electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898

A software system for rigid-body modelling of solution scattering data

M.B. Kozin and D.I. Svergun

J. Appl. Cryst. (2000). **33**, 775–777

A software system for rigid-body modelling of solution scattering data

Mikhail B. Kozin[abc] and Dmitri I. Svergun[ab*]

[a]*European Molecular Biology Laboratory, Notkestrasse 85, D-22603 Hamburg, Germany*
[b]*Institute of Crystallography, Leninsky pr. 59, 117333 Moscow, Russia*
[c]*GKSS Research Center, Geesthacht, Germany*
Email:svergun@embl-hamburg.de

A computer system for rigid body modelling against solution scattering data is described. Fast algorithms to compute scattering from a complex of two arbitrary positioned subunits are implemented and coupled with the graphics program *ASSA* (Kozin, Volkov & Svergun, 1997, *J. Appl. Cryst.* **30**, 811-815). Mutual positions and orientations of the subunits (represented by low-resolution envelopes or by atomic models) can be determined by interactively fitting the experimental scattering curve from the complex. The system runs on the major Unix platforms (SUN, SGI and DEC workstations).

1. Introduction.

Small-angle X-ray and neutron scattering (SAS) are widely used to analyse low-resolution structure and conformational changes of native biological macromolecules in solution. SAS studies of large macromolecular complexes are of particular interest as it is often difficult to obtain crystals of such complexes suitable for X-ray crystallographic studies. Instead, crystals are grown of individual domains or subunits composing the complex to obtain their structure at atomic resolution. A model of the entire complex can then be constructed using SAS with rigid body modelling (Ashton *et al.*, 1997; Krueger *et al.*, 1997; Svergun *et al.*, 1998). Recently, new approaches for *ab initio* quaternary structure analysis and advanced modelling have been developed (Svergun, 1994, Svergun *et al.*, 1996, 1997). In particular, methods to accurately compute solution scattering patterns from atomic models were proposed (Svergun *et al.*, 1995, 1998) and algorithms for fast computation of the scattering from complex particles were implemented (Svergun, 1994; Svergun *et al.*, 1997).

An automated rigid body modelling can be performed based on these algorithms *e.g.* by an exhaustive search of positional parameters describing the subunit organisation to fit the experimental scattering from the complex. It is often difficult to formalise biological and chemical requirements that should be taken into account in such an automated subunit positioning. The straightforward automated search may thus yield a structure that fits the data but has little biochemical sense. The biochemical information (*e.g.* on contacts between the domains) can be best accounted for using an interactive search based on visual criteria. Integrating the computational modules in a single modelling system with feedback and graphics interface will allow the users to answer their questions with a maximum of transparency.

The implementation of the interactive modelling approach is described in the present paper. First, the rigid body algorithm for subunits positioning from solution scattering data is presented. Then, the incorporation of the algorithm into the 3D graphics package *ASSA* (Kozin, Volkov & Svergun, 1997) and the details of the user interface are discussed. Finally, the application of the developed system is illustrated on the study of the quaternary structure of the *E.coli* F1 ATP syntase (Svergun *et al.* 1998).

2. Rigid body refinement algorithm

The idea behind the rigid body refinement algorithm can be formulated as follows. Let us consider a complex consisting of two subunits A and B, and denote the scattering amplitudes from the two subunits centred at the origin in the reference orientations as $A(\boldsymbol{s})$ and $B(\boldsymbol{s})$ respectively (here, $\mathbf{s}$ is the momentum transfer vector; $s=|\boldsymbol{s}|=(4\pi/\lambda)\sin\theta$, 2θ is the scattering angle and λ is the wavelength). To construct an arbitrary complex, the first subunit can be fixed and the second rotated and moved. The rotation is described by the Euler angles α, β and γ (Edmonds, 1957) and the shift by the vector $\boldsymbol{u} = (u_x, u_y, u_z)$, so that the entire operation is described by six parameters. These parameters can be evaluated by minimising the discrepancy between the experimental scattering from the complex and that calculated as described below.

The scattering from a complex of the two subunits is expressed as (Svergun, 1994)

$$I(s) = I_A(s) + I_B(s) + 2\left\langle A(\mathbf{s})C^*(\mathbf{s})\right\rangle_\Omega , \qquad (1)$$

where $C(\boldsymbol{s})$ is the scattering amplitude from the displaced second subunit, and $\langle\rangle_\Omega$ denotes spherical average in reciprocal space.

Assume that the scattering amplitudes from the subunits are known and expand them into the series

$$A(\mathbf{s}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}(s)Y_{lm}(\Omega) , \qquad (2)$$

$$B(\mathbf{s}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} B_{lm}(s)Y_{lm}(\Omega) , \qquad (3)$$

where $Y_{lm}(\Omega)$ are spherical harmonics. Due to the orthogonal properties of spherical harmonics, Eq. (1) reduces to

$$I(s) = 2\pi^2\sum_{0}^{\infty}\sum_{-l}^{l}\left(|A_{lm}(s)|^2 + |B_{lm}(s)|^2\right) + 4\pi^2\sum_{0}^{\infty}\sum_{-l}^{l}\mathrm{Re}\left[A_{lm}(s)C^*_{lm}(s)\right] \qquad (4)$$

If the entire complex is aligned so that the direction of the vector $\boldsymbol{u}$ coincides with the Z-axis, then the partial amplitudes of the shifted second subunit are expressed as (Svergun et al., 1997)

$$C^0_{lm}(s,u) = (-1)^m\sum_{p=0}^{\infty} j_p(su)\sum_{k=|l-p|}^{l+p} d_{lm}(k,p)\sum_{j=-k}^{k} B_{kj}(s), \qquad (5)$$

where $j_p(x)$ are the spherical Bessel functions and the coefficients $d_{lm}(k,p)$ are represented through the 3j Wigner symbols (Edmonds, 1957)

$$d_{lm}(k,p) = i^p (2p+1)\sqrt{(2l+1)(2k+1)} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ -m & 0 & m \end{pmatrix}. \quad (6)$$

The effect of rotation is expressed as

$$C_{lm}(s) = \sum_{k=-l}^{l} D^{l}_{mk}(\alpha,\beta,\gamma) C^{0}_{lk}(s,u), \quad (7)$$

where $D_{mk}^{l}(\alpha, \beta, \gamma)$ denote the elements of the finite rotation matrix (Edmonds, 1957).

As the structures of the subunits and their partial amplitudes are known, the scattering intensity from the complex can be rapidly evaluated using Eqs. (5-7). The six positional parameters (α, β, γ, u_x, u_y, u_z) of the second subunit can be determined to minimise the discrepancy between the experimental and calculated curves:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left[\frac{I(s_j) - I_{\exp}(s_j)}{\sigma(s_j)} \right]^2, \quad (8)$$

where $I_{exp}(s)$ is the experimental intensity specified at N points s_j, j = 1, ... N, and $\sigma(s_j)$ is the correspondent standard deviation.

The program *ALM22INT* written in FORTRAN calculates the scattering intensity from a complex particle for the given positional parameters using Eqs. (4-7), and optionally implements an exhaustive search in real space to determine the optimal subunit positions by minimising the discrepancy (8).

3. Interactive positioning of domains

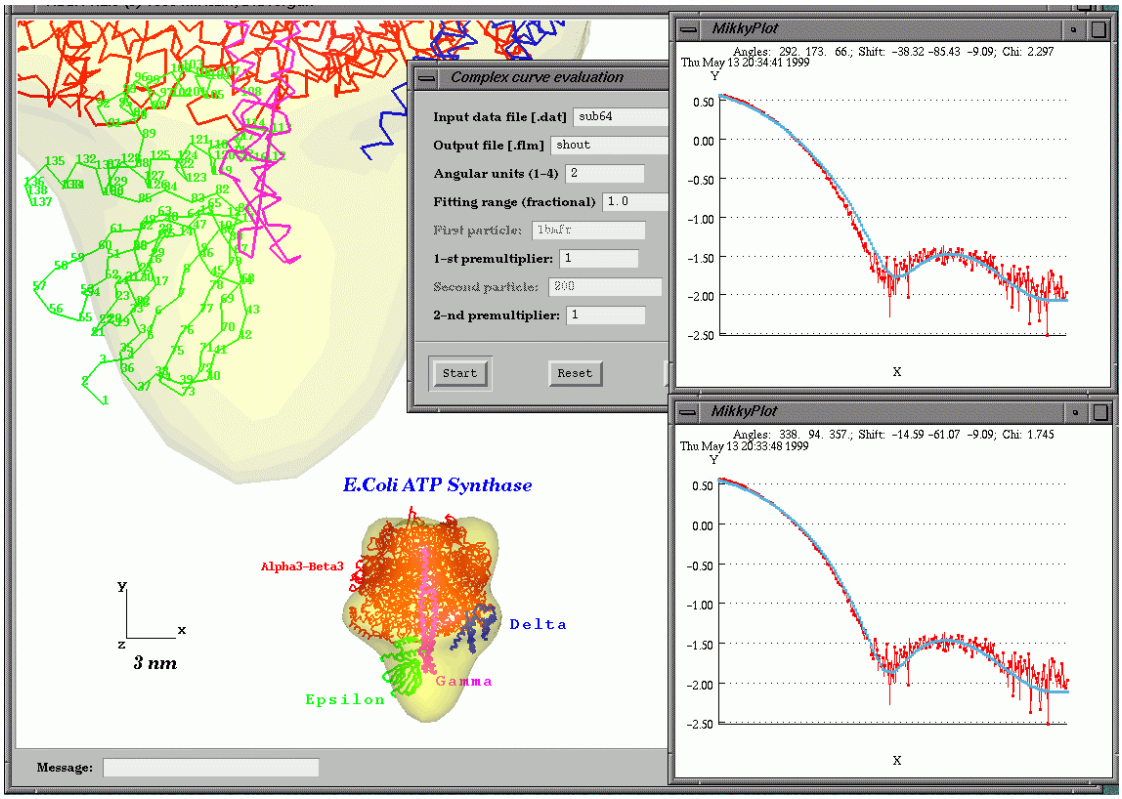
The optimal set of positional parameters should correspond to the global minimum of the functional (8) and the problem of global minimisation is rather complicated. Moreover, the formally best solution may make little biochemical sense. In fact, a constrained minimisation should be implemented, but, as mentioned above, biochemical requirements can hardly be formalised as mathematical constraints. In most practical cases visual criteria play an important role. In order to take into account such information as the mutual arrangement of certain residues or the range of feasible conformational changes (for instance, domain movements opening/closing active sites) the rigid body refinement procedure should be incorporated into a graphical package with a feedback interface. This would allow to interactively analyse the influence of the subunit movement and rotation on the goodness of fit to the experimental data.

The 3D graphics package *ASSA* (Kozin, Volkov & Svergun, 1997) was used as a frame for the interactive control over the program *ALM22INT*. The graphical features of *ASSA* fulfil the necessary requirements: the package is open for coupling with other programs, and it allows to analyse structure and mutual position of models, specified in different formats, *e.g.* SAS low-resolution models (Svergun & Stuhrmann, 1991) and standard Protein Data Bank (PDB) atomic structures (Bernstein *et al.*, 1977). Recently a new version of *ASSA* has been developed running on SUN SPARC, SGI and DEC Unix workstations. The new version for SGI and DEC uses standard OpenGL library for 3D graphics, Motif and Xt Intrinsic libraries for the menu-driven user interface and has a number of new features. In particular, new tools to represent homodimeric structures and to display the packing of macromolecules in the crystal are added.

The program *ALM22INT* is invoked from *ASSA* by a pop-up command window (Fig. 1). The user specifies the experimental data file and picks up two objects, that can be low-resolution envelope models represented by spherical harmonics or atomic models in the PDB format. A group of several objects can also be merged and specified as a single subunit. To start the procedure the system requires the multipole components of scattering amplitudes of the selected subunits precomputed and stored as separate files. This is performed by the programs *FLM2ALM* for the envelope models and *CRYSOL* (Svergun, Barberato & Koch, 1995) for atomic models and groups of objects.

The intensity evaluation of the complex is then started as an independent child process. The parameters describing mutual position and orientation of the selected objects are transferred to *ALM22INT* through information pipes upon request. To speed up the calculations, *ASSA* computes and transmits the coordinates of the complex rotated to make the vector $\boldsymbol{u}$ coincide with the Z-axis. The fit to the experimental data is plotted in a separate window and these plots can be kept on the screen (Fig. 1). The effect of changes in the subunit positions to the goodness of fit can thus directly be seen and the current value of the discrepancy χ (8) is displayed in the plots.
The available biochemical information can be used to put limits on the allowable movements of subunits. In particular, for each individual subunit the rotation point can be set to a specified residue, selected either by its number or per mouse click; moreover, the direction of the rotation axis can also be chosen by the user. This permits to preserve the mutual arrangement of certain residuals during the modelling. The absolute position of the geometrical centre of every subunit can also be displayed on the screen, thus allowing to impose shift restrictions. The sensitivity of the interactive search in the vicinity of the local minimum of the target function (8) can be tuned by adjusting magnitudes of the angular and spatial steps. The exact configuration corresponding to the local minimum in the vicinity of the current position can further be found by an automated local refinement procedure.

Fig.1 illustrates the application of the interactive modelling in the study of the quaternary structure of the *E.coli* F1 ATPase (Svergun *et al.* 1998). This complex contains five different subunits (α, β, γ, δ, ε), and the structures of $\alpha_3\beta_3\gamma$ subcomplex as well as δ and ε subunits were determined by X-ray crystallography and NMR. The low-resolution model of the particle envelope obtained *ab initio* from SAS data allowed to unequivocally identify the volume occupied by the $\alpha_3\beta_3\gamma$ subcomplex. The location of the δ subunit near the bottom of the $\alpha_3\beta_3$ hexamer was found from investigation of β-δ disulphide formation in the mutant *E.coli* F1 ATPase (Grueber & Capaldi, 1996). The ε subunit was positioned assuming the $\alpha_3\beta_3\gamma\delta$ assembly as a single second subunit of the complex. The movements of the ε subunit were constrained to lie within the low-resolution



envelope found for the entire complex and to agree with the available information about the cross-links between the β and ε subunits, reported by Dallmann, Flynn & Dunn (1992). The solution in Fig.1 provides the discrepancy χ=1.68 to the experimental data.

The result of the interactive subunit positioning described above does not necessarily correspond to the best χ value that would have been provided by an exhaustive search in the case of an automatic rigid body refinement. For *E.coli* F1 ATPase, the global minimum of the target function (8) would correspond to a position of the ε subunit outside the particle envelope, which makes little sense. Incorporating biochemical information implies a certain trade-off between the goodness of fit and biological relevance: the domain organisation should correspond to a local minimum of the discrepancy functional for all feasible configurations. The statistical significance of the solution can be assessed using the *F* statistics (Bevington, 1969) as follows: moving the ε subunit by approximately 5 Å from its position shown on Fig. 1 worsens the χ by 0.1, which is statistically significant with a probability of 0.997.

The evaluation and display of a single fit takes less than a second of CPU time on an average UNIX workstation. The executable modules of the system for SUN, SGI and DEC workstations along with the user manual are available from http://www.embl-hamburg.de/ExternalInfo/Research/Sax/index.html.

The authors are indebted to Dr. M. H. J. Koch for useful comments. MK is grateful to the GKSS Research Centre and EMBL Hamburg Outstation for financial support of the project.

Figure 1
Interactive positioning of the ε-subunit of *E.coli* F1 ATPase. The pop-up window used to start the process from *ASSA* is displayed together with the two fits to the experimental data. The discrepancy value (8) is printed in the upper-right corner of the plots. The lower plot corresponds to the rendered position of the ε-subunit found to be optimal.

References

Ashton, A.W., Boehm, M.K., Gallimore, J.R., Pepys, M.B. & Perkins, S.J. (1997) *J. Mol. Biol.* **272**, 408-422.
Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535-542.
Bevington, P.B. (1969) *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New-York.
Dallmann, H.G., Flynn, T.G. & Dunn S.D. (1992) *J. Biol. Chem.* **270**, 24609-24614.
Edmonds, A.R. (1957) *Angular Momentum in Quantum Mechanics*. Princeton: Princeton University Press.
Grueber, G. and Capaldi, R.A. (1996) *J. Biol. Chem.* **271**, 32623-32628.
Kozin, M.B, Volkov, V.V. & Svergun, D.I. (1997). *J. Appl. Cryst.* **30**, 811-815.
Krueger, J.K., Olah, G.A., Rokop, S.E., Zhi, G., Stull, J.T., & Trewhella, J. (1997). *Biochemistry*, **36,** 6017-6023.
Svergun, D.I. (1994). *Acta Cryst.* **A50**, 391-402.
Svergun, D.I., Aldag, I., Sieck, T., Altendorf, K.-H., Koch, M.H.J., Clarke, R.J., Kane, D.J., Kozin, M.B. & Grüber G. (1998) *Biophysical Journal*, **75**, 2212-2219.
Svergun, D.I., Barberato, C. & Koch, M.H.J. (1995). *J. Appl. Cryst.* **28**, 768-773
Svergun, D.I., Volkov, V.V., Kozin, M.B. & Stuhrmann H.B. (1996). *Acta Cryst.* **A52**, 419-426.
Svergun, D.I., Volkov, V.V., Kozin, M.B., Stuhrmann, H.B., Barberato, C. & Koch, M.H.J. (1997). *J. Appl. Cryst.* **30**, 798-802.
Svergun, D.I. & Stuhrmann, H.B. (1991). *Acta Cryst.* **A47**, 736-744.

 Received 17 May 1999 • Accepted 24 January 2000