Advances in structure analysis using small-angle scattering in solution

Dmitri I. Svergun[1,2)] and Michel H. J. Koch[1)]

[1)] *EMBL, Hamburg Outstation, Notkestraße 85, D-22603 Hamburg, Germany*
[2)] *Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia*

Corresponding author: Dmitri Svergun, EMBL c/o DESY
Notkestrasse 85, D-22603 Hamburg, GERMANY
Tel: +49 40 89902 125, Fax: +49 40 89902 149
E-mail: Svergun@EMBL-Hamburg.DE

Running title: Structure analysis from solution scattering

Keywords: X-ray scattering, neutron scattering, protein structure, *ab initio* methods, rigid body modelling, protein folding, RNA folding

Summary

The resolution and reliability of solution scattering models have been significantly improved by *ab initio* shape and domain structure determination and detailed modelling of macromolecular complexes using rigid body refinement. There is also substantial progress in the quantitative analysis and modelling of assembly and folding processes and intermolecular interactions.

Introduction

Small-angle scattering (SAS) of X-rays and neutrons is a fundamental tool in studies of the solution structure of biological macromolecules with sizes ranging between a few KDa and several MDa, which does not require special sample preparation [1]. This advantage becomes overwhelming in the investigation of large conformational transitions that occur upon binding of effectors or changes in physicochemical conditions and assembly or (un)folding processes.

In a scattering experiment, a solution of macromolecules is exposed to X-rays (with wavelength $\lambda \approx 0.15$ nm) or thermal neutrons ($\lambda \approx 0.5$ nm). The scattered intensity *I(s)* is recorded as a function of momentum transfer s ($s = 4\pi\sin\theta/\lambda$, where 2θ is the angle between the incident and scattered radiation) and the solvent scattering is subtracted. The random positions and orientations of particles result in an isotropic intensity distribution which, for monodisperse solutions of non-interacting particles, is proportional to the scattering from a single particle averaged over all orientations. The net particle scattering is proportional to the squared difference in scattering length density (electron density for X-rays and nuclear/spin density for neutrons) between particle and solvent - the so-called contrast. The latter can be effectively varied in neutron scattering using H_2O/D_2O mixtures or selective deuteration to yield additional information.

The information content of SAS data is qualitatively illustrated in Fig. 1, which represents X-ray patterns from proteins with different folds and molecular masses. At low angles (2-3 nm resolution) the curves are rapidly decaying functions of s essentially determined by the particle shape, which clearly differ. At medium resolution (2 to 0.5 nm) the differences are already less pronounced and above 0.5 nm resolution all curves are very similar. SAS thus contains information about the gross structural features – shape, quaternary and tertiary structure - but is not suitable for the analysis of the atomic structure.

Shape determination

The aim of *ab initio* analysis of SAS data is to recover the three-dimensional structure from the one-dimensional scattering pattern. Except for the trivial case of a spherical particle, the solution is clearly not unique and many different models could yield the same SAS curve. To constrain the solution, homogeneous models are often used. This drastic simplification is reasonable in the analysis of the low-angle part of the scattering patterns from single component particles at sufficiently high contrasts, such as proteins in solution.

In the past, low resolution models were built by trial-and-error using *a priori* information from electron microscopy or other methods. In the first general *ab initio* approach [2] an angular envelope function of the particle $r=F(\omega)$, where (r,ω) are spherical coordinates, is described by a series of spherical harmonics. The low resolution shape is thus defined by a few parameters - the coefficients of this series – which fit the scattering data. This approach was further developed [3] and implemented in the computer program SASHA [4]. It was demonstrated that under certain circumstances a unique envelope can be extracted from the scattering data (of course, up to an enantiomorphic shape – this ambiguity holds for all *ab initio* methods!) [5].

The use of such envelopes is limited to globular particles with relatively simple shapes without significant internal cavities. More detailed models can be constructed *ab initio* using different types of Monte-Carlo searches. Such a search in a confined volume was first implemented in the program DALAI_GA [6,7]. A sphere with diameter equal to the maximum particle size D_{max}, which is directly determined from the scattering data, is filled with a large number, M, of densely packed beads. Each bead belongs either to the particle (index=1) or to the solvent (index=0), and the shape is thus described by a binary string of length M. Starting from a random string, a genetic algorithm searches for a model that fits the data. In a more general approach [8], the beads may belong to different components so that the shape and internal structure of multicomponent particles (e.g. nucleoproteins) can be reconstructed by simultaneously fitting scattering data at different contrasts [*9]. For single component particles, the procedure degenerates to an *ab initio* shape determination. Compactness and connectivity constraints are imposed in the search, implemented in the program DAMMIN. The ‘give-n-take’ procedure SAXS3D [10] and the program SASMODEL [11] based on interconnected ellipsoids are *ab initio* Monte Carlo approaches without limitation of the search space.

The Monte-Carlo based models contain hundreds or thousands parameters, and caution is required to avoid overinterpretation. A common approach is to align a set of models resulting from independent shape reconstruction runs to obtain an average model retaining the most persistent- and conceivably also most reliable- features (e.g. using the program SUPCOMB [12]). Particle symmetry, if known, provides very useful constraints, which can be imposed in the programs SASHA and DAMMIN and in the program GASBOR described below.

Domain structure analysis

For all its simplicity the assumption of particle homogeneity is a severe limitation. A new approach [**13] represents a protein by an assembly of dummy residues (DRs) and uses simulated annealing to build a locally "chain-compatible" DR-model inside a sphere of diameter D_{max}. Compared to *ab initio* shape determination, DR-modelling (implemented in the program GASBOR) employs fewer free parameters to account for more experimental information. Fig. 2 presents *ab initio* models of bovine serum albumin reconstructed using different methods. Comparison with the atomic model of a homologous protein indicates that all methods adequately represent the low resolution structure. The less detailed shape models only fit the scattering at very low angles, whereas the more detailed DR model neatly fits the entire curve up to 0.5 nm resolution.

Adding missing loops or domains

Inherent flexibility and conformational heterogeneity in proteins often result in the absence of loops and even entire domains in crystallographic or NMR models. Such missing fragments still contribute to the SAS intensity and their probable configurations can be found by fixing the known part of the structure and adding the missing parts to fit the SAS pattern from the entire particle. The DR approach has recently been extended and algorithms for adding missing loops or domains are implemented in the program suite CREDO (Petoukhov, Eady, Brown and Svergun, unpublished).

Rigid body modelling

Solution scattering patterns of multi-domain proteins and macromolecular complexes can also be fitted using models built from high resolution structures of individual domains or subunits assuming that their tertiary structure is preserved. The 'automated constrained fit' procedure [14] generates thousands of possible bead models in an exhaustive search for the best fit. In another approach [15,16], the domains are first represented by triaxial ellipsoids to find an approximate arrangement and the atomic models are positioned based on information from other methods. An alternative set of modelling tools operates directly on atomic models using spherical harmonics to accurately compute the scattering from individual domains (programs CRYSOL [17] for X-rays and CRYSON [18] for neutrons. Algorithms for rapid computation of the scattering from the complex [19] are coupled with display and manipulation programs (ASSA [20] and MASSHA [*21]) for interactive and/or automated fitting of the experimental data.

Recent applications

Ab initio methods are now routinely used by several research groups. A non-exhaustive list of examples includes a MoFe protein of *Klebsiella pneumoniae* nitrogenase [*22], structural changes in *Manduca sexta* midgut V1 ATPase [*23]; the P-end-capping protein of actin-tropomyosin [24], the effects of ligand binding to retinoic acid receptors [*25], the sap1 protein from Schizosaccharomyces pombe [26], the C-propeptide domain from human procollagen [*27] and the apo form of *Escherichia coli* haemoglobin protease [28]. Shape determination methods are also applicable to nucleic acids [29].

Applications of rigid body modelling include analysis of similarities between crystal and solution structures of diphosphate-dependent enzymes [*30], the HslUV protease-chaperone complex [*31], and the *E. coli* ATCase [32,*33]. A novel quaternary structure interface for a dimeric α-crystalline domain was proposed [34] and a folded-back model of monomeric factor H of human complement was deduced from X-ray and neutron SAS and ultracentrifugation [35]. A combination of NMR, X-ray SAS and molecular dynamics simulations revealed large differences between the crystal and solution structure of an adaptor protein Grb2 [36]. A model of heterodimeric cAMP-dependent protein kinase was constructed using homology modelling and energy minimization [**37]. *Ab initio* and rigid body modelling were applied to the bacteriophage PRD1 vertex complex [*38].

X-ray SAS patterns have recently been used to correct the Fourier amplitudes in cryo-electron microscopy [39-41] and to constrain the models of protein fold built by structure prediction algorithms [42]. *Ab initio* envelopes have been employed for phasing low resolution reflections in protein crystallography [43,44], and more detailed DR-models should be even more useful to this end.

Assembly and Folding - progress in time-resolved scattering

The more conventional forms of time resolved SAS are becoming increasingly popular in assembly studies of viruses or viral capsids (e.g. [45]) but the main driving force behind the developments is now provided by the folding problem.

Nearly every classical perturbation technique has been combined with SAS and brought to bear on this problem in recent years: fast mixing with stopped or continuous flow to monitor the effect of jumps in denaturant [46], salt [*47] or H^+ concentration [*48], pressure [49] and light intensity [50]. Time resolutions down to the sub-ms range have been reached using undulator radiation with microfabricated mixing devices [46, *48].

Equilibrium studies are, however, still important. One of the most thorough ones during the period of review is that of thermal unfolding of neocarzinostatin (NCS) [**51]. It clearly illustrates the shortcomings of the Guinier analysis used in many recent papers (but not some of the earlier ones e.g. [52]!) on (un)folding of biopolymers. Changes in the forward scattering I(0) observed in several systems [46, *48] even in equilibrium

conditions [**51] are usually attributed to variations of the partial specific volume or solvation during (un)folding but this may not be the only cause. Numerical techniques like singular value decomposition allow one to determine the presence of intermediates [**51,53]. In several cases thermodynamic considerations indicate that the folding intermediates do not represent thermodynamically well-defined states and unfolded states differ depending on whether they have been obtained by thermal or chemical denaturation [**51,54].

The unfolding pathways in a given solvent and the unfolded states in different solvents are thus clearly not unique. Moreover, the number of intermediate states – or ensembles of intermediates- depends on the observation methods, hence the need to use complementary techniques [55]. Equilibrium studies have been made on a number of systems in recent years and pressure induced effects on denaturation and assembly have been investigated on lactate dehydrogenase [56] and staphylococcal nuclease [57].

Denaturant-induced unfolding of NCS monitored with different techniques yields sigmoid curves, which can in first approximation be described by a two-state process [54] but correspond to very different Gibbs free energies of unfolding, indicative of multiple transitions. In the case of cytochrome c, which contains mainly α-helices, the mechanism for pH-induced unfolding has two intermediate states [*48]. Two ensembles of unfolded states had also been found previously in an equilibrium denaturation study with guanidine hydrochloride [58].

An area related to protein folding concerns the formation of amyloid-like cross-β structures (reviewed in [59]). The process has been studied on phosphoglycerate kinase [60] and more medically relevant examples like immunoglobulin light chains [61].

Studies on nucleic acid folding are also rapidly progressing. Following initial studies on tRNAPhe and Rnase P RNA from *Bacillus subtilis* [62] and on the Tetrahymena group I ribozyme [63,64] a model of the global folding process of the ribozyme covering five orders of magnitude in time was obtained from X-ray scattering and computer simulations [*47,*65]). The RNA forms non-specifically collapsed intermediate(s), which are the analogue of the molten globule intermediates in protein folding, and searches for its tertiary contacts in this highly restricted conformational sub-space. This is perhaps a hint for *in silico* folders.

The individual observations made so far on various proteins and nucleic acids suggest that some general rules are emerging and this emphasizes the need for a general theoretical framework for the discussion of these phenomena, which despite genuine progress, is still missing. This should provide great opportunities for the more theoretically minded in coming years.

Conclusions

Many successful applications of new data analysis methods during recent years have further illustrated how much the popular view that solution scattering yields only the

molecular mass and radius of gyration of a particle is a misconception. SAS is also an increasingly important tool in the study of intermolecular interactions in relation to crystallization [*66] and of - perhaps less prestigious, but not less important - biomolecules like lipids and polysaccharides and their assemblies.

The programs SASHA, DAMMIN, GASBOR, CREDO, CRYSOL, CRYSON, ASSA, MASSHA, SUPCOMB can be downloaded from www.embl-hamburg.de/ExternalInfo/Research/Sax/. The program DALAI_GA can be obtained from akilonia.cib.csic.es/DALAI_GA2/, the program SAXS3D from www.cmpharm.ucsf.edu/~walther/saxs/.

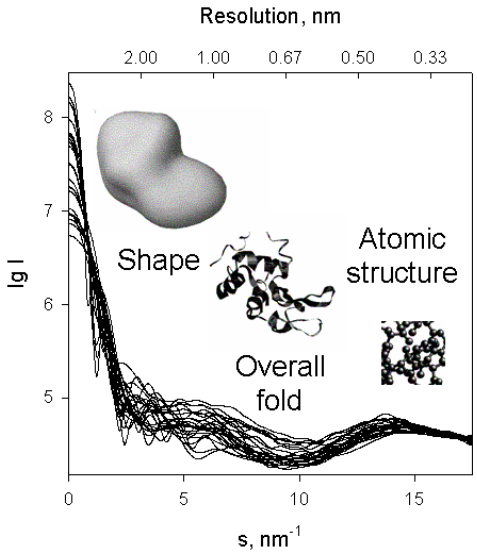

Fig. 1. X-ray solution scattering curves computed from atomic models of twenty-five different proteins with molecular masses between 10 and 300 KDa. The scattering intensities are plotted on a logarithmic scale, the upper axis displays the spatial resolution $\Delta=2\pi/s$ and the labels indicate the levels of structural organization characteristic of this resolution.

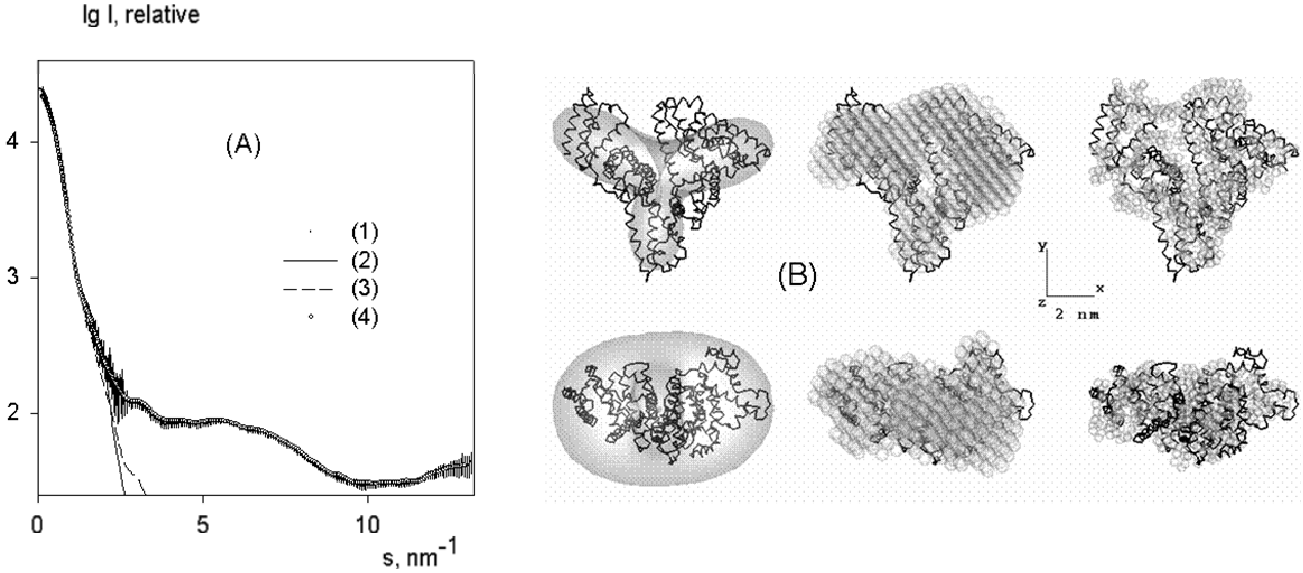

Fig. 2. (A) Synchrotron X-ray scattering from BSA (1) and scattering from the *ab initio* models: (2) envelope model (SASHA); (3) bead model (DAMMIN); (4) dummy residue model (GASBOR). (B) Atomic model of ligand-bound HSA (C_α chain, Protein Data Bank entry 1bke [67]) superimposed to the models of BSA obtained by SASHA (left column, semi-transparent envelope), DAMMIN (middle column, semi-transparent beads) and GASBOR (right column, semi-transparent dummy residues). The models superimposed by SUPCOMB [12] were displayed on an SGI Workstation using ASSA [20]. The middle and bottom rows are rotated counterclockwise by 90^0 around X and Y, respectively.

References

1. Feigin LA, Svergun DI: *Structure analysis by small-angle x-ray and neutron scattering*. New York: Plenum Press; 1987.
2. Stuhrmann HB: **Ein neues Verfahren zur Bestimmung der Oberflaechenform und der inneren Struktur von geloesten globularen Proteinen aus Roentgenkleinwinkelmessungen.** *Zeitschr. Physik. Chem. Neue Folge* 1970, **72**:177-198.
3. Svergun DI, Stuhrmann HB: **New developments in direct shape determination from small-angle scattering 1. Theory and model calculations.** *Acta Crystallogr.* 1991, **A47**:736-744.
4. Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB, Barberato C, Koch MHJ: **Shape determination from solution scattering of biopolymers**. *J. Appl. Crystallogr.* 1997, **30**:798-802.
5. Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB: **New developments in direct shape determination from small-angle scattering 2. Uniqueness.** *Acta Crystallogr.* 1996, **A52**:419-426.
6. Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM: **Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm**. *Biophys J* 1998, **74**:2760-2775.
7. Chacon P, Diaz JF, Moran F, Andreu JM: **Reconstruction of protein form with X-ray solution scattering and a genetic algorithm**. *J Mol Biol* 2000, **299**:1289-1302.
8. Svergun DI: **Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing**. *Biophys J* 1999, **76**:2879-2886.
*9. Svergun DI, Nierhaus KH: **A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome**. *J Biol Chem* 2000, **275**:14432-14439.

First application of simulated annealing to the analysis of contrast variation data from a multicomponent macromolecular complex consisting of 42 X-ray and neutron solution scattering curves from reconstituted *E. coli* ribosomes, where the proteins and rRNA moieties in the subunits were either protonated or deuterated in all possible combinations. The search volume defined by a cryo-electron microscopic model of the ribosome is divided into almost 8000 densely packed 0.5 nm radius spheres. Each sphere is assigned to either solvent, protein or ribosomal RNA (rRNA) moieties to simultaneously fit all the scattering curves. The resulting 3 nm resolution model represents the volumes occupied by the rRNA and protein moieties in the entire ribosome. The protein-rRNA map predicted from small-angle scattering is in remarkably good agreement with the later high-resolution crystallographic models of the ribosomal subunits from other species.

10. Walther D, Cohen FE, Doniach S: **Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules**. *J. Appl. Crystallogr.* 2000, **33**:350-363.
11. Vigil D, Gallagher SC, Trewhella J, Garcia AE: **Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin**. *Biophys J* 2001, **80**:2082-2092.
12. Kozin MB, Svergun DI: **Automated matching of high- and low-resolution structural models**. *J. Appl. Crystallogr.* 2001, **34**:33-41.

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

The *ab initio* approach allowing one to build protein models using X-ray SAS data up to 0.5 nm resolution. The search model is an assembly of dummy residues centred at the C_α positions. The number of DR is usually known from the DNA or protein sequence, whereas their coordinates are initially randomly distributed inside a sphere of diameter D_{max}. During simulated annealing a randomly chosen DR is moved to a point 0.38 nm (typical distance between adjacent residues) away from another randomly selected DR. The final model is constrained to be locally "chain-compatible" (in particular, each DR should have two neighbours at a distance of 0.38 nm). The efficiency of the method is illustrated by the reconstruction of several proteins, with known and unknown crystal structure, from experimental scattering data. The method substantially improves the resolution and reliability of models compared to *ab initio* shape determination.

14. Boehm MK, Woof JM, Kerr MA, Perkins SJ: **The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling**. *J Mol Biol* 1999, **286**:1421-1447.

15. Krueger JK, Zhi G, Stull JT, Trewhella J: **Neutron-scattering studies reveal further details of the Ca2+/calmodulin-dependent activation mechanism of myosin light chain kinase**. *Biochemistry* 1998, **37**:13997-14004.

16. Wall ME, Gallagher SC, Trewhella J: **Large-Scale Shape Changes in Proteins and Macromolecular Complexes**. *Annu Rev Phys Chem* 2000, **51**:355-380.

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

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

19. Svergun DI: **Solution scattering from biopolymers: advanced contrast variation data analysis**. *Acta Crystallogr.* 1994, **A50**:391-402.

20. Kozin MB, Svergun DI: **A software system for automated and interactive rigid body modeling of solution scattering data**. *J. Appl. Crystallogr.* 2000, **33**:775-777.

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

The Wintel-based program MASSHA for three-dimensional rendering and rigid body refinement is similar to the UNIX-based program ASSA [20]. It allows display and manipulation of high resolution atomic structures and low resolution models represented as smooth envelopes or ensembles of beads. The program is coupled with computational modules for advanced modelling, such as alignment of different types of models by invoking the program SUPCOMB [12]. MASSHA provides both interactive and automated refinement to position subunits in a heterodimeric complex (general case) and in a homodimeric complex with a two-fold symmetry axis.

*22. Grossmann JG, Hasnain SS, Yousafzai FK, Eady RR: **Evidence for the selective population of FeMoco sites in MoFe protein and its molecular recognition by**

the Fe protein in transition-state complex analogues of nitrogenase. *J Biol Chem* 2000.

This study illustrates that low resolution models provide useful information about the relationship between structure and function. The *ab initio* angular envelopes indicate that the conformation of the MoFe protein containing one Mo-atom per $\alpha_2\beta_2$ tetramer differs from that of the protein with full occupancy (two Mo-atoms per molecule). This is consistent with the existence of MoFe protein molecules that contain only one occupied FeMo cofactor site and have 50% lower specific activity. The restored shape of a complex of MoFe protein lacking one of the cofactors with the Fe protein points to a 1:1 stoichiometry and suggests that the Fe protein interacts with the FeMo cofactor-binding α-subunit of the MoFe protein. The authors conclude that the conformation of the α-subunit or the αβ subunit pair that lacks the FeMo cofactor is altered and that the change is recognized by the Fe protein.

*23. Grueber G, Svergun DI, Godovac-Zimmermann J, Harvey WR, Wieczorek H, Koch MHJ: **Evidence for major structural changes in the manduca sexta midgut V1 ATPase due to redox modulation. A small-angle x-ray scattering study**. *J Biol Chem* 2000, **275**:30082-30087.

An example of *ab initio* shape reconstruction by simulated annealing with symmetry restrictions (a three-fold symmetry axis). The mushroom-like low resolution shapes of the oxidized and reduced form of the enzyme deduced from X-ray SAS indicate that the main structural alteration occurs in the head piece, where the major subunits A and B are located, and at the bottom of the stalk. The existence of a major structural change due to redox modulation is corroborated by a lower susceptibility to tryptic digestion and tryptophan fluorescence of the reduced V_1 ATPase.

24. Fujisawa T, Kostyukova A, Maeda Y: **The shapes and sizes of two domains of tropomodulin, the P-end-capping protein of actin-tropomyosin**. *FEBS Lett* 2001, **498**:67-71.

*25. Egea PF, Rochel N, Birck C, Vachette P, Timmins PA, Moras D: **Effects of Ligand Binding on the Association Properties and Conformation in Solution of Retinoic Acid Receptors RXR and RAR.** *J. Mol. Biol.* 2001, **307**:557-576.

An extensive biophysical solution study of the ligand-binding domains (LBD) of retinoid nuclear receptors RARα and RXRα using analytical ultracentrifugation, gel-filtration and X-ray and neutron scattering. The effect of retinoic acid (RA) binding on the association properties and conformation is analyzed. The shapes of monomeric RARα and RXRα LBDs restored using simulated annealing correlate well with the available crystal structures of RXRs and RARs LBDs. Three conformational states are observed in RARα/RXRα LBD heterodimers depending on the progressive binding of the RA stereoisomers. Unliganded RXRα LBD forms homodimers and homotetramers in solution and the model of the latter is constructed *ab initio* and by rigid body refinement.

26. Bada M, Walther D, Arcangioli B, Doniach S, Delarue M: **Solution structural studies and low-resolution model of the Schizosaccharomyces pombe sap1 protein**. *J Mol Biol* 2000, **300**:563-574.

*27. Bernocco S, Finet S, Ebel C, Eichenberger D, Mazzorana M, Farjanel J, Hulmes DJ: **Biophysical characterization of the C-propeptide trimer from human procollagen III reveals a tri-lobed structure**. *J Biol Chem* 2001, **276**:48930-48936.

The shape of the recombinant C-propeptide trimer is reconstructed *ab initio* using angular envelopes and a genetic algorithm. Both techniques yield an elongated cruciform model with three large lobes and a minor lobe. Based on these shapes, a model is proposed for the C-terminal region of the procollagen molecule.

28. Scott DJ, Grossmann JG, Tame JR, Byron O, Wilson KS, Otto BR: **Low resolution solution structure of the Apo form of Escherichia coli haemoglobin protease Hbp**. *J Mol Biol* 2002, **315**:1179-1187.
29. Funari SS, Rapp G, Perbandt M, Dierks K, Vallazza M, Betzel C, Erdmann VA, Svergun DI: **Structure of free thermus flavus 5 S rRNA at 1.3 nm resolution from synchrotron X-ray solution scattering**. *J Biol Chem* 2000, **275**:31283-31288.

*30. Svergun DI, Petoukhov MV, Koch MHJ, Koenig S: **Crystal versus solution structures of thiamine diphosphate-dependent enzymes**. *J. Biol. Chem.* 2000, **275**:297-302.

A systematic study of differences between the quaternary crystal and solution structures of the thiamine diphosphate-dependent enzymes. The x-ray scattering data collected from five multisubunit enzymes in the crystallization buffers are compared with the curves computed from the available crystallographic structures. For all enzymes except the very compact pyruvate decarboxylase from *Z. mobili*s, differences are observed between the experimental and calculated profiles. The changes in relative position of the subunits in solution are determined by rigid body refinement. For pyruvate oxidase from *L. plantarum* and transketolase from *S. cerevisiae*, which have tight intersubunit contacts in the crystal, relatively small modifications of the quaternary structure with a root mean square displacement (rmsd) of about 0.25 nm suffice to fit the experimental data. For the enzymes with looser contacts (the native and activated forms of yeast pyruvate decarboxylase), modifications of the crystallographic models with rmsd up to 1.5 nm are required. The magnitude of the distortions induced by the crystal environment is thus correlated with the interfacial area between subunits.

*31. Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB: **Crystal and solution structures of an HslUV protease-chaperone complex**. *Cell* 2000, **103**:633-643.

The paper presents and validates the crystal structure a “prokaryotic proteasome” HslUV from *H. influenzae* composed of the HslV protease and the HslU ATPase at 0.34 nm resolution. Two hexameric ATP binding rings of HslU bind to the opposite sides of the HslV protease whereas the HslU intermediate domains extend outward from the complex. The scattering calculated from the crystallographic model gives a good fit to the experimental X-ray SAS data recorded from solutions under conditions where the complex is assembled and active. The fit can further be improved by adding structural fragments from the crystallographic model of HslU (about 50 residues missing in the current model). It is interesting that the scattering computed from the earlier atomic model of *E.coli* HslUV in the crystal (showing different arrangement of subunits) fails to fit the SAS pattern.

32. Svergun DI, Barberato C, Koch MHJ, Fetler L, Vachette P: **Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state**. *Proteins* 1997, **27**:110-117.

*33. Fetler L, Vachette P: **The allosteric activator Mg-ATP modifies the quaternary structure of the R-state of *E.coli* Aspartate Transcarbamylase without altering the T<–>R equilibrium.** *J. Mol. Biol.* 2001, **309**:817-832.

The quaternary structure of the R-state of *E. coli* ATCase, a paradigm of allosteric enzymes, has previously been shown to display large differences between the solution and the crystal [32]. The present paper demonstrates that Mg-ATP, an allosteric activator, modifies this quaternary structure into a more expanded conformation, which is modelled using rigid body movements. Interestingly, Mg-ATP binding does not alter the T↔R equilibrium. Beyond rigid body modelling, these results, together with previous studies, shed new light on the molecular mechanisms behind the regulatory properties of ATCase.

34. Feil IK, Malfois M, Hendle J, van Der Zandt H, Svergun DI: **A novel quaternary structure of the dimeric alpha-crystallin domain with chaperone-like activity**. *J Biol Chem* 2001, **276**:12024-12029.

35. Aslam M, Perkins SJ: **Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling**. *J Mol Biol* 2001, **309**:1117-1138.

36. Yuzawa S, Yokochi M, Hatanaka H, Ogura K, Kataoka M, Miura K, Mandiyan V, Schlessinger J, Inagaki F: **Solution Structure of Grb2 Reveals Extensive Flexibility Necessary for Target Recognition**. *J. Mol. Biol.* 2001, **306**:527-537.

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

An excellent example of detailed model building constrained by low resolution SAS data. Previous neutron scattering analysis of selectively deuterated complexes (reviewed in [16]) yielded a model of the quaternary structure of the holoenzyme. The docking of homology models of the subunits was constrained by the dimensions and positions of the ellipsoids in the neutron-derived model and mutagenesis data. The “best fit” model of the heterodimer was further refined using molecular dynamics and energy minimization to establish the side chain packing at the heterodimer interface. The calculations using the two different homology models predicted similar interfaces. The resulting model provides an explanation for the stability of both the nondissociated protein kinase and its dissociated subunits in aqueous solutions.

*38. Sokolova A, Malfois M, Caldentey J, Svergun DI, Koch MHJJ, Bamford DH, Tuma R: **Solution structure of bacteriophage PRD1 vertex complex**. *J Biol Chem* 2001, **27**:27.

A combination of *ab initio* modelling and rigid body refinement methods in an x-ray study of a large protein assembly. Low resolution models of the vertex proteins P5, P2 and P31 comprising the spike structure of the bacteriophage PRD1 are reconstructed *ab initio* using simulated annealing with symmetry restrictions. Further, rigid body modelling is employed to build the *in vitro* complexes $P5_6$:P31 and $P5_9$. These assemblies are combined in a plausible model of the vertex within the context of the virion. This model is similar to the adenovirus spike assembly but reveals structural differences possibly related to the modes of DNA delivery in PRD1 and adenovirus.

39. Thuman-Commike PA, Tsuruta H, Greene B, Prevelige PE, Jr., King J, Chiu W: **Solution x-ray scattering-based estimation of electron cryomicroscopy**

imaging parameters for reconstruction of virus particles. *Biophys J* 1999, **76**:2249-2261.

40. Gabashvili IS, Agrawal RK, Spahn CM, Grassucci RA, Svergun DI, Frank J, Penczek P: **Solution structure of the E. coli 70S ribosome at 11.5 A resolution**. *Cell* 2000, **100**:537-549.
41. Saad A, Ludtke SJ, Jakana J, Rixon FJ, Tsuruta H, Chiu W: **Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination**. *J Struct Biol* 2001, **133**:32-42.
42. Zheng W, Doniach S: **Protein structure prediction constrained by solution X-ray scattering data and structural homology identification**. *J Mol Biol* 2002, **316**:173-187.
43. Hao Q, Dodd FE, Grossmann JG, Hasnain SS: **Ab initio phasing using molecular envelope from solution X-ray scattering**. *Acta Crystallogr D Biol Crystallogr* 1999, **55**:243-246.
44. Hao Q: **Phasing from an envelope**. *Acta Crystallogr D Biol Crystallogr* 2001, **57**:1410-1414.
45. Canady MA, Tsuruta H, Johnson JE: **Analysis of rapid, large-scale protein quaternary structural changes: time-resolved X-ray solution scattering of Nudaurelia capensis omega virus (NomegaV) maturation**. *J Mol Biol* 2001, **311**:803-814.
46. Pollack L, Tate MW, Finnefrock AC, Kalidas C, Trotter S, Darnton NC, Lurio L, Austin RH, Batt CA, Gruner SM, et al.: **Time Resolved Collapse of a Folding Protein Observed with Small Angle X-Ray Scattering**. *Phys. Rev. Lett.* 2001, **86**:4962-4965.

*47. Russell R, Millett IS, Tate MW, Kwok LW, Nakatani B, Gruner SM, Mochrie SG, Pande V, Doniach S, Herschlag D, et al.: **Rapid compaction during RNA folding**. *Proc Natl Acad Sci U S A* 2002, **99**:4266-4271.

A picture of the global folding of the Tetrahymena ribozyme covering five decades in time is obtained by X-ray SAS and a coarse grained computer model. Substantial compaction occurring in the low ms range yields a kind of molten globule and is followed by overall compaction and shape changes, which are completed within 1s before any tertiary contacts are formed.

*48. Akiyama S, Takahashi S, Kimura T, Ishimori K, Morishima I, Nishikawa Y, Fujisawa T: **Conformational landscape of cytochrome c folding studied by microsecond- resolved small-angle x-ray scattering**. *Proc Natl Acad Sci U S A* 2002, **99**:1329-1334.

Two intermediates are detected by rapid mixing (using a continuous flow device for times shorter than 22 ms, stopped-flow beyond) during the pH-induced refolding of cytochrome c. Intermediate I consists of an extended region and a clustered region formed in less than 160μs by specific collapse around a region of the main chain and the heme. Intermediate II, which forms after about 500 μs, would correspond to the kinetic molten globule with a well-structured core and a small fraction of extended structure and transforms into the folded state in about 15 ms.

49. Woenckhaus J, Kohling R, Thiyagarajan P, Littrell KC, Seifert S, Royer CA, Winter R: **Pressure-jump small-angle x-ray scattering detected kinetics of staphylococcal nuclease folding**. *Biophys J* 2001, **80**:1518-1523.

50. Sasaki J, Kumauchi M, Hamada N, Oka T, Tokunaga F: **Light-induced unfolding of photoactive yellow protein mutant M100L**. *Biochemistry* 2002, **41**:1915-1922.

51. Perez J, Vachette P, Russo D, Desmadril M, Durand D: **Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J Mol Biol* 2001, **308**:721-743.

Both model calculations and experiment indicate that as soon as the protein unfolds the values of the radius of gyration (R_g) obtained by the Guinier approximation are very significantly smaller than those obtained with the Debye formula for an infinitely thin Gaussian chain, which are also preferable to those obtained from the distance distribution function. This does not affect the mostly qualitative conclusions in recent work, but should be taken into account in more quantitative interpretations. The paper contains a good application of singular value decomposition and an interesting thermodynamic analysis.

52. Damaschun G, Damaschun H, Gast K, Misselwitz R, Muller JJ, Pfeil W, Zirwer D: **Cold denaturation-induced conformational changes in phosphoglycerate kinase from yeast**. *Biochemistry* 1993, **32**:7739-7746.

53. Segel DJ, Eliezer D, Uversky V, Fink AL, Hodgson KO, Doniach S: **Transient dimer in the refolding kinetics of cytochrome c characterized by small-angle X-ray scattering**. *Biochemistry* 1999, **38**:15352-15359.

54. Russo D, Durand D, Calmettes P, Desmadril M: **Characterization of the denatured states distribution of neocarzinostatin by small-angle neutron scattering and differential scanning calorimetry**. *Biochemistry* 2001, **40**:3958-3966.

55. Brooks CL, 3rd: **Viewing protein folding from many perspectives**. *Proc Natl Acad Sci U S A* 2002, **99**:1099-1100.

56. Fujisawa T, Kato M, Inoko Y: **Structural characterization of lactate dehydrogenase dissociation under high pressure studied by synchrotron high-pressure small-angle X-ray scattering**. *Biochemistry* 1999, **38**:6411-6418.

57. Panick G, Vidugiris GJ, Malessa R, Rapp G, Winter R, Royer CA: **Exploring the temperature-pressure phase diagram of staphylococcal nuclease**. *Biochemistry* 1999, **38**:4157-4164.

58. Segel DJ, Fink AL, Hodgson KO, Doniach S: **Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c**. *Biochemistry* 1998, **37**:12443-12451.

59. Lynn DG, Meredith SC: **Review: model peptides and the physicochemical approach to beta- amyloids**. *J Struct Biol* 2000, **130**:153-173.

60. Damaschun G, Damaschun H, Fabian H, Gast K, Krober R, Wieske M, Zirwer D: **Conversion of yeast phosphoglycerate kinase into amyloid-like structure**. *Proteins* 2000, **39**:204-211.

61. Khurana R, Gillespie JR, Talapatra A, Minert LJ, Ionescu-Zanetti C, Millett I, Fink AL: **Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates**. *Biochemistry* 2001, **40**:3525-3535.

62. Fang X, Littrell K, Yang XJ, Henderson SJ, Siefert S, Thiyagarajan P, Pan T, Sosnick TR: **Mg2+-dependent compaction and folding of yeast tRNAPhe and the catalytic domain of the B. subtilis RNase P RNA determined by small- angle X-ray scattering**. *Biochemistry* 2000, **39**:11107-11113.

63. Russell R, Millett IS, Doniach S, Herschlag D: **Small angle X-ray scattering reveals a compact intermediate in RNA folding**. *Nat Struct Biol* 2000, **7**:367-370.
64. Fang XW, Golden BL, Littrell K, Shelton V, Thiyagarajan P, Pan T, Sosnick TR: **The thermodynamic origin of the stability of a thermophilic ribozyme**. *Proc Natl Acad Sci U S A* 2001, **98**:4355-4360.
*65. Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D: **Exploring the folding landscape of a structured RNA**. *Proc Natl Acad Sci U S A* 2002, **99**:155-160.
The folding of Tetrahymena ribozyme is followed starting from distinct regions in the folding landscape. The results of fluorescence energy transfer and SAS reveal discrete folding pathways lying in deep channels in the landscape and separated by high energy barriers. Chemical protection and mutagenesis are used to establish the features determining the selection of the pathway.
*66. Casselyn M, Perez J, Tardieu A, Vachette P, Witz J, Delacroix H: **Spherical plant viruses: interactions in solution, phase diagrams and crystallization of brome mosaic virus**. *Acta Crystallogr D Biol Crystallogr* 2001, **57**:1799-1812.
A systematic study of the interactions between brome mosaic virus particles in solution to obtain optimal crystallization conditions.
67. Curry S, Mandelkow H, Brick P, Franks N: **Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites**. *Nat Struct Biol* 1998, **5**:827-835.

This article was published in

Cur. Opin. Struct. Biol. **12**, as Svergun, D. I. & Koch, M. H. J. Advances in structure analysis using small-angle scattering in solution. 654–660, Copyright (2002),

and is posted with permission from Elsevier