Eur Biophys J (2006) 35: 567–576
DOI 10.1007/s00249-006-0063-9

ARTICLE

Maxim V. Petoukhov · Dmitri I. Svergun

Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution

Received: 31 January 2006 / Revised: 16 March 2006 / Accepted: 22 March 2006 / Published online: 25 April 2006

Abstract Novel techniques for simultaneous analysis of X-ray and neutron scattering patterns from macromolecular complexes in solution are presented. They include ab initio shape and internal structure determination of multicomponent particles and more detailed rigid body modeling of complexes using high resolution structures of subunits. The methods fit simultaneously X-ray and neutron scattering curves including contrast variation data sets from selectively deuterated complexes. Biochemically sound interconnected models without steric clashes between the components displaying a pre-defined symmetry are generated. For rigid body modeling, distance restraints between specified residues/nucleotides or their ranges are taken into account. The efficiency of the methods is demonstrated in model examples, and potential sources of ambiguity are discussed.

Introduction

Small-angle scattering of X-rays and neutrons (SAXS and SANS) is widely used to study low resolution structure of non-crystalline systems of different nature, including biological macromolecules (Feigin and Svergun 1987). Recent developments in SAXS data analysis methods made it possible to retrieve significantly more structural information from the high quality scattering patterns than previously believed [see e.g., Svergun and Koch (2003) for a review] and the method is now being employed to address increasingly complicated questions. The absence of radiation damage and possibility of contrast variation by solvent exchange (H_2O/D_2O) (Ibel and Stuhrmann 1975) or specific deuteration (Engelman and Moore 1972) makes SANS an extremely useful complementary tool to SAXS. Using contrast variation, multiple data sets are recorded from the same object, which are different because of different contrasts of the components in the particle. Contrast of the component is defined as $\Delta\rho = <\rho(\mathbf{r})> - \rho_s$, where $<\rho(\mathbf{r})>$ is the average scattering length density of the component, ρ_s is that of the solvent. With neutrons, the former quantity can be changed by selective deuteration, while the latter is conveniently varied in a broad range by H_2O/ D_2O exchange in solution. The additional information gained by isotopic H/D substitution is especially important in the study of functional complexes, which are rapidly becoming major targets of modern structural biology (Aloy et al. 2004; Sali et al. 2003). The remarkable progress in instrumentation and biological sample preparation procedures with high yield production of deuterated material (Vanatalu et al. 1993) needs to be accompanied by adequate developments in SANS data analysis methods to transform the wealth of contrast variation data into meaningful models. In the present paper, novel data analysis techniques allowing simultaneous interpretation of X-ray and neutron scattering patterns from macromolecular solutions will be presented. These methods include ab initio shape and internal structure determination of multicomponent particles and yet more detailed modeling of macromolecular complexes using rigid body refinement.

Ab initio shape determination is one of the most frequently used tools in SAXS/SANS data analysis. Modern shape determination algorithms employ Monte–Carlo based search to generate macromolecular models consisting of beads or dummy residues (Chacon et al. 1998; Svergun 1999; Svergun et al. 2001; Vigil et al. 2001;

M. V. Petoukhov · D. I. Svergun (✉)
European Molecular Biology Laboratory, Hamburg Outstation,
c/o DESY, Notkestraße 85, 22603 Hamburg, Germany
E-mail: Svergun@EMBL-Hamburg.DE
Tel.: +49-40-89902125
Fax: +49-40-89902149

M. V. Petoukhov · D. I. Svergun
Institute of Crystallography, Russian Academy of Sciences,
Leninsky pr. 59, 117333 Moscow, Russia

Walther et al. 2000). When using contrast variation data in SANS, not only shape but also internal structure of multicomponent particles can be retrieved ab initio from the scattering data. Thus, a low resolution map of protein–RNA distribution in the 70S *E. coli* ribosome was obtained (Svergun and Nierhaus 2000) before the high resolution crystallographic structures appeared and the a posteriori comparison proves that the SANS-based structure was a very good prediction (Koch et al. 2003). We present here an enhanced method for ab initio multiphase bead modeling against X-ray and neutron scattering data, implemented in the program MONSA, where, in particular, a possibility is included to account for the symmetry of the complex. Another practically important approach for the structure analysis of macromolecular complexes in solution is rigid body modeling in terms of the high resolution models. Thanks to the tremendous progress in large scale structure determination of individual macromolecules or their subunits by crystallography and NMR, rigid body modeling using lower resolution methods (cryo-EM or SAXS/SANS) has become an extremely valuable tool in the study of complexes (Nixon et al. 2005; Rosano et al. 2004; Sabatucci et al. 2005). Advanced methods developed for interactive (Konarev et al. 2001; Kozin et al. 1997) and automated (Petoukhov and Svergun 2005) modeling of the complexes based on SAXS data have already been successfully used in practice. The major bottleneck of SAXS-based rigid body modeling lies in the possibility of obtaining multiple solutions, which fit the data equally well (Petoukhov and Svergun 2005). The ambiguity can be reduced by using additional information like distance restrains between specific residues or nucleotides, particle symmetry, orientational restrains from residual dipolar coupling (RDC) in NMR etc. (Grishaev et al. 2005; Krueger et al. 2000; Mattinen et al. 2002; Tung et al. 2002). Extremely important information is provided by contrast variation in SANS, and fitting of multiple data sets has already been employed by several groups (Furtado et al. 2004; Heller et al. 2003; King et al. 2005). We present here a new version of the global rigid body modeling program SASREF [originally developed for SAXS-based analysis (Petoukhov and Svergun 2005)], which allows simultaneous fitting of multiple SAXS and SANS data sets while accounting for particle symmetry, distance restrains, perdeuteration etc. The effectiveness of the new methods is validated in simulated examples but also possible sources of obtaining ambiguous models are discussed.

Theory and methods

Scattering from a multicomponent particle

Let us consider a complex consisting of $K > 1$ components (subunits). Solution scattering intensity $I(s)$ of such a particle is expressed as (Svergun 1997):

$$I(s) = \left\langle \left| \sum_{k=1}^{K} A^{(k)}(\boldsymbol{s}) \right|^2 \right\rangle_{\Omega}, \tag{1}$$

where $\boldsymbol{s}$ is the scattering vector in reciprocal space, $s = 4\pi \sin(\theta)/\lambda$, 2θ is the scattering angle and λ is the wavelength, $A^{(k)}(\boldsymbol{s})$ denotes the scattering amplitude of the kth subunit at the given position and $\langle \ldots \rangle_{\Omega}$ stands for the spherical average in reciprocal space.

The amplitudes can be represented using the spherical harmonics $Y_{lm}(\Omega)$ to allow for analytical spherical averaging leading to a convenient analytical representation of the scattering intensity in the form:

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| \sum_{k=1}^{K} A_{lm}^{(k)}(s) \right|^2. \tag{2}$$

Here, the complex functions $A^{(k)}{}_{lm}(s)$ are the partial scattering amplitudes of the kth component.

Structure modeling of multisubunit particles against solution scattering data aims at finding the spatial arrangement of the components (and their shapes in ab initio case), scattering from which would best fit the experimental scattering from the entire complex. If multiple data sets are available, e.g., from contrast variation and/or from partial constructs, these sets can be fitted simultaneously. To construct physically sound models fitting (multiple) scattering data, the two modeling algorithms presented below minimize a target function F in the form:

$$F = \frac{1}{N_c} \sum_{i=1}^{N_c} \chi_i^2 + \text{penalties}. \tag{3}$$

The first term in Eq. 3 ensures minimization of the overall discrepancy in multiple data sets (N_c is the total number of scattering curves) from contrast variation series and/or from partial constructs, whereas the penalty term formulates physical restrictions on the model (compactness, connectivity, distance restrains etc. depending on the algorithm). The discrepancy from individual scattering curve $I^i_{\text{exp}}(s)$ is given by the equation:

$$\chi_i^2 = \frac{1}{N_i - 1} \sum_j \left[\frac{I^i_{\text{exp}}(s_j) - c_i I^i(s_j)}{\sigma_i(s_j)} \right]^2, \tag{4}$$

where N_i is the number of experimental points, c_i is a scaling factor and $\sigma_i(s_j)$ is the experimental error at the momentum transfer s_j.

The minimization of the target function F in both ab initio and rigid body approaches is performed using simulated annealing (SA) (Kirkpatrick et al. 1983). The main idea in this method is to perform random modifications of the system while moving always to the configurations that decrease the target function, but sometimes also to those that increase F. The probability of accepting the latter moves decreases in the course of the minimization (the system is cooled). At the beginning, the temperature is high and the changes almost

random, whereas at the end a configuration with nearly minimum energy is reached. The algorithm was implemented in its faster simulated quenching (Ingber 1993; Press et al. 1992) version.

Ab initio analysis

Several approaches were proposed for ab initio shape determination using SAXS data (Chacon et al. 1998; Svergun 1999; Vigil et al. 2001; Walther et al. 2000), where the model is represented as a "binary" (particle–solvent) system. The SA-based bead modeling algorithm (Svergun 1999) also allowed one to use neutron scattering data for constructing models of multicomponent particles. The use of SANS with contrast variation allows one to separate the information about shape and internal structure by changing solvent density and selective labeling furthermore permits to visualize specific structural fragments. The concept of multiphase bead modeling as implemented in the program MONSA represents the particle as a collection of M densely packed beads inside a sphere with the diameter equal to the maximum size of the particle. The bead radius r_0 is selected to yield the total number of beads M of several thousands. To describe a multicomponent complex, a phase index is assigned to each bead corresponding either to the solvent (index = 0) or to one of the components (up to four components are currently supported). The particle is therefore represented at low resolution by a string of length M containing the phase index for each bead. Representing the scattering amplitude of the kth phase in the form $A^{(k)}(\boldsymbol{s}) = \Delta\rho_k A^{(k)}_{\mathrm{un}}(\boldsymbol{s})$ where $\Delta\rho_k$ and $A^{(k)}_{\mathrm{un}}(\boldsymbol{s})$ are its contrast and shape scattering amplitude (with unitary density), respectively, one obtains for the total scattering intensity:

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left\{ \sum_{k=1}^{K} \left[\Delta\rho_k A^{(k)}_{lm}(s)\right]^2 + 2 \sum_{n>k} \Delta\rho_k A^{(k)}_{lm}(s)\, \Delta\rho_n \left[A^{(n)}_{lm}(s)\right]^* \right\}. \quad (5)$$

Here, the partial amplitudes from the volume occupied by the kth phase are

$$A^{(k)}_{lm}(s) = i^l \sqrt{2/\pi} f(s) \sum_{j=1}^{N_k} j_l(sr_j)\, Y^*_{lm}(\omega_j), \quad (6)$$

where the sum runs over the dummy atoms of the kth phase, r_j, ω_j are their polar coordinates, $j_l(x)$ the spherical Bessel function and $f(s)$ is the bead form factor. Using Eqs. 2, 5 and 6, one can compute the scattering curves from a multiphase bead model for an arbitrary configuration of the string and arbitrary contrasts $\Delta\rho_k$.

The penalty terms in Eq. 3 used in the program MONSA ensures low resolution with respect to r_0 and interconnectivity of the entire bead model (Svergun 1999). The interconnectivity is controlled by using graphs, which are defined as a set of beads of the given phase where one can move from any bead to any other bead staying within this phase whereas a move is always made within the first coordination sphere. Starting from a random string, SA is employed to search for a model composed by interconnected compact phases, which simultaneously fits multiple scattering curves in the contrast variation series to minimize the target function. A particular case of this approach, program DAMMIN (Svergun 1999) performs ab initio shape determination (one component, one scattering curve). The difference between the "black and white" shape determination against one data set implemented in DAMMIN and ab initio analysis using contrast variation and multiphase bead model is illustrated in Fig. 1.

Biological macromolecules and their complexes often form symmetric assemblies. The program DAMMIN allows one to account for various types of particle symmetry introduced as a rigid constraint. The symmetry imposes selection rules onto spherical harmonics in Eq. 5, which significantly speeds up the calculations but also allows one to obtain more adequate models of symmetric particles. Symmetry is actively used for shape analysis with DAMMIN (Borges et al. 2005; Goda et al. 2005; Goodridge et al. 2005; Hashimoto et al. 2005; Nagar and Kuriyan 2005; Nixon et al. 2005; Rosenberg et al. 2005; Shi et al. 2005). The multiphase program MONSA was up to now working only in P1, and, to allow for the use of symmetry in SANS analysis, symmetry was included in the last MONSA version (v.142). Now, the generation of the initial model, its randomization and phase modifications by each mutation during the SA search are performed for the group of symmetry related beads simultaneously to continuously keep the specific symmetry of the model. The list of symmetries currently supported in MONSA includes groups Pnk, $n = 1, ..., 6$, $k = 1, ..., 2$. The components, for which interconnectivity is not required, are allowed to contain the number of individual graphs belonging to the correspondent phase that does not exceed nk.

In a SANS experiment, the experimental data is usually smeared because of unfocussed beam and polychromatic (usually up to 10%) radiation used. If required, MONSA smears the theoretical curves computed from the model by the instrument resolution function introduced by (Pedersen et al. 1990), and these smeared curves enter into Eq. 4 for the computation of discrepancy. This allows one to avoid the desmearing step and to work directly with the measured experimental data.

Rigid body modeling

The program SASREF (Petoukhov and Svergun 2005) performs global rigid body modeling of the quaternary structure of macromolecular complexes against solution scattering data. Given that the high resolution models of the individual components of the multisubunit complex

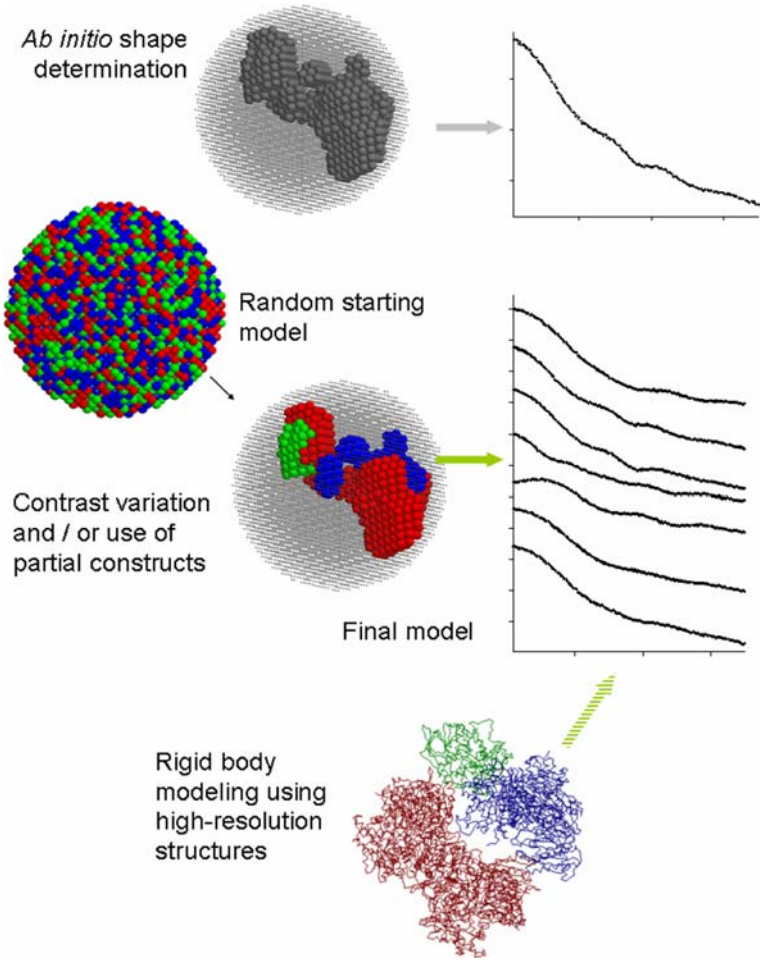

Fig. 1 Comparison of single-phase shape determination against one curve with multiple curve fitting using an ab initio multi-phase model and with rigid body refinement. The positions occupied by the solvent are displayed as *dots*, those occupied by the components of the particle as *beads*

are known and assuming that their tertiary structure remains unchanged upon complex formation, the task of rigid body modeling is to determine their optimal spatial arrangement that yields the best fit to the experimental scattering pattern of the entire complex. SA is employed to find the positions and orientations of the subunits forming interconnected assembly without overlaps while minimizing the discrepancy between the calculated and the experimental scattering profiles. The minimization procedure starts from an arbitrary arrangement of subunits, e.g., from that in a tentative model of the complex or just from all subunits centered at the origin in their reference orientations. It is possible to fix selected subunits at their starting positions and orientations to preserve known sub-structures. A single modification of the assembly is done by rotation of a randomly selected subunit by an arbitrary angle $\phi < \phi_{\max}$ about a rotation axis followed by a random shift $r < r_{\max}$ along an arbitrary direction. If the scattering data sets from the partial constructs are also available the program performs simultaneous data fitting with the assumption of the same architecture in all the constructs, which significantly increases the information content and the reliability of the modeling. The penalty terms (Eq. 3) in SASREF allow one to avoid steric clashes and to ensure interconnectivity of the subunits, and, moreover, the program can account for known interfaces (e.g., binding sites) between subunits by restraining correspondent inter-residue distances [see Petoukhov and Svergun (2005) for more detail].

The contrast variation in SANS is provides valuable additional information about complexes (Koch and Stuhrmann 1979; Wall et al. 2000; Zaccai and Jacrot 1983). The contrasts of the individual subunits can be effectively varied using H_2O/D_2O mixtures and selective deuteration. To fully use the advantages of SANS, the original SAXS-oriented version of SASREF (Petoukhov and Svergun 2005) was extended with the capability of simultaneous fitting of X-ray scattering data and contrast variation neutron series. Our aim was not only to account for the neutron scattering curves recorded at different solvent compositions (volume fractions of D_2O) but also for possible perdeuteration of specific subunits in distinct series of measurements. Of course, it has to be assumed that deuteration does not affect the complex formation, but the H/D equilibrium substitution in the exchangeable groups and the hydration effects have to be taken into account. For adequate computation of the scattering from the high resolution structures, programs CRYSOL (Svergun et al. 1995) for

the X-ray case and CRYSON (Svergun et al. 1998) for neutrons are used.

To compute the X-ray curves and neutron scattering profiles from the atomic models for arbitrary solvent composition and subunits perdeuteration, four basic scattering amplitudes ($\boldsymbol{A}^{(k)}$, $\boldsymbol{B}^{(k)}$, $\boldsymbol{C}^{(k)}$, $\boldsymbol{D}^{(k)}$) are associated with each subunit. These amplitudes depend on the corresponding amplitudes of the given subunit computed at its initial position and orientation ($\boldsymbol{A}_0^{(k)}$, $\boldsymbol{B}_0^{(k)}$, $\boldsymbol{C}_0^{(k)}$, $\boldsymbol{D}_0^{(k)}$) and on six spatial parameters (three rotations and three shifts) describing its displacement in the current configuration as described elsewhere (Svergun 1997; Svergun et al. 1997). Here, $\boldsymbol{A}^{(k)}$ are the X-ray scattering amplitudes, whereas $\boldsymbol{B}^{(k)}$, $\boldsymbol{C}^{(k)}$ and $\boldsymbol{D}^{(k)}$ are neutron scattering amplitudes of protonated subunit in pure water, protonated subunit in 100% D_2O and fully perdeuterated subunit in H_2O, respectively. The resulting neutron scattering amplitude of the subunit for the arbitrary volume fraction $0 \leq x \leq 1$ of the D_2O in solution and subunit perdeuteration $0 \leq y \leq 1$ is computed as

$$E^{(k)}(x,y) = (1-x-y)B^{(k)} + xC^{(k)} + yD^{(k)}. \quad (7)$$

CRYSOL and CRYSON provide the reference scattering amplitudes in terms of the partial amplitudes ($\boldsymbol{A}_0^{(k)}{}_{lm}$, $\boldsymbol{B}_0^{(k)}{}_{lm}$, $\boldsymbol{C}_0^{(k)}{}_{lm}$, $\boldsymbol{D}_0^{(k)}{}_{lm}$), which permit one to rapidly compute the intensities from the given configurations and the overall discrepancy following Eqs. 2–4, and 7. Similarly to MONSA, SASREF includes smearing of the theoretical curves computed from the rigid body model using the resolution function to account for the instrumental effects.

In the presence of symmetry, this information can be taken into account as a hard constrain, which significantly reduces the number of free parameters describing the system. The symmetry groups Pn and $Pn2$ ($n = 1, ..., 6$) are currently supported whereby the n-fold axis and the twofold axis of $Pn2$ coincide with Z and Y, respectively. The present version of SASREF is also able to handle the situation when the sub-structures possess a lower symmetry than the entire complex (e.g., the full structure is a hexamer with P32 symmetry and the set of sub-structures includes trimers and dimers with three and twofold symmetry axes, respectively).

Generation of simulated data for testing

The new versions of programs MONSA and SASREF run on IBM PC-compatible machines under Windows 9x/NT/2000/XP, Linux and Mac OSX as well as on major Unix platforms. The programs were tested on synthetic model examples generated from two nucleoprotein complexes taken from PDB (Berman et al. 2000). Theoretical scattering patterns were generated from available atomic structures and randomized to yield a constant relative error of 3% in each data point, and the complexes were broken into subunits. The X-ray and neutron scattering amplitudes from the subunits were computed using CRYSOL and CRYSON, respectively. The structures of the complexes were restored by fitting multiple scattering patterns by both programs independently to explore their functionality and capabilities. The angular ranges of the synthetic data were taken to be typical for a real SAXS/SANS experiment, $0.1 < s < 1.8$ nm^{-1} for ab initio modeling and $0.1 < s < 3.5$ nm^{-1} for rigid body analysis (given the lower resolution of bead models, a shorter range was used for the ab initio modeling). In all cases, random initial approximations were employed, i.e., the tests were run without knowledge about the actual structure of the complexes.

Results

A symmetric complex

The developed ab initio and rigid body modeling algorithms were first tested on a symmetric complex consisting of one protein subunit and one RNA molecule in asymmetric part. Biologically active molecule of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp) (PDB entry 1asy, Ruff et al. 1991) is a dimer with the two monomers related by a twofold symmetry axis (Fig. 2). The crystallographic dimeric complex has the molecular weight of 160 kDa and contains 490 amino acids and 75 bases per monomer, whereby the dimerization interface is formed by protein subunits. The two X-ray scattering curves from the entire complex and from the dimeric protein alone and the five neutron scattering profiles from the contrast variation series on the complex corresponding to 0, 40, 55, 70 and 100% D_2O in the solvent (Fig. 3) were generated and fitted simultaneously with the P2 symmetry constraint by the two approaches.

Multiple runs of the ab initio program MONSA yielded models where the protein shape was well reconstructed but displayed variations in the appearance of the RNA moiety. In Fig. 2, two typical ab initio models are presented yielding equally good overall fits to the simulated data (Table 1, Fig. 3). The model (a) in the upper row demonstrates good agreement with the high resolution structure. The model (b) in the second row displays correct shapes of the protein and individual RNA's (and even the distance between the centers of the two RNA molecules), but their mutual arrangement is different from that in the original model.

For the rigid body modeling, the atomic structures of monomeric protein and single tRNA were used as two rigid bodies. Similarly to the ab initio case, multiple SASREF reconstructions were performed and compared, and, again, two groups of solutions were discovered providing nearly the same quality of fit to the data. (Table 1, Fig. 3). Both solutions presented in Fig. 1 display correct positions of the subunits in the complex, but only one of them (c) has also proper orientations of all the subunits yielding r.m.s.d. = 0.22 nm

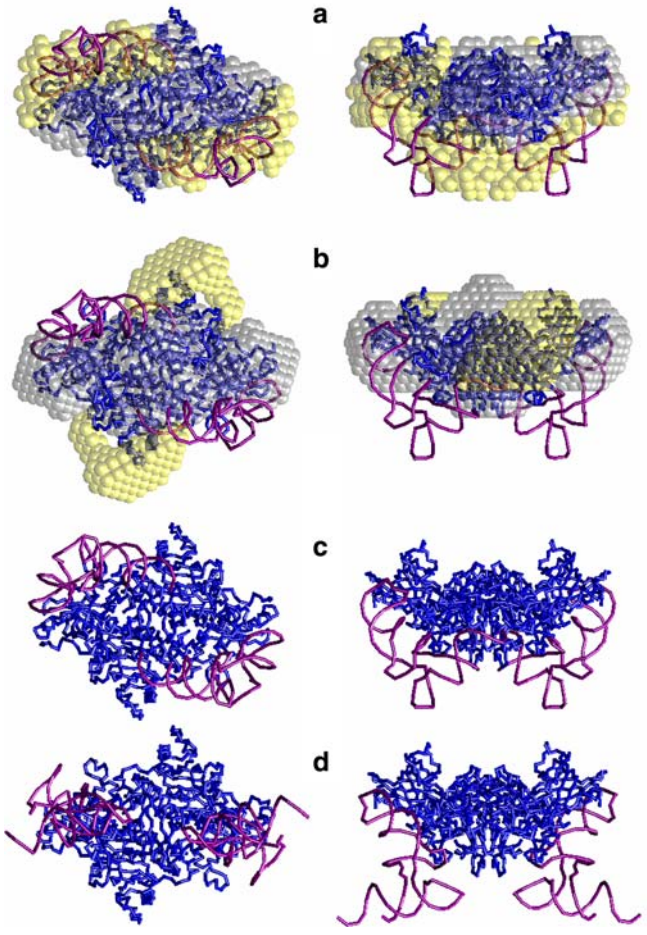

Fig. 2 Aspartyl-tRNA synthetase complexed with tRNA. **a**, **b** Comparisons of the crystal structure with the ab initio bead models generated by MONSA. In the high resolution model, the protein and tRNA are shown as *blue* and *magenta backbones*, in the bead model corresponding phases are presented in *gray* and *yellow*, respectively. **c** Best rigid body model generated by SASREF. **d** A SASREF model with different orientations of tRNA. *Right view* is rotated by 90° about horizontal axis

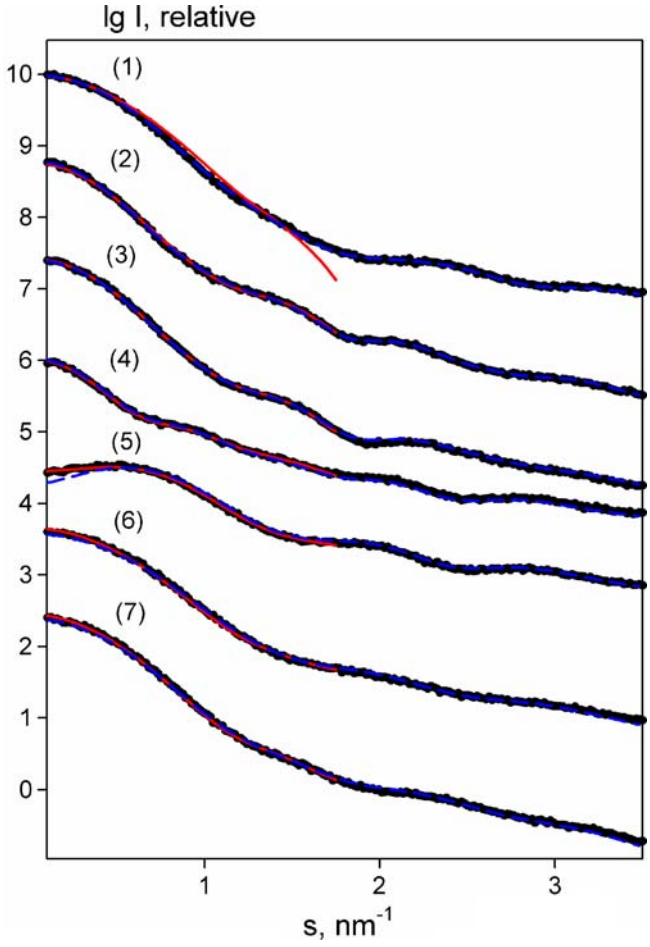

Fig. 3 Scattering profiles from the Aspartyl-tRNA synthetase complex with tRNA. The simulated data are shown by *dots*, the fits obtained by MONSA and SASREF are displayed as *red solid* and *blue dashed lines*, respectively. *1* and *2* are X-ray scattering curves of the dimeric protein and the entire complex, respectively. *3–7* are neutron scattering patterns at 0, 40, 55, 70 and 100% D_2O, respectively. The patterns are displaced in logarithmic scale for better visualization

to the atomic coordinates of the crystal structure. The model (d) reveals a correct architecture of the protein and also the correct overall shape of the complex, but both RNA molecules are flipped within the correct volume.

A complex with selective perdeuteration

In another test example, the complex consisting of one double-stranded DNA and two protein subunits was chosen to demonstrate the efficiency of the selective deuteration of individual components. The T7 DNA polymerase ternary complex with dCTP at the insertion site (PDB entry 1t8e, Brieba et al. 2004) contains polymerase molecule, thioredoxin subunit and the nucleic acid (Fig. 4) with molecular weights of 78, 11 and 14 kDa, respectively. Eleven scattering patterns generated for the rigid body modeling (Fig. 5, curves 4–14), included an X-ray scattering curve from the entire complex, and ten SANS sets at the same D_2O concentrations as for the above example (five neutron profiles for fully protonated complex and five of the complex containing perdeuterated thioredoxin). For the ab initio modeling with MONSA, the three additional SAXS patterns from individual components were also employed (Fig. 5, curves 1–3).

The program MONSA was run several times against 14 scattering curves and yielded reproducible solutions. A typical ab initio 3-phase model (Fig. 4a) provides good fits to the entire set of scattering data (Table 2, Fig. 5). Its comparison with the initial high resolution model in Fig. 4 reveals not only correct reconstruction of the overall shape but also very accurate determination of the internal architecture of the complex.

Table 1 Summary of χ values for Aspartyl-tRNA synthetase complex with tRNA

Model	Data							F
	X-ray		Neutron scattering data from the complex at specified D_2O concentration (%)					
	Protein	Complex	0	40	55	70	100	
MONSA, model (a)	6.76	1.98	1.33	1.21	1.81	4.75	3.72	14.1
MONSA, model (b)	5.30	1.63	1.13	1.05	3.76	6.16	5.09	14.7
SASREF, model (c)	1.35	1.22	1.39	1.37	3.25	1.88	1.92	3.63
SASREF, model (d)	1.33	2.06	1.84	2.05	2.89	1.97	2.38	4.50

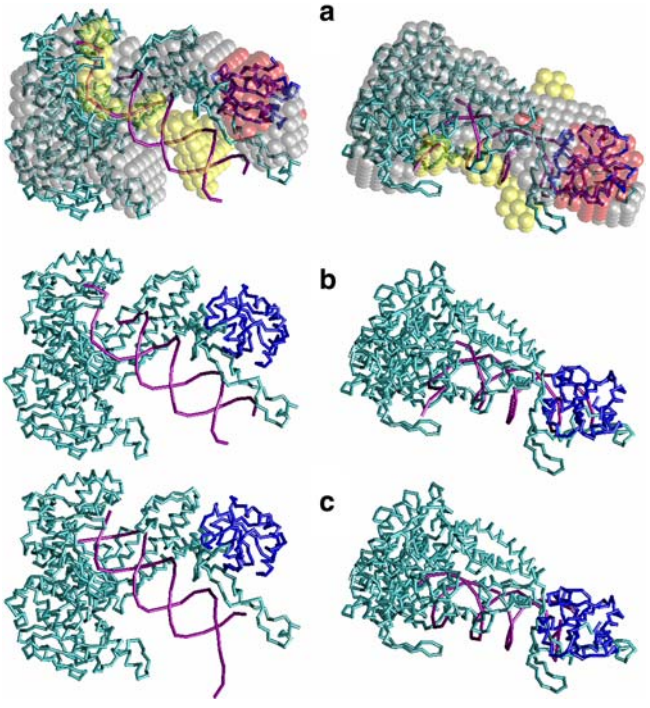

Fig. 4 Ternary complex of T7 DNA polymerase with dCTP at the insertion site. **a** Superposition of the typical ab initio and the crystallographic models. The three-phase MONSA solution is shown as *gray*, *yellow* and *red beads* for polymerase, DNA and thioredoxin, respectively, and these components in the high resolution model are displayed as *cyan*, *magenta* and *blue backbones*, respectively. **b** The best rigid body model generated by SASREF. **c** A SASREF model with the opposite direction of DNA. *Right view* in all panels is rotated by 90° about horizontal axis

Rigid body modeling of the ternary complex was done by SASREF based on the atomic coordinates of the two protein subunits and of the DNA molecule. One and the same rigid body (though with different scattering amplitudes according to Eq. 7) was used to represent protonated and perdeuterated thioredoxin. The multiple reconstructions demonstrated the same architecture of the complex similar to that of the crystallographic model and agree well with the simulated scattering data (Table 2, Fig. 5). The typical rigid body model (Fig. 4b) yields r.m.s.d. = 0.41 nm to the crystallographic coordinates, (where most of the difference comes the from a slightly different orientation of rather globular thioredoxin molecule. Still, there was an ambiguity in DNA orientation, i.e., models with opposite direction of the long axis were also generated by SASREF. Figure 4c displays a typical model with such inverse orientation of DNA (yielding only slightly worse fit (Table 2) to the data compared to the correct model.

Discussion

A set of methods presented here can be used for simultaneous analysis of X-ray and neutron scattering data from complexes of proteins, nucleic acids and other biological macromolecules. The new versions of programs MONSA and SASREF allow one to perform ab initio analysis and to construct rigid body models with minimum user intervention while fitting multiple scattering data sets. Given the limited resolution of SAS data one should however bear in mind the possibility of non-unique solutions compatible with the experimental data. Sometimes in literature methods are presented utilizing even less informative scattering data to construct rather complicated models without sufficient word of caution to the potential users. We have therefore deliberately presented here possible ambiguity of interpretation, even when rich in content contrast variation data are available. Our test results should however not be taken as a proof that unambiguous reconstruction of complexes from SAS is impossible. On the contrary, the use of contrast variation and specific deuteration in neutron scattering together with additional information (symmetry, distance restraints from fluorescence or muagenesis etc.), which is foreseen within the methods presented here, opens the way of modeling macromolecular complexes in solution with unprecedented accuracy.

One of the potential sources of ambiguity is an enanthiomorphous solution. Clearly, a mirror image of the entire structure yields the same scattering intensity, and this ambiguity has always to be accounted for when considering low resolution SAS-generated shapes. For multi-component systems, such an inversion of one of the components may lead to a similar effect, whereby the scattering patterns even at different contrasts of the components may display little changes. Indeed, in this

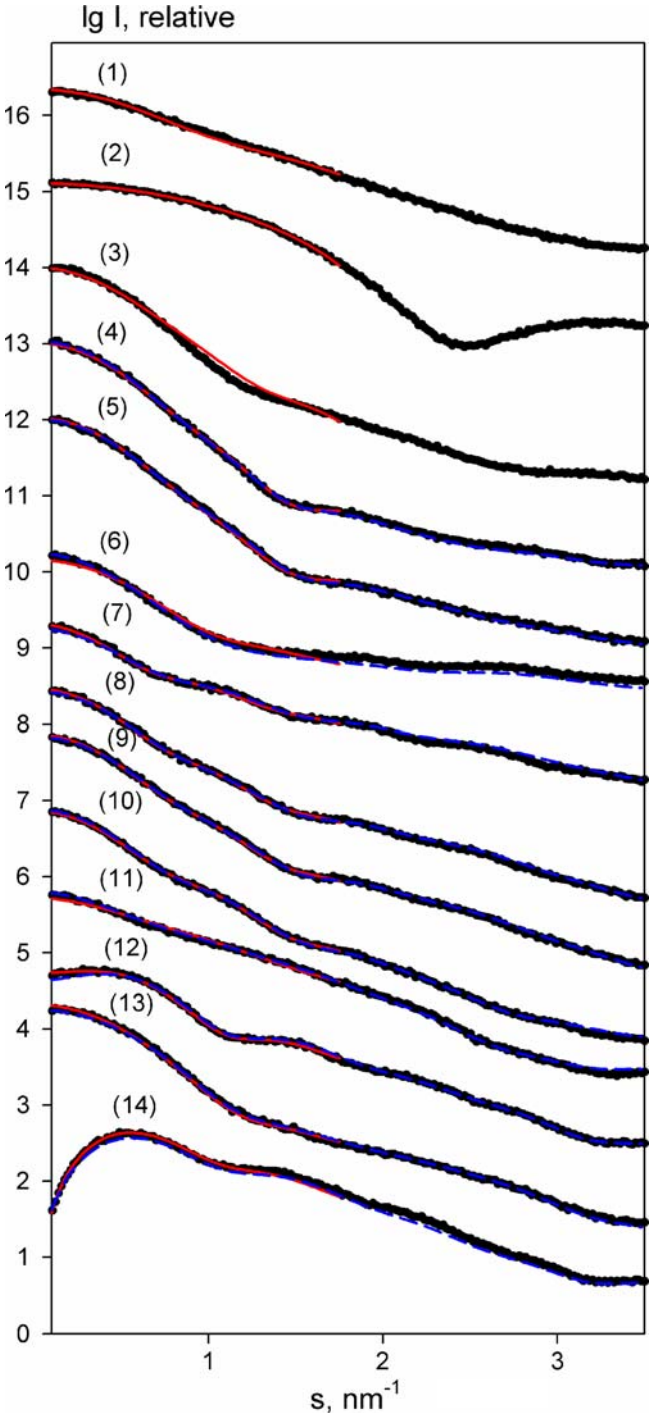

Fig. 5 Scattering profiles from the ternary complex. *1–4* are X-ray scattering curves of the DNA, thioredoxin, polymerase and the entire complex, respectively; *5–9* are neutron scattering patterns from protonated complex in 0, 40, 55, 70 and 100% D_2O, respectively; *10–14* are neutron scattering patterns from the complex with perdeuterated thioredoxin in 0, 40, 70, 100 and 55% D_2O, respectively. Notations of the simulated data and fits are as in Fig. 3. The patterns are displaced in logarithmic scale for better visualization

case, the intensities of the components will remain invariant and only the cross-term would change. These changes can be minimal for certain rotation angles, especially for symmetric structures, if the distance between the centers of components does not change. In particular, in the first example of the symmetric binary complex, taking an enantiomoprph of the tRNA moiety rotated by 180 around the axis perpendicular to the twofold axis leads to the same cross-terms and thus to exactly the same intensities. This leads to the possibility of multiple tRNA arrangements in the first example of the symmetric binary complex displayed in Fig. 2. For ab initio reconstructions, the enanthiomorph-related symmetric arrangements are indistinguishable, but for rigid body modeling false solutions are readily avoided if appropriate additional information is available. Thus, in the example in Fig. 2, requirement of proximity of amino acid residue ASN227 in aspartyl-tRNA synthetase and nucleotide U612 in the tRNA molecule yielded models consistently converging to the correct solutions similar to that in Fig. 2c.

For the second, non-symmetric example on a ternary system with specific deuteration, both ab initio and rigid body approaches allow one to reliably reconstruct the initial structure. The ambiguity in the polarity of the DNA (Fig. 4b, c) was, of course, to be expected, given that at low resolution the molecule is nearly cylindrical, but this issue can be readily resolved, e.g., from DNA footprinting experiments. This example vividly underlines the power of SANS and in particular of specific deuteration. Clearly, no unambiguous structure reconstruction is possible based on X-ray data only, but also with conventional contrast variation on protonated samples, thioredoxin molecule could have not be positioned inside the complex.

The discrepancy values of the fits in Tables 1 and 2 indicate that there are systematic deviations between the simulated data and the scattering computed from the models (most χ values noticeably exceed 1.0). For MONSA solutions, this is explained by the fact that the bead models do not take into account the internal inhomogeneities inside each component. For SASREF solutions, the discrepancy is due to the fact that the hydration shell of the particle is differently generated CRYSOL (or CRYSON) for the entire complex and for the individual components. In the latter case, each component has its own hydration layer, whereas in the former case the layer surrounds the entire particle. The contribution from the shell is relatively small but may play a role at low contrasts (cf. curves at 55% D_2O in Tables 1, 2). Of course, model computations without accounting for the shell lead to the fits to the data with χ around 1 for the correct solution, but we felt it was important to test the program in conditions when systematic deviations are present.

The methods presented here account for a large body of experimental information to construct biochemically sound models of macromolecular complexes. They fit simultaneously X-ray and neutron scattering curves, including contrast variation data sets from selectively deuterated complexes. The models are always interconnected, without steric clashes between the components,

Table 2 Summary of χ values for polymerase ternary complex with DNA and thioredoxin

Model	Data														*F*
	X-ray curves (1 = DNA, 2 = thiotedoxin, 3 = polymerase and 4 = entire complex)				Neutron scattering data from the complex at specified D_2O concentration (%)					Neutron scattering data from the complex with perdeuterated thioredoxin at specified D_2O concentration (%)					
	1	2	3	4	0	40	55	70	100	0	40	55	70	100	
MONSA model (a)	1.89	7.05	4.95	3.02	1.54	3.80	5.66	2.51	1.55	1.13	2.78	2.69	2.25	2.68	16.3
SASREF model (b)	–	–	–	1.63	1.22	1.38	2.17	1.68	1.43	1.37	1.83	11.49	2.14	1.68	14.6
SASREF model (c)	–	–	–	1.49	1.32	1.44	2.11	1.45	1.30	1.50	1.88	11.61	2.19	1.39	14.8

and display the pre-defined symmetry. For rigid body modeling, distance restraints between specified residues/nucleotides or their ranges are readily included as described (Petoukhov and Svergun 2005), which significantly reduces ambiguity of model building. Another way of reducing this ambiguity could also be a "jackknife" approach by performing the modeling against a subset of available scattering patterns leaving out one or several scattering curves. An a posteriori analysis of the fits to the data sets not used for the model building may help to selection the correct solution. Unfortunately, the most efficient way of using the "jackknife" procedure would be to leave out the most sensible curves, i.e. those around the matching point, and these curves are usually the least accurate. The RDC information from NMR (Grishaev et al. 2005; Mattinen et al. 2002), if available, further restrains the mutual orientation of the subunits in rigid body analysis, and this option has also been recently added to SASREF (Konarev et al. 2006).

As usual in experiments employing isotopic substitution, care must also be taken to ensure that the H/D exchange does not significantly alter the structure and/or behavior of macromolecules. Indeed, D_2O is known to affect protein stabilization, vibrational and librational frequencies, and ionic solvation. For proteins, stronger hydration–bond interactions are observed in D_2O, and, moreover, lower solubility of apolar groups is observed in D_2O than in H_2O, which favors the hydrophobic interaction, and thus often increases the tendency to aggregate. In some cases, changes in chemical composition/ionic strength of the buffer may be useful when measuring at higher D_2O concentrations (see, e.g., Svergun et al. 1998). A good way of controlling the integrity of the samples upon isotopic substitution is to perform X-ray scattering measurements as the H/D exchange should have practically no effect on the X-ray pattern.

The joint use of X-rays and neutrons for the structural studies of functional complexes in solution has a rich history in structural biology (e.g., ribosome studies). We hope that the results and tools presented in this manuscript will become useful in future SAXS/SANS structural studies of functional complexes. The programs MONSA and SASREF described here can be downloaded as precompiled executables for all major computer platforms from the URL http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html.

References

Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin AC, Bork P, Superti-Furga G, Serrano L, Russell RB (2004) Structure-based assembly of protein complexes in yeast. Science 303:2026–2029

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Borges JC, Fischer H, Craievich AF, Ramos CHI (2005) Low resolution structural study of two human HSP40 chaperones in solution—DjA1 from subfamily A and DjB4 from subfamily B have different quaternary structures. J Biol Chem 280:13671–13681

Brieba LG, Eichman BF, Kokoska RJ, Doublie S, Kunkel TA, Ellenberger T (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. EMBO J 23:3452–3461

Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. Biophys J 74:2760–75

Engelman DM, Moore PB (1972) A new method for the determination of biological quarternary structure by neutron scattering. Proc Natl Acad Sci USA 69:1997–1999

Feigin LA, Svergun DI (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum, New York

Furtado PB, Whitty PW, Robertson A, Eaton JT, Almogren A, Kerr MA, Woof JM, Perkins SJ (2004) Solution structure determination of monomeric human IgA2 by X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling: a comparison with monomeric human IgA1. J Mol Biol 338:921–941

Goda S, Kojima M, Nishikawa Y, Kujo C, Kawakami R, Kuramitsu S, Sakuraba H, Hiragi Y, Ohshima T (2005) Intersubunit interaction induced by subunit rearrangement is essential for the catalytic activity of the hyperthermophilic glutamate dehydrogenase from Pyrobaculum islandicum. Biochemistry 44:15304–15313

Goodridge HS, Stepek G, Harnett W, Harnett MM (2005) Signalling mechanisms underlying subversion of the immune response by the filarial nematode secreted product ES-62. Immunology 115:296–304

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

Hashimoto H, Shimizu T, Imasaki T, Kato M, Shichijo N, Kita K, Sato M (2005) Crystal structures of type II restriction endonuclease Eco0109I and its complex with cognate DNA. J Biol Chem 280:5605–5610

Heller WT, Finley NL, Dong WJ, Timmins P, Cheung HC, Rosevear PR, Trewhella J (2003) Small-angle neutron scattering with contrast variation reveals spatial relationships between the three subunits in the ternary cardiac troponin complex and the effects of troponin I phosphorylation. Biochemistry 42:7790–7800

Ibel K, Stuhrmann HB (1975) Comparison of neutron and X-ray scattering of dilute myoglobin solutions. J Mol Biol 93:255–265

Ingber L (1993) Simulated annealing: practice versus theory. Math Comp Model 18:29–57

King WA, Stone DB, Timmins PA, Narayanan T, von Brasch AA, Mendelson RA, Curmi PM (2005) Solution structure of the chicken skeletal muscle troponin complex via small-angle neutron and X-ray scattering. J Mol Biol 345(4):797–815

Kirkpatrick S, Gelatt CD Jr, Vecci MP (1983) Optimization by simulated annealing. Science 220:671–680

Koch MHJ, Stuhrmann HB (1979) Neutron-scattering studies of ribosomes. Methods Enzymol 59:670–706

Koch MHJ, Vachette P, Svergun DI (2003) Small angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quart. Rev. Biophys. 36:147–227

Konarev PV, Petoukhov MV, Svergun DI (2001) MASSHA—a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. J. Appl. Crystallogr. 34:527–532

Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. J Appl Crystallogr 39:277–286

Kozin MB, Volkov VV, Svergun DI (1997) ASSA—a program for three-dimensional rendering in solution scattering from biopolymers. J Appl Crystallogr 30:811–815

Krueger JK, Gallagher SC, Wang CA, Trewhella J (2000) Calmodulin remains extended upon binding to smooth muscle caldesmon: a combined small-angle scattering and Fourier transform infrared spectroscopy study. Biochemistry 39:3979–3987

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

Nagar B, Kuriyan J (2005) SAXS and the working protein. Structure 13:169–170

Nixon BT, Yennawar HP, Doucleff M, Pelton JG, Wemmer DE, Krueger S, Kondrashkina E (2005) SAS solution structures of the apo and Mg (2 +)/BeF3-bound receiver domain of DctD, from Sinorhizobium meliloti. Biochemistry 44:13962–13969

Pedersen JS, Posselt D, Mortensen K (1990) Analytical treatment of the resolution function for small-angle scattering. J Appl Crystallogr 23:321–333

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

Press WH, Teukolsky SA, Wetterling WT, Flannery BP (1992) Numerical Recipes. University Press, Cambridge

Rosano C, Zuccotti S, Cobucci-Ponzano B, Mazzone M, Rossi M, Moracci M, Petoukhov MV, Svergun DI, Bolognesi M (2004) Structural characterization of the nonameric assembly of an Archaeal alpha-L-fucosidase by synchrotron small angle X-ray scattering. Biochem Biophys Res Commun 320:176–182

Rosenberg OS, Deindl S, Sung RJ, Nairn AC, Kuriyan J (2005) Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. Cell 123:849–860

Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science 252:1682–1689

Sabatucci A, Vachette P, Beltramini M, Salvato B, Dainese E (2005) Comparative structural analysis of low-molecular mass fragments of Rapana venosa hemocyanin obtained using two different procedures. J Struct Biol 149:127–137

Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. Nature 422:216–225

Shi YY, Hong XG, Wang CC (2005) The C-terminal (331–376) Sequence of *Escherichia coli* DnaJ Is Essential for Dimerization and Chaperone Activity: a small angle X-ray scattering study in solution. J Biol Chem 280:22761–22768

Svergun DI (1997) Restoring three-dimensional structure of biopolymers from solution scattering. J Appl Crystallogr 30:792–797

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

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

Svergun DI, Nierhaus KH (2000) A map of protein-rRNA distribution in the 70 S *Escherichia coli* ribosome. J Biol Chem 275:14432–14439

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

Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB, Barberato C, Koch MHJ (1997) Shape determination from solution scattering of biopolymers. J Appl Crystallogr 30:798–802

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

Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80:2946–2953

Tung CS, Walsh DA, Trewhella J (2002) A structural model of the catalytic subunit-regulatory subunit dimeric complex of the cAMP-dependent protein kinase. J Biol Chem 277:12423–12431

Vanatalu K, Paalme T, Vilu R, Burkhardt N, Junemann R, May R, Ruhl M, Wadzack J, Nierhaus KH (1993) Large-scale preparation of fully deuterated cell components. Ribosomes from *Escherichia coli* with high biological activity. Eur J Biochem 216:315–321

Vigil D, Gallagher SC, Trewhella J, Garcia AE (2001) Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin. Biophys J 80:2082–2092

Wall ME, Gallagher SC, Trewhella J (2000) Large-scale shape changes in proteins and macromolecular complexes. Annu Rev Phys Chem 51:355–380

Walther D, Cohen FE, Doniach S (2000) Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. J Appl Crystallogr 33:350–363

Zaccai G, Jacrot B (1983) Small angle neutron scattering. Annu Rev Biophys Bioeng 12:139–157