

FoXS: a web server for rapid computation and fitting of SAXS profiles

Dina Schneidman-Duhovny¹, Michal Hammel² and Andrej Sali^{1,*}

¹Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California at San Francisco, CA 94158 and ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received January 28, 2010; Revised April 30, 2010; Accepted May 11, 2010

ABSTRACT

Small angle X-ray scattering (SAXS) is an increasingly common technique for low-resolution structural characterization of molecules in solution. SAXS experiment determines the scattering intensity of a molecule as a function of spatial frequency, termed SAXS profile. SAXS profiles can contribute to many applications, such as comparing a conformation in solution with the corresponding X-ray structure, modeling a flexible or multi-modular protein, and assembling a macromolecular complex from its subunits. These applications require rapid computation of a SAXS profile from a molecular structure. FoXS (Fast X-Ray Scattering) is a rapid method for computing a SAXS profile of a given structure and for matching of the computed and experimental profiles. Here, we describe the interface and capabilities of the FoXS web server (<http://salilab.org/foxs>).

INTRODUCTION

SAXS is becoming a widely used technique for low-resolution structural characterization of molecules in solution (1–6). Unlike EM, NMR and X-ray crystallography, the key strength of SAXS is that it can be performed under a wide variety of solution conditions, including near physiological conditions. The experiment is performed with ~1.0 mg/ml of a macromolecular sample in a ~15 µl volume, and usually takes only a few minutes on a well-equipped synchrotron beam line (4). The SAXS profile of a macromolecule, $I(q)$, is computed by subtracting the SAXS profile of the buffer from the SAXS profile of the macromolecule in the buffer. The profile can be converted into an approximate distribution of pairwise atomic distances of the macromolecule (i.e. the pair-distribution function) via a Fourier transform.

Computational approaches for modeling a macromolecular structure based on its SAXS profile can be classified into *ab initio* and rigid body modeling methods (3). On the one hand, the *ab initio* methods search for coarse 3D shapes represented by dummy atoms (beads) that fit the experimental profile (7–9). On the other hand, rigid body modeling approaches refine an atomic model of the molecule with the aim to fit the computed SAXS profile to the experimental one (10). Therefore, rigid body modeling can be used only if an approximate structure of the studied molecule or its components are available.

Rigid body modeling approaches require the computation of the SAXS profile of a given atomic structure and its comparison with the experimental profile. Here, we describe a web server (FoXS) that performs this task. FoXS can be used as a tool for numerous SAXS-based modeling applications, such as comparing solution and crystal structures (4), modeling of a perturbed conformation (e.g. modeling active conformation starting from non-active conformation) (11), structural characterization of flexible proteins (12,13), assembly of multi domain proteins starting from single domain structures (14), assembly of multi protein complexes (10), fold recognition and comparative modeling (15,16), modeling of missing regions in the high-resolution structure (17), and determination of biologically relevant states from the crystal (18,19).

There are several methods and software tools for calculating a SAXS profile of a given molecular structure. The methods differ in the use of the inter-atomic distances and in the treatment of the solvation layer (20). Inter-atomic distances can be computed and used explicitly (5,14). The most popular method CRY SOL (21), also available as a web server, uses multipole expansion for fast calculation of the spherically averaged scattering profile. Another widely used approach is Monte Carlo sampling of the distances in the model (22). Coarse graining that combines several atoms in a single scattering

*To whom correspondence should be addressed. Tel: +1 415 514 4227; Fax: +1 415 514 4231; Email: sali@salilab.org

center can also be used to speed up the calculation (23,24). The solvation layer can be treated explicitly by introducing water molecules (24,25) or implicitly by a continuous envelope surrounding the model (21).

FoXS is a rapid and accurate web server for calculating a SAXS profile of a given molecular structure that explicitly computes all inter-atomic distances as well as models the first solvation layer based on the atomic solvent accessible areas. The server also provides an optimization of the hydration layer density as well as the excluded volume of the protein, to maximize the fit of the computed profile to the experimental profile. Additional optimization is achieved by adjusting the 'background' of the experimental profile for wider scattering angles (26). Next, we describe the method implemented in FoXS and its interface.

FOXS METHOD

The method for computing SAXS profiles is based on the Debye formula (27):

$$I(q) = \sum_{i=1}^N \sum_{j=1}^N f_i(q)f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}} \quad (1)$$

where the intensity, $I(q)$ is a function of the momentum transfer $q = (4\pi \sin \theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam; d_{ij} is the distance between atoms i and j , and N is the number of atoms in the molecule. In our model, the form factor $f_i(q)$ takes into account the displaced solvent as well as the hydration layer:

$$f_i(q) = f_v(q) - c_1 f_s(q) + c_2 s_i f_w(q) \quad (2)$$

where $f_v(q)$ is the atomic form factor *in vacuo* (21), $f_s(q)$ is the form factor of the dummy atom that represents the displaced solvent (28), s_i is the fraction of solvent accessible surface of the atom i (29), and $f_w(q)$ is the water form factor. The parameter c_1 is used to adjust the total excluded volume of the atoms (default value = 1.0) and c_2 is used to adjust the density of the water in the hydration layer (default value = 0.0). $I(q)$ is calculated rapidly per Equation 1 as described in detail in (14).

The computed profile is fitted to a given experimental SAXS profile by minimizing the χ function with respect to c , c_1 and c_2 :

$$\chi = \sqrt{\frac{1}{M} \sum_{i=1}^M \left(\frac{I_{\text{exp}}(q_i) - cI(q_i)}{\sigma(q_i)} \right)^2} \quad (3)$$

where $I_{\text{exp}}(q)$ and $I(q)$ are the experimental and computed profiles, respectively, $\sigma(q)$ is the experimental error of the measured profile, M is the number of points in the profile, and c is the scale factor. The minimal value of χ is found by enumerating c_1 and c_2 ($0.95 \leq c_1 \leq 1.12$ and $0 \leq c_2 \leq 4.0$), in steps of 0.005 and 0.1 respectively, and performing the linear least-squares minimization to find the value of c that minimizes χ for each c_1, c_2 combination.

In addition, FoXS has an option to adjust the background of the experimental profile (26).

FoXS was successfully tested with all PDB structures (30) that have an experimental SAXS profile in the open access SAXS database (<http://bioisis.net/>) (4) as well as a number of additional cases (in preparation).

FOXS WEB SERVER

The web server has one mandatory input, a structure file in the PDB format (or a zip archive with multiple PDB files). The server assigns form factors for the standard PDB protein and nucleic acid atoms. For other types of molecule, the form factors are assigned using the element field in the PDB format. Hydrogen atoms are considered implicitly for proteins and nucleic acids by adding their form factors to that of their bound heavy atom. If the input structure includes other groups, such as lipids or sugars, it is recommended to add all the hydrogens explicitly and turn off the 'implicit hydrogens' option on the FoXS input form.

The optional input includes an experimental SAXS profile, which must be obtained for the exact molecule specified in the input PDB file (including all loops, linkers and His tags). The profile is specified in a text file with three columns:

q , $I(q)$, $\sigma(q)$

q intensity error

0.00000 3280247.73 1904.037

0.00060 3280164.59 1417.031

0.00120 3279915.19 1840.032

where $q = (4\pi \sin \theta)/\lambda$, 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam. It is recommended to provide an accurate error estimate of $\sigma(q)$ because it is used for the profile fitting. If the third column is not given, FoXS will assume the error is distributed according to the Poisson distribution with $\lambda = 10$.

The server also has several optional input parameters. A user can specify an e-mail address for emailing a link to the results page. Maximal q value determines the range for calculating the profile (default $q_{\text{max}} = 0.5 \text{ \AA}^{-1}$). A user can also control the sampling resolution of the profile by setting the number of points in the profile. The profile will be sampled at the resolution equal to the maximal q value divided by the number of profile points. For example, if the q_{max} value is 0.5 \AA^{-1} and a user asks for 1000 profile points, the resulting profile will be uniformly sampled at the interval of 0.0005 \AA^{-1} . In addition, a user can also decide whether or not to include the hydration layer in the profile computation (included by default). It is also possible to decide whether or not to adjust the excluded volume of the protein (c_1 value—adjusted by default) and the background of the experimental profile (not adjusted by default).

Once the inputs are defined, FoXS computes the SAXS profile of the input PDB structure(s) and fits it to the experimental profile, if provided. For fitting, the computed profile is resampled at the q values sampled by the

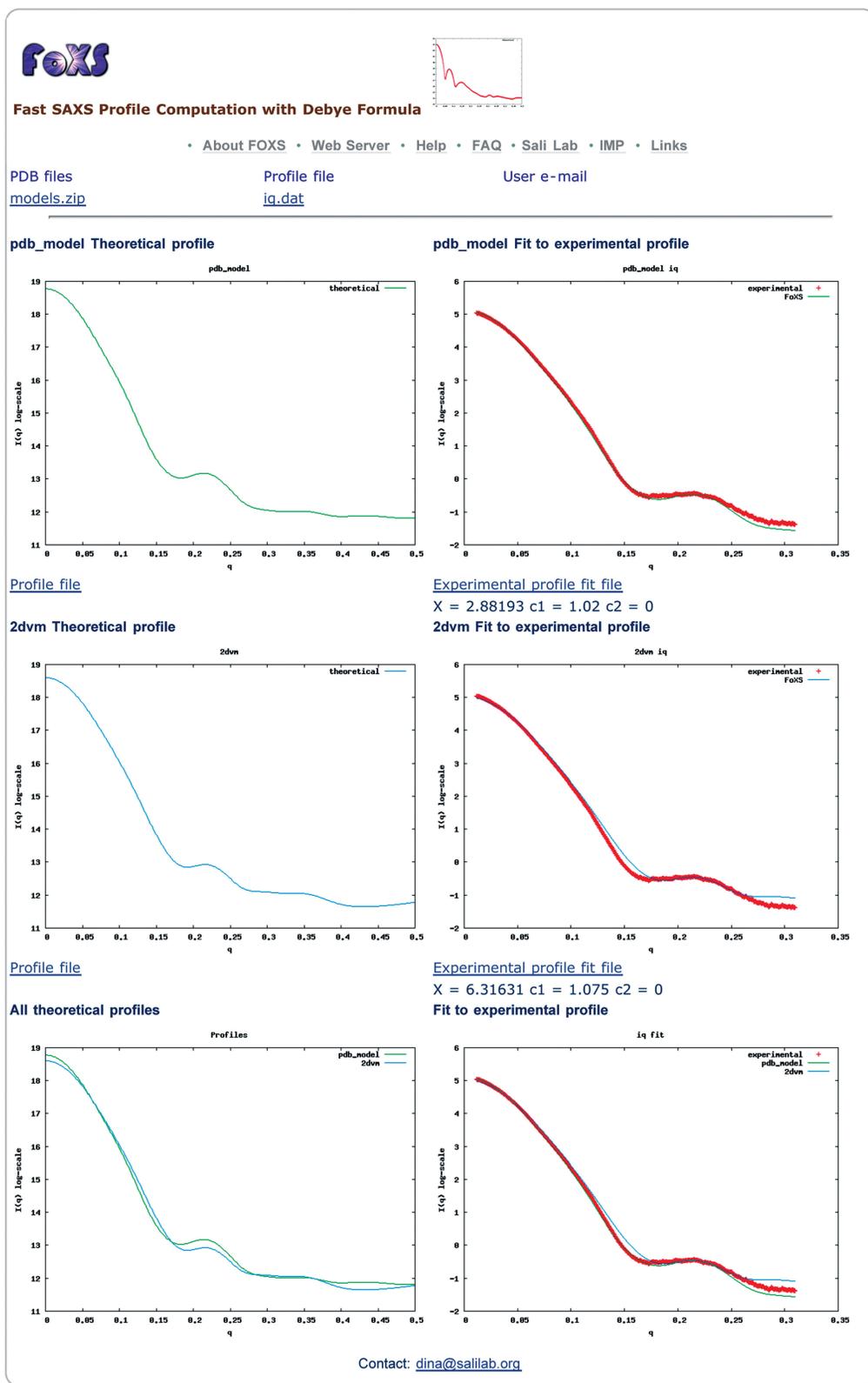


Figure 1. Snapshot of a FoXS output page. Computed profiles of two PDB files are compared to the experimental SAXS profile of malic enzyme (data from <http://bioisis.net/>, PF1026). The first structure (pdb_model) includes a model of the unfolded His tag region (35 residues), while the second structure (2dvm) does not. The server was run with the default parameters and the hydration layer modeling was disabled. Plots on the left display the theoretical profiles and plots on the right display their fit to the experimental profile. The top two plots are for the structure with the modeled unfolded region (pdb_model), the middle two plots are for the original PDB file (2dvm), the bottom left plot overlay the profiles for the two input structures, and the bottom right plot shows their fit to the experimental profile. The structure with the modeled unfolded region shows a better fit to the experimental profile with the value of $\chi = 2.88$, compared to $\chi = 6.33$ for the original crystallographic structure. The user can follow the links to download the computed profiles and their fittings.

experimental profile, with the computed intensities estimated using linear interpolation. The server does not modify the input experimental profile, unless background adjustment is requested.

The computation is performed in real time and the server page is updated once the calculation has finished. The typical running time is less than a second for a system of a thousand of atoms, and can extend to a few minutes for tens of thousands of atoms.

The output page displays a plot of the computed profile as well as a plot of the computed profile fitted to the experimental profile (Figure 1). In addition, the values of χ , c_1 and c_2 for the current profile fit are displayed. The profiles and their fit can be downloaded. In case of multiple PDB files in the input, the server will display computed profile plots for each PDB file, as well as plots with their fit to the experimental profile. In addition, a single plot with all the computed profiles and a plot with computed profiles fitted to the experimental profile are displayed.

CONCLUSIONS

A SAXS profile can provide significant insight into the structures of macromolecules, especially when combined with other information. SAXS experiments are gaining in popularity due to the technological advances that allow rapid and accurate data collection for a relatively small amount of the sample. Rapid and accurate computational methods are required for interpretation of the SAXS profiles. Here, we described a rapid, accurate, and user-friendly web server for calculating a SAXS profile of a given atomic structure and its fitting to the experimental profile. The web server can be used in a wide range of SAXS, such as comparing solution and crystal structures, modeling of a perturbed conformation (e.g. modeling active conformation starting from non-active conformation), structural characterization of flexible proteins, assembly of multi domain proteins starting from single domain structures, assembly of multi protein complexes, fold recognition and comparative modeling, modeling of missing regions in the high-resolution structure, and determination of biologically relevant states from the crystal.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Hiro Tsuruta for discussions about SAXS. The authors thank Ursula Pieper for help with the server setup. The authors are also grateful for computer hardware gifts from Ron Conway, Mike Homer, Intel, Hewlett-Packard, IBM and Netapp.

FUNDING

Weizmann Institute Advancing Women in Science Postdoctoral Fellowship to DSD; Sandler Family Supporting Foundation, National Institutes of Health (R01 GM083960); National Institutes of Health (U54

RR022220); National Institutes of Health (PN2 EY016525); DOE program Integrated Diffraction Analysis Technologies (IDAT), Pfizer Inc. SIBYLS beamline at Lawrence Berkeley National Laboratory. Funding for open access charge: National Institutes of Health (R01 GM083960).

Conflict of interest statement. None declared.

REFERENCES

- Koch, M.H., Vachette, P. and Svergun, D.I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.*, **36**, 147–227.
- Petoukhov, M.V. and Svergun, D.I. (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.*, **17**, 562–571.
- Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.*, **40**, 191–285.
- Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L. 2nd, Tsutakawa, S.E., Jenney, F.E. Jr, Classen, S., Frankel, K.A., Hopkins, R.C. *et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*, **6**, 606–612.
- Tiede, D.M., Mardis, K.L. and Zuo, X. (2009) X-ray scattering combined with coordinate-based analyses for applications in natural and artificial photosynthesis. *Photosynth. Res.*, **102**, 267–279.
- Whitten, A.E. and Trehwella, J. (2009) Small-angle scattering and neutron contrast variation for studying bio-molecular complexes. *Methods Mol. Biol.*, **544**, 307–323.
- Chacon, P., Moran, F., Diaz, J.F., Pantos, E. and Andreu, J.M. (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.*, **74**, 2760–2775.
- Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, **76**, 2879–2886.
- Svergun, D.I., Petoukhov, M.V. and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946–2953.
- Petoukhov, M.V. and Svergun, D.I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, **89**, 1237–1250.
- Nishimura, N., Hitomi, K., Arvai, A.S., Rambo, R.P., Hitomi, C., Cutler, S.R., Schroeder, J.I. and Getzoff, E.D. (2009) Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science*, **326**, 1373–1379.
- Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
- Pelikan, M., Hura, G.L. and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.*, **28**, 174–189.
- Förster, F., Webb, B., Krukenberg, K.A., Tsuruta, H., Agard, D.A. and Sali, A. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. *J. Mol. Biol.*, **382**, 1089–1106.
- Zheng, W. and Doniach, S. (2002) Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *J. Mol. Biol.*, **316**, 173–187.
- Zheng, W. and Doniach, S. (2005) Fold recognition aided by constraints from small angle X-ray scattering data. *Protein Eng. Des. Sel.*, **18**, 209–219.
- Petoukhov, M.V., Eady, N.A., Brown, K.A. and Svergun, D.I. (2002) Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophys. J.*, **83**, 3113–3125.

18. Hammel,M., Kriechbaum,M., Gries,A., Kostner,G.M., Laggner,P. and Prassl,R. (2002) Solution structure of human and bovine beta(2)-glycoprotein I revealed by small-angle X-ray scattering. *J. Mol. Biol.*, **321**, 85–97.
19. Datta,A.B., Hura,G.L. and Wolberger,C. (2009) The structure and conformation of Lys63-linked tetraubiquitin. *J. Mol. Biol.*, **392**, 1117–1124.
20. Perkins,S.J. (2001) X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modelling fits. *Biophys. Chem.*, **93**, 129–139.
21. Svergun,D., Barberato,C. and Koch,M.H.J. (1995) CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
22. Tjioe,E. and Heller,W.T. (2007) ORNL_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes. *J. Appl. Crystallogr.*, **40**, 782–785.
23. Grishaev,A., Wu,J., Trewella,J. and Bax,A. (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.*, **127**, 16621–16628.
24. Yang,S., Park,S., Makowski,L. and Roux,B. (2009) A rapid coarse residue-based computational method for X-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys. J.*, **96**, 4449–4463.
25. Merzel,F. and Smith,J.C. (2002) SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 242–249.
26. Ciccariello,S., Goodisman,J. and Brumberger,H. (1988) On the Porod law. *J. Appl. Crystallogr.*, **21**, 117–128.
27. Debye,P. (1915) Zerstreuung von Röntgenstrahlen. *Annalen der Physik*, **351**, 809–823.
28. Fraser,R.D.B., MacRae,T.P. and Suzuki,E. (1978) An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules. *J. Appl. Crystallogr.*, **11**, 693–694.
29. Connolly,M.L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
30. Berman,H., Henrick,K. and Nakamura,H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.