimer, whereas the SAXS selects the monomer. In both of these cases, the SAXS-selected oligomer agrees with the oligomer observed by gel filtration. For the majority of cases, with the exceptions of Samples 6 and 9, the PISA prediction was in good agreement with the SAXS-derived oligomer. In Sample 6, PISA predicted an elongated dimer and in Sample 9, three dimers in solution were predicted. Neither of these cases were supported by the SAXS data. In certain cases, no oligomer provided by the PDB or via symmetry operation was found to be consistent with the SAXS data when comparing the R_g , D_{max} , and overall fit to the curve. In this study, as the atomic structure is already known, the results can be analyzed as a mixture of oligomers. Samples 4 and 11 showed clear evidence of oligomer mixtures with Sample 4 consisting of 63% dimer and 37% tetramer and Sample 11 consisting of 47% dimer and 53% tetramer. These are discussed below.

Sensitivity to Different Constructs

Alternate constructs were available for two samples. The first sample, a putative hydrogenase, had crystallographic structures for both constructs, where SAXS data was collected, Samples 14 and 15 with 316 and 290 residues, respectively. Interestingly, for Sample 15, the construct with fewer residues, a significantly larger D_{max} (79.7 Å) for the construct with fewer residues, compared with the D_{max} (69.2 Å) for the construct with more residues. The corresponding crystallographic structure for Sample 15 shows a dimer in the PDB, where one monomer has two fewer residues in the electron density than the adjacent monomer. In the adjacent monomer, these two residues appear to form a beta strand secondary structural element, whereas in the adjacent monomer, lacking these two residues, this element is not present. This may reflect a higher level of disorder for these and neighboring residues in solution and subsequently for the five additional residues absent from this terminus. The SAXS envelope for this sample fits well to the overall crystallographic structure with the exception of an additional region present on only one side of the dimer. The disordered residues present in one monomer may be occupying this area. However, SAXS is a technique sensitive to aggregation and this extension of the SAXS envelope by an additional 10 Å compared

with the similar construct may result from minor levels of aggregation present in solution that has escaped detection via static light scattering and Guinier analysis of multiple concentrations. Without further data, it is not possible to distinguish the source of this difference.

The second example of multiple constructs, the protein Alr3790, has a single crystallographic structure (PDB id 3HIX), the Alr3790 protein, for the two constructs, whereas SAXS data for each construct, Samples 16 and 17, are clearly different. These constructs had 105 residues (providing the crystallographic structure) and 141 residues, respectively (out of 151 in the protein). The 3HIX structural model shows a trimer in the asymmetric unit. This trimer did not fit the SAXS data for either construct. Breaking the trimer into two separate dimers, D1 and D2, showed that each construct forms a structurally distinct dimer in solution. Sample 16 contains 36 fewer residues than Sample 17 and these extra residues are located precisely at the D1 dimer interface. A possible explanation for the two solution states is that these extra residues impede D1 dimer formation in Sample 17, but not being present in Sample 16, allow the formation of dimer D2. The comparison of the calculated scattering for each possible dimer configuration with the experimental SAXS data clearly distinguishes the correct dimer formation for each construct. The envelope for Sample 17 appears to underestimate the volume of the entire D1 dimer. Given that analytical gel filtration data and Porod molecular weight indicated a monomer in solution, possibly the monomer form may have a significant population in solution. A mixture analysis using both monomer and dimer only gave marginal improvements to the fit ($\chi = 2.1$ – 2.5 , respectively), and no improvement to the size parameters (data not shown). If a monomer population is also present at a low concentration, it does not appear to greatly affect the SAXS curve. The PISA analysis predicted a stable hexamer consisting of a trimer of dimers for Samples 14 and 15. Although the hexamer is not shown to exist from the SAXS data, the dimer is present. For Samples 16 and 17, the PISA analysis predicts both dimers to be equally stable. The SAXS data for Sample 16 shows one dimer, whereas for Sample 17 the SAXS data shows the other. For both examples, the observed SAXS data, structural fit to the observed data, and the envelopes with overlaid crystallographic structure are shown in Figure 2. Again the globular region of these proteins is well represented by the SAXS-derived ab initio molecular envelope.

NMR and SAXS Comparison

The SAXS D_{max} was consistently smaller than that derived from the corresponding solution NMR structure. In each case, the NMR structural data consists of the 20 lowest energy conformers from 100 that were calculated. The D_{max}

in the NMR case is calculated from the maximum dimension of the total envelope of all 20 conformers. As such, it can lead to an overestimate of D_{\max} as it measures the extremes within this set of conformers and does not take into account relative populations of these conformers or their dynamics. Although it is possible to obtain such population distributions from NMR studies, this information is not available from these NMR structural ensembles. For all the samples except Sample 25, the calculated and measured R_g 's are similar. The R_g is defined as the root mean square distance of the atoms in the molecule from their common center of gravity. As such, it is less sensitive to extremes within the population of conformers derived from the NMR data. Of note are Samples 18 and 19, where the same SAXS data is compared against two NMR structures. The first, Sample 18, was compared with a structure with no residual dipolar coupling information and the second, Sample 19, was compared with a structure making use of residual dipolar coupling in the refinement process. Although these two NMR structures are similar, with backbone root mean square deviation (RMSD) between the mean coordinates of each ensemble of 4.6 Å (1.5 Å for the well defined residues, 20–75), the fit of the SAXS data is significantly better to the latter. Figure 3 shows the SAXS data, structural fit to the data, and the NMR structures overlaid on each SAXS envelope. For Samples 20–24, results similar to those observed in comparing SAXS data to the crystallographic structures are seen. The SAXS envelope accurately contains the globular portion of the NMR model, where the SAXS and NMR model diverge is consistent with the expected location of disordered residues. An exception to this appears to occur in Sample 21, where the NMR model indicates disordered structure extending away from the top right of the ordered portion but the SAXS envelope indicates that structural envelope is predominately to the right of the ordered portion of the molecule. Both samples 22 and 25 show large structurally disordered regions. SAXS is a technique that is sensitive to the time- and ensemble-averaged volume occupied by a protein, but there will be a case where the amount of time a protein molecule is in a particular position or the percentage of molecules in that position is too small to produce a signal that can be interpreted as the envelope and not noise. Although NMR can be used to characterize distributions of conformations in disordered regions by interpreting the data as arising from ensemble averaging, these methods were not used for these NMR structures and the distributions of conformations in disordered regions cannot be interpreted as representative of the true conformational distributions in solution. In this case, limitations of each technique must be realized and a balance needs to be made between the limitations of both techniques.

The SAXS data shows that Samples 18, 19, and 24 are in the monomeric state, which is in disagreement with the oligomeric state determined by analytical gel filtration. NMR 1D N_{15} T_1/T_2 measurements are a more accurate technique than gel filtration for oligomer determination. NMR 1D N_{15} T_1/T_2 measurements on Samples 18, 19, and 24 confirm the monomeric state, in agreement with the SAXS data.

The Combination of Crystallography and NMR with SAXS

For two samples, 27 and 28, both crystallographic and NMR structures were available. In each of these cases, the SAXS envelopes were in good agreement with the crystallographic and NMR structures. For Sample 27, the R_g and D_{\max} from the SAXS data were each within ~ 1 Å of the NMR structure. The R_g was within ~ 1 Å of the crystallographic structure, but the D_{\max} measures 15.8 Å greater when SAXS data is compared with the crystallographic structure. This is consistent with the crystallographic structure having 18 missing residues, $\sim 10\%$ of the structure, whereas NMR accounted for all the residues. For Sample 28, the R_g for the NMR data was in exact agreement with the SAXS data but the D_{\max} from the SAXS data was ~ 36 Å less. The crystallographic structure had 13 missing residues, 13% of the structure, which accounts for a smaller R_g and D_{\max} when compared with the SAXS values. The difference in D_{\max} from the NMR structure and SAXS data is discussed above.

The observed SAXS data and structural fit to the observed data and ab initio envelopes with structures overlaid are shown in Figure 4. The NMR and crystallographic structures are similar and fit well into the SAXS-derived molecular envelopes. In Sample 27, no residues are missing from the NMR structure (a) and 18 are missing from the crystallographic structure (b). For the NMR structure, regions of structural disorder are consistent with envelope regions otherwise not explained by the NMR structure. Similarly, in the X-ray structural case, missing residues are represented by the molecular envelope density consistent with the position and number of those residues. Sample 28 is missing 13 residues in the crystallographic structure (a), but can be positioned using the SAXS envelope. The NMR data (b), while fitting the experimental SAXS data better than the crystallographic (accounting for these missing residues), appears to place the bulk of the disordered region in a different location than the SAXS envelope suggests. When comparing these two examples, 18 missing residues can make little difference to the overall calculated curve in cases such as Sample 27, whereas a similar number of missing residues can have a great impact on the calculated curve, as seen in Sample 28. This may be due to location of the missing residues as well as their size

compared with the size of the particle as a whole. The PISA analysis predicted the same dimer organization for Sample 27 as determined by the SAXS data. However, the dimer predicted by PISA for Sample 28 was not seen in the SAXS data.

Mixtures

This study has been used to determine how well SAXS *ab initio* molecular envelope reconstructions represent known structures and from this gain an idea of the accuracy of cases where no structural information is present. However, having this structural information also allows to analyze samples as mixtures. Samples 4 and 11 were determined to be mixtures of oligomers from the SAXS data (Figure 5).

In these examples, scattering from each oligomer was calculated and estimates of volume fractions present in solution were carried out. For Sample 4, the fit to the curve improved from $\chi = 7.8$ for the dimer (not shown) to $\chi = 2.6$ for the dimer–tetramer mixture. This is seen primarily in the improvement of the low q -region of the curve, corresponding to the overall size of particles in solution. This D_{\max} value reflect this as well, 58.7 Å for the dimer (not shown), 81.2 Å for the tetramer, in good agreement with the SAXS estimated value of 82.7 Å. The 28 residues of the dimer, and 56 residues of the tetramer that were missing in the crystallographic structure, may explain the poor fit beyond about 0.13 \AA^{-1} for the mixture. For Sample 11, the fit improved dramatically from $\chi = 13.5$ for the dimer and $\chi = 10.4$ for the tetramer (not shown) to $\chi = 1.4$ for the mixture. Similarly, the D_{\max} for the dimer is only 71.0 Å, but for the tetramer, it increases to 80.8 Å, closer to the SAXS-derived D_{\max} of 89.7 Å. The PISA analysis indicated that both dimers were in stable oligomeric states but did not identify either tetramer.

Without knowledge of the structure, one is unable to determine volume fractions of oligomers in solution. *Ab initio* reconstructions for mixtures should not be carried out because most algorithms, including that used in the DAMMIF reconstructions in this study, assume a monodisperse solution and are not suited for polydisperse mixtures.⁵³ Attempts at reconstructing *ab initio* envelopes when samples are polydisperse are shown in Figure 5. It is readily seen that in some cases, for example, Sample 4, the envelope is a poor representation of either oligomer in solution, whereas in Sample 11, the envelope appears to be able to accommodate most of the tetramer. This illustrates that although an *ab initio* model may be constructed, it is not reliable for either oligomer if the solution is polydisperse. In a polydisperse solution containing multiple oligomers of the same basic quaternary unit, the intramolecular distances within the basic quaternary unit will be similar for each oligomer and thus

contribute similarly to the intensity profile as a monodisperse solution. Only the additional intramolecular distances present in the larger oligomer that are not present in the basic quaternary unit will contribute to the scattering differently than the monodisperse solution. This highlights the importance of oligomer screening before SAXS data collection, or the use of biophysical or biochemical separation techniques to ensure a single monodisperse oligomer population distribution, especially when no other structural information is available. If prior structural information exists, this illustrates the strength of the application of SAXS for mixed-oligomer analysis to characterize solutions containing mixtures of quaternary structures.

DISCUSSION

SAXS is not a new biophysical technique but it has only recently been applied to high-throughput structural biology.⁵⁶ In this article, SAXS has been approached from a different perspective, that of a high-throughput crystallization screening laboratory. SAXS has been used to characterize remnants of samples that remained after crystallization screening. More than 500 different proteins from this group have been characterized to date. From these samples, we have presented a subset of cases where crystallographic and/or NMR structural information was available. In some cases, this was known before SAXS, in other cases it became known subsequently. In all cases, SAXS studies using minimal amounts of sample at multiple concentrations but a single buffer condition, produced molecular envelopes that were consistent with crystallographic and NMR-based structural knowledge. We acknowledge the limitations of SAXS; for example, disordered regions may be averaged to a single area that is not representative of the actual molecular structure. Similarly, the SAXS envelope may not be completely sensitive to highly dynamic regions of a structure and in extreme cases could insufficiently sample and subsequently incorrectly represent the volume occupied by the flexible portion of the molecule. SAXS experiments can be performed on all the greater than 30,000 soluble, purified samples produced by the US PSI. SAXS could be used to structurally characterize the majority of these samples. We have demonstrated that these envelopes appear to be highly consistent with known structural information. If these samples could be characterized structurally, albeit at low resolution, they would significantly increase the amount of structural knowledge that is currently available.

The fact that our envelopes are in good agreement with known structures does not imply that envelopes for samples recalcitrant to crystallization will necessarily be representative

of the structures of these samples. There could be significant biochemical, biophysical, or structural reasons for failure to crystallize. However, SAXS is a powerful technique for characterizing samples in solution. It can distinguish between natively unfolded samples, those with flexible disordered regions and those that may have multiple globular regions with flexible linkers. We can identify these problems and limit our analysis to those samples that are well behaving. In doing so, we can have reasonable confidence that the molecular envelope produced from SAXS data reflects the molecular structure. However, reasonably confident is not completely confident. Without complementary structural, or biochemical knowledge, we can never be 100% certain of the accuracy of the envelope. We have to remain wary and have to settle for the fact that most of what we see from envelope reconstruction is correct, but this is not always going to be the case.

An important note in this study is the observation of two cases of mixtures. We have a limited sample set that has been well characterized on preparation but then cycled through freeze thaw cycles both before and after crystallization trials before SAXS analysis. Samples should be as fresh as possible and homogeneous. One approach that is clearly recommended is the use of size-exclusion chromatography and light scattering techniques immediately before SAXS data collection to monitor monodispersity.⁵⁷

SAXS is clearly complementary to high-resolution structural techniques such as crystallography and NMR spectroscopy. We have demonstrated that it provides unique quaternary structural information from the solution state that can be leveraged into biological knowledge that is not determined using independent methodologies. This is exemplified by the identification of oligomer organizations for Samples 2–6, 9, 11, 16, and 17 that are alternatives to those seen in the crystallographic structures. Binding energy and entropy of dissociation can be calculated, where structural information is present enabling prediction of the biological oligomer with services such as PISA.⁴⁷ This approach is successful in 80–90% of cases, a similar success rate seen with our data. However, SAXS can directly identify these oligomers to support or challenge the prediction.

In NMR, special data collection and analysis methods are required to determine the correct representation of highly disordered regions. In such disordered regions, SAXS data indicates a more compact structure than that indicated by the reported NMR conformational ensemble. To some extent, this is an issue with the calculation of D_{\max} from an ensemble of conformers, but there are clear cases where this alone does not fully explain the difference in D_{\max} values. Specific modeling of SAXS sensitivity is needed to resolve this case. There are methods to treat molecules or parts of

molecules as ensembles of conformers within the SAXS analysis. The ensemble optimization method (EOM)⁵⁸ randomly generates conformers, bins them to create ensembles, and using a genetic algorithm, optimizes the ensembles by comparing the average scattering profile of their conformers to the experimental data. Using an increasing number of conformers per ensemble, and an analysis of the deviation of experimental data from predicted data, SAXS analysis can be used to study dynamic structural regions. In this study, we have not made use of these methods due to the number of samples examined and the computational resources required for each case. On the other hand, the NMR methods used by the NESG consortium are not aimed at accurate representation of conformational distributions in disordered regions, which requires special methods and considerations.

The structural and biochemical data used in this study are publically available. We are happy to provide the SAXS data associated with this study to groups that may use it for further development. We have used SAXS to complement high-throughput crystallization screening and are uniquely positioned with the availability of a large number of well-behaved and well-characterized sample courtesy of the NESG efforts. We have presented a top-level overview of our initial results on a subset of samples, where structural information was already available. SAXS data has been useful and provided additional information in these cases.

The strength of SAXS shown by our results causes us to echo the conclusions of Hura et al.⁵⁶ in adopting the method for high-throughput structural genomic studies and to go one step further in suggesting that it is in fact essential. Although X-ray crystallography and NMR are clearly powerful structural techniques, when SAXS analysis is added, the synergistic relationship between the techniques provides a far greater understanding of the biological system as a whole.

Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program (P41RR001209) and the National Institute of General Medical Sciences. The authors acknowledge those responsible for the structural information they have used from the PDB and Dr. George DeTitta for access to the high-throughput screening laboratory and remnants of samples left after crystallization screening. The referees are acknowledged for useful comments. This work was supported in part by NIH grants R01 GM088396 to EHS, and Protein Structure Initiative grants U54 GM074958 and U54 GM094597 to GTM, and U54 GM074899 to George DeTitta. Dr. E. Lattman is acknowledged for useful discussions.

REFERENCES

- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* 2000, 28, 235–242.
- Watson, J. D.; Crick, F. H. *Nature* 1953, 171, 737–738.
- Hodgkin, D. C.; Kamper, J.; Lindsey, J.; MacKay, M.; Pickworth, J.; Robertson, J. H.; Shoemaker, C. B.; White, J. G.; Prosen, R. J.; Trueblood, K. N. *Proc R Soc Lond A* 1957, 242, 228–263.
- Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *J Mol Biol* 1984, 180, 385–398.
- Abrahams, J. P.; Leslie, A. G.; Lutter, R.; Walker, J. E. *Nature* 1994, 370, 621–628.
- Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* 1998, 280, 69–77.
- Cramer, P.; Bushnell, D. A.; Kornberg, R. D. *Science* 2001, 292, 1863–1876.
- Schlünzen, F.; Hansen, H. A. S.; Thygesen, J.; Bennett, W.S.; Volkman, N.; Harms, J.; Bartels, H.; Krumbholz, S.; Levin, I.; Zaytzev-Bashan, A.; Geva, M.; Weinstein, S.; Agmon, R. Sharon, R.; Dribin, A.; Maltz, E.; Peretz, M.; Weinrich, V.; Franceschi, F.; Bötdeker, N.; Morlang, S.; Berkovitch-Yellin, Z.; Yonath, A.; Sagi, I. *Biochem Cell Biol* 1995, 73, 739–749.
- Chen, L.; Oughtred, R.; Berman, H. M.; Westbrook, J. *Bioinformatics* 2004, 20, 2860–2862.
- Luft, J. R.; Collins, R. J.; Fehrman, N. A.; Lauricella, A. M.; Veatch, C. K.; DeTitta, G. T. *J Struct Biol* 2003, 142, 170–179.
- Chayen, N. E.; Shaw Stewart, P. D.; Blow, D. M. *J Cryst Growth* 1992, 122, 176–180.
- Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Q Rev Biophys* 2007, 40, 191–285.
- Xiao, R.; Anderson, S.; Aramini, J.; Belote, R.; Buchwald, W. A.; Ciccocanti, C.; Conover, K.; Everett, J. K.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Jiang, M.; Kornhaber, G. J.; Lee, D. Y.; Locke, J. Y.; Ma, L. C.; Maglaqui, M.; Mao, L.; Mitra, S.; Patel, D.; Rossi, P.; Sahdev, S.; Sharma, S.; Shastry, R.; Swapna, G. V.; Tong, S. N.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T.; Acton, T. B. *J Struct Biol* 2010, 172, 21–33.
- Acton, T. B.; Gunsalus, K. C.; Xiao, R.; Ma, L. C.; Aramini, J.; Baran, M. C.; Chiang, Y. W.; Climent, T.; Cooper, B.; Denissova, N. G.; Douglas, S. M.; Everett, J. K.; Ho, C. K.; Macapagal, D.; Rajan, P. K.; Shastry, R.; Shih, L. Y.; Swapna, G. V.; Wilson, M.; Wu, M.; Gerstein, M.; Inouye, M.; Hunt, J. F.; Montelione, G. T. *Methods Enzymol* 2005, 394, 210–243.
- Acton, T. B.; Xiao, R.; Anderson, S.; Aramini, J. M.; Buchwald, W.; Ciccocanti, C.; Conover, K.; Everett, J. K.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Kornhaber, G. J.; Lau, J.; Lee, D. Y.; Liu, G.; Maglaqui, M.; Ma, L.-C.; Mao, L.; Patel, D.; Rossi, P.; Sahdev, S.; Sharma, S.; Shastry, R.; Swapna, G. V. T.; Tang, Y.; Tong, S. N.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T. *Methods Enzymol* 2011, 493, 21–60.
- Forouhar, F.; Lew, S.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Maglaqui, M.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3hz7/pdb.
- Seetharaman, J.; Su, M.; Wang, H.; Foote, E. L.; Mao, L.; Nair, R.; Rost, B.; Acton, T. B.; Xiao, R.; Everett, J. K.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3h9w/pdb.
- Forouhar, F.; Lew, S.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Lee, D.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3lmf/pdb.
- Vorobiev, S.; Neely, H.; Seetharaman, J.; Wang, D.; Ciccocanti, C.; Mao, L.; Xiao, R.; Acton, T. B.; Everett, J. K.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3mjg/pdb.
- Forouhar, F.; Abashidze, M.; Seetharaman, J.; Mao, M.; Xiao, R.; Ciccocanti, C.; Lee, D.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3mfx/pdb.
- Seetharaman, J.; Lew, S.; Wang, D.; Janjua, H.; Cunningham, K.; Owens, L.; Xiao, R.; Liu, J.; Baran, M. C.; Acton, T. B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3lyy/pdb.
- Seetharaman, J.; Chen, Y.; Wang, D.; Janjua, H.; Cunningham, K.; Owens, L.; Xiao, R.; Liu, J.; Baran, M. C.; Acton, T. B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3mjg/pdb.
- Seetharaman, J.; Abashidze, M.; Forouhar, F.; Janjua, H.; Xiao, R.; Ciccocanti, C.; Foote, E. L.; Acton, T. B.; Rost, B.; Montelione, G. T.; Hunt, J. F.; Tong, L. 2009, DOI:10.2210/pdb3i24/pdb.
- Kuzin, A.; Chen, Y.; Seetharaman, J.; Mao, M.; Xiao, R.; Ciccocanti, C.; Foote, E. L.; Wang, H.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3icl/pdb.
- Vorobiev, S.; Neely, H.; Seetharaman, J.; Wang, H.; Foote, E. L.; Ciccocanti, C.; Sahdev, S.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3ign/pdb.
- Kuzin, A.; Su, M.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Maglaqui, M.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Hunt, J. F.; Tong, L. 2009, DOI:10.2210/pdb3ha2/pdb.
- Kuzin, A.; Scott, L.; Forouhar, F.; Abashidze, M.; Seetharaman, J.; Mao, M.; Xiao, R.; Ciccocanti, C.; Wang, H.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Hunt, J. F.; Tong, L. 2010, DOI:10.2210/pdb3ljx/pdb.
- Forouhar, F.; Neely, H.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Maglaqui, M.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Hunt, J. F.; Tong, L. 2009, DOI:10.2210/pdb3hxl/pdb.
- Forouhar, F.; Abashidze, M.; Seetharaman, J.; Mao, M.; Xiao, R.; Ciccocanti, C.; Foote, E. L.; Belote, R. L.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F., 2010, DOI:10.2210/pdb3lrx/pdb.
- Forouhar, F.; Abashidze, M.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Foote, E. L.; Ciccocanti, C.; Belote, R. L.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2010, DOI:10.2210/pdb3lyu/pdb.
- Vorobiev, S.; Chen, Y.; Forouhar, F.; Maglaqui, M.; Ciccocanti, C.; Mao, L.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3hix/pdb.
- Liu, G.; Shastry, R.; Ciccocanti, C.; Janjua, H.; Acton, T. B.; Xiao, R.; Mao, B.; Everett, J. K.; Montelione, G. T. 2010, DOI:10.2210/pdb2kw2/pdb.
- Liu, G.; Xiao, R.; Janjua, J.; Acton, T. B.; Mao, B.; Everett, J.; Montelione, G. T. 2010, DOI:10.2210/pdb2kvu/pdb.
- Cort, J. R.; Ramelot, T. A.; Lee, D.; Ciccocanti, C.; Janjua, H.; Acton, T. B.; Xiao, R.; Everett, J. K.; Montelione, G. T.; Kennedy, M. A. 2010, DOI:10.2210/pdb2kvz/pdb.

35. Liu, G.; Tong, S.; Xiao, R.; Acton, T. B.; Everett, J. K.; Montelione, G. T. 2010, DOI:10.2210/pdb210b/pdb.
36. Liu, G.; Janjua, H.; Xiao, R.; Ciccocanti, C.; Shastry, R.; Acton, T. B.; Tong, S.; Everett, J. K.; Montelione, G. T. 2010, DOI:10.2210/pdb2kz5/pdb.
37. Lee, H.; Montelione, G. T.; Prestegard, J. H. 2009, DOI:10.2210/pdb2kl1/pdb.
38. Cort, J.; Lee, D.; Ciccocanti, C.; Janjua, H.; Acton, T. B.; Xiao, R.; Everett, J. K.; Montelione, G. T.; Kennedy, M. A. 2010, DOI:10.2210/pdb210d/pdb.
39. Mills, J. L.; Eletsky, D.; Lee, C.; Lee, K.; Ciccocanti, T.; Hamilton, R.; Acton, J. B.; Xiao, G.; Everett, T. K.; Prestegard, J. G.; Montelione, G. T.; Szyperski, T. 2010, DOI:10.2210/pdb2kzw/pdb.
40. Eletsky, A.; Mills, J. L.; Lee, D.; Ciccocanti, C.; Hamilton, K.; Acton, T. B.; Xiao, R.; Everett, J. K.; Montelione, G. T.; Szyperski, T. 2010, DOI:10.2210/pdb2kw7/pdb.
41. Eletsky, A.; Garcia, E.; Wang, H.; Ciccocanti, C.; Jiang, M.; Nair, R.; Rost, B.; Acton, T. B.; Xiao, R.; Everett, J. K.; Lee, H.; Prestegard, J.; Montelione, G. T.; Szyperski, T. 2009, DOI:10.2210/pdb2ko1/pdb.
42. Vorobiev, S.; Su, M.; Seetharaman, J.; Janjua, J.; Xiao, R.; Ciccocanti, C.; Wang, H.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Montelione, G. T.; Tong, L.; Hunt, J. F. 2009, DOI:10.2210/pdb3ibw/pdb.
43. Tang, Y.; Xiao, R.; Ciccocanti, C.; Janjua, H.; Lee, D. Y.; Everett, J. K.; Swapna, G. V.; Acton, T. B.; Rost, B.; Montelione, G. T. *Proteins* 2010, 78, 2563–2568.
44. Vorobiev, S.; Chen, Y.; Lee, D.; Patel, D. J.; Ciccocanti, C.; Sahdev, S.; Acton, T. B.; Xiao, R.; Everett, J. K.; Montelione, G. T.; Hunt, J. F.; Tong, L. 2010, DOI:10.2210/pdb3ld7/pdb.
45. Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *Proteins* 2003, 52, 573–584.
46. Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem Rev* 2006, 106, 3055–3079.
47. Krissinel, E.; Henrick, K. *J Mol Biol* 2007, 372, 774–797.
48. Smolsky, I. L.; Liu, P.; Niebuhr, M.; Ito, K.; Weiss, T. M.; Tsuruta, H. *J Appl Crystallogr* 2007, 40, S453–S458.
49. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J Appl Crystallogr* 2003, 36, 1277–1282.
50. Guinier, A.; Foumet, F. *Small Angle Scattering of X-rays*; Wiley Interscience: New York, 1955.
51. Svergun, D. I. *J Appl Crystallogr* 1992, 25, 495–503.
52. Petoukhov, M. V.; Konarev, P. V.; Kikhney, A. G.; Svergun, D. I. *J Appl Crystallogr* 2007, 40, S223–S228.
53. Franke, D.; Svergun, D. I. *J Appl Crystallogr* 2009, 42, 342–346.
54. Volkov, V. V.; Svergun, D. I. *J Appl Crystallogr* 2003, 36, 860–864.
55. Svergun, D.; Barberato, C.; Koch, M. H. J. *J Appl Crystallogr* 1995, 28, 768–773.
56. Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., II; Tsutakawa, S. E.; Jenney, F. E., Jr.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S. J.; Scott, J. W.; Dillard, B. D.; Adams, M. W.; Tainer, J. A. *Nat Methods* 2009, 6, 606–612.
57. Rambo, R. P.; Tainer, J. A. *RNA* 2010, 16, 638–646.
58. Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J Am Chem Soc* 2007, 129, 5656–5664.

Reviewing Editor: Sarah A. Woodson