

Small angle neutron and X-ray scattering in structural biology: recent examples from the literature

Cameron Neylon

Received: 4 October 2007 / Revised: 20 December 2007 / Accepted: 21 December 2007 / Published online: 23 January 2008
© EBSA 2008

Abstract Small angle scattering can provide unique structural information on the shape, domain organisation, and interactions of biomacromolecules in solution. Small angle neutron scattering (SANS) combined with deuterium labelling makes it possible to define the positions of specific components within a complex while small angle X-ray scattering (SAXS) provides more precise data on the overall shape. Here I review four recent publications, three of which were presented at the Neutrons in Biology meeting at the STFC Rutherford Appleton Laboratory in July 2007, that utilise SANS, SAXS, and complementary techniques to define the solution structure of large multi-domain proteins and macromolecular complexes. These four papers emphasise the critical importance of sample quality and characterisation as well as the important role played by complementary techniques in building structural models based on small angle scattering data. They show the ability of SANS and SAXS in determining solution structures provides an important complementary structural technique for large, flexible, and glycosylated proteins where high resolution structural techniques, such as crystallography and NMR, cannot be applied.

Keywords SANS · SAXS · Structural biology · Solution structure · Rigid body modelling · Ab initio shape reconstruction · Constrained modelling

Introduction

Small angle neutron and X-ray scattering can provide large-scale structural data on molecules in solution. They can generally be regarded as providing information on the size and shape of molecules in solution. The size range that the data reports on is insufficient to allow the placement of amino acid residues or, in most cases, to define secondary structure or the path of the backbone through the protein. However, while the scale of structures that can be determined is large, the precision with which they can be determined is very high. Therefore while small angle scattering is usually described as a low-resolution structural technique it is perhaps more appropriate to describe it as high precision technique for determining large distances. These distances can act as very tight constraints on the modelling and refinement of the three dimensional structure of biomacromolecules.

X-rays and neutrons interact in a fundamentally different way with matter. X-rays interact with the electrons of atoms leading to the well known monotonic increase of X-ray sensitivity to atoms with atomic number. X-rays are much more sensitive to heavy metals than to carbon and nitrogen and (apart from the case of X-ray crystal structures determined to resolutions better than 1.2 Å) hydrogen positions in protein are difficult to determine from X-ray data. Neutrons however interact with the nucleus of atoms in a way, which shows no clear tendency with respect to atomic number. This has two important consequences for small angle scattering. The

Advanced neutron scattering and complementary techniques to study biological systems. Contributions from the meetings, “Neutrons in Biology”, STFC Rutherford Appleton Laboratory, Didcot, UK, 11–13 July and “Proteins At Work 2007”, Perugia, Italy, 28–30 May 2007.

C. Neylon (✉)
Science and Technology Facilities Council Rutherford Appleton
Laboratory, Didcot OX 11 0QX, UK
e-mail: c.neylon@rl.ac.uk

first is that light elements are as visible to neutrons as heavy elements, with carbon, nitrogen, and oxygen all having similar scattering length densities (SLD). The second important factor is that, because neutrons interact with the nucleus they can be sensitive to isotopic substitution. For nitrogen and carbon the differences between common isotopes are small and have not been exploited experimentally. However for the hydrogen isotopes, hydrogen and deuterium, there is a very large difference in SLD.

Hydrogen has a *negative* SLD whereas deuterium has a strongly positive SLD, similar to that of other important biological elements. The negative SLD of hydrogen, along with the fact that the solvent for biological experiments is generally water means that the SLD of the solvent can be varied in a systematic manner from a value of -0.6 to $6.4 \times 10^{10} \text{ cm}^{-2}$. This range encompasses the SLD of the majority of hydrogenated biological molecules as well as that generally reached by proteins deuterated by expression in minimal media in D_2O with a hydrogenated carbon source. As the intensity of small angle scattering is determined largely by the difference in SLD between solute molecule and solvent, the contrast, it is therefore possible to largely remove the contribution of specific molecules to the experimental scattering pattern, in effect making them ‘disappear’. Furthermore by deuterating one component of a complex system it is possible to examine just that component by ‘matching out’ the remaining parts, or to determine the scattering from different components using data obtained at a set of different contrasts.

Small angle neutron scattering therefore has immense potential in dissecting the structure of biological systems with multiple components, either multiple domain proteins, or protein–protein complexes. However, due to challenges in preparing high-quality samples, obtaining high quality data and the difficulties of data analysis SANS has made a relatively limited contribution to structural biology compared to other techniques. Four recent papers (Bonner et al. 2007; Callow et al. 2007; Comoletti et al. 2007; Whitten et al. 2007), three of them presented at the Neutrons in Biology meeting held at the Rutherford Appleton Laboratory in July 2007 illustrate the growing body of work exploiting the potential of SANS, generally in combination with SAXS, to tackle significant and challenging problems in structural biology and point the way forward for the expansion of the field as new instruments and data analysis approaches are developed. For more general coverage of the area the reader is referred to other recent reviews (Trehwella 2006; Petoukhov and Svergun 2006, 2007; Stuhmann 2004).

Importance of sample quality and characterisation

Small angle scattering (SAS) is a weak effect and samples for SANS generally need to be at concentrations greater than 3 mg mL^{-1} to obtain data of sufficient quality for the kinds of structural analysis described in the papers reviewed here. Samples volumes of around $200\text{--}500 \mu\text{L}$ are required making the minimum samples requirements for SANS around $1\text{--}10 \text{ mg}$. X-ray scattering is more sensitive; however, samples of $\sim 1 \text{ mg mL}^{-1}$ are still required. The requirement for high sample concentration brings with it two potential problems which must be addressed before full data analysis. The first is the problem of aggregation. SAS is very sensitive to large particles. Thus small amounts of aggregate in a sample can have a significant effect on the scattering pattern. Aggregation therefore needs to be rigorously excluded. This is usually best demonstrated by complementary analytical techniques such as analytical ultracentrifugation, light scattering, or gel filtration chromatography. Most samples for small angle scattering analysis are purified by gel filtration in the final stage. Additional characterisation by light scattering and AUC are commonly used. Evidence for lack of aggregation can also be provided for the solution scattering data itself via linearity of the low Q region of the data in a Guinier plot ($\log_e(I)$ vs. Q^2). This plot is very sensitive to aggregation or lack of monodispersity. However, while this is a necessary piece of evidence for ruling out aggregation it is not sufficient to demonstrate it.

The second potential problem for high concentration samples is that of interaction between molecules in solution. At high concentrations interactions between molecules will impose a supramolecular structure on the solution, which will be detected in small angle scattering. In general this will lead to a suppression of low angle scattering and an underestimation of the size of system. Whitten et al, and Comoletti et al. emphasise strongly the importance of excluding this possibility. Comoletti et al. describe in detail the use of concentration series to rigorously exclude interparticle interactions. Their approach is to take data for a series of concentrations, determine the R_g from a Guinier plot and then extrapolate the R_g to infinite dilution. This is done on a lab-based X-ray instrument. The conditions utilised for synchrotron or neutron scattering experiments are those where the determined R_g is insignificantly different to that at infinite dilution.

An important aspect of the experiment which is often not considered is the approach for determining sample concentration. If the sample is known to be rigorously monodisperse and the molecular weight is known then the concentration can be determined via the extrapolated intensity at zero angle. However, it is more common that

this region is used to provide evidence that the sample is monodisperse or to identify the possibility of oligomerisation of the protein. For this to be successful it is critical that the protein concentration is known accurately. Colorimetric assays such as the Bradford assay are generally inadequate for this. Accurate determination of concentration via absorbance at 280 nm is also more challenging than is often assumed and requires the protein to contain a reasonable number of tryptophan residues (Gill and Hippel 1989; Pace et al. 1995; Perkins 1986). It is noteworthy that Comoletti et al. resort to amino acid analysis to obtain an accurate concentration for the neuroligin–neurexin complexes. See also (Ashton et al. 1997) for the confusion that can arise when absorption coefficients do not behave as expected.

Contrasting data analysis for SAS

Small angle scattering of biomacromolecules is generally carried out in solution samples with all molecules within the sample oriented at random. While data is often collected on a two dimensional detector it is then radially averaged (after removing any artefacts due to the beam stop or other instrument factors) to generate a one-dimensional data set. This one-dimensional data is then used to attempt to reconstruct a three dimensional structure. It is important to realise that this is mathematically impossible without further information. Multiple three-dimensional structures can give rise to the same one-dimensional scattering pattern. Therefore further information is required for three-dimensional reconstruction. All data analysis approaches therefore use further information, either structural information from other techniques, or make assumptions about the type of structures possible when fitting the data. In the case of rigid body modelling existing structures are used which are then allowed to move with respect to each other. Where *ab initio* structure reconstruction approaches are used constraints are applied such as compactness and connectivity, which are assumed to be characteristic of most proteins. This places limits on the use of *ab initio* approaches to determine the structure of non-compact biomolecules.

Rigid body modelling

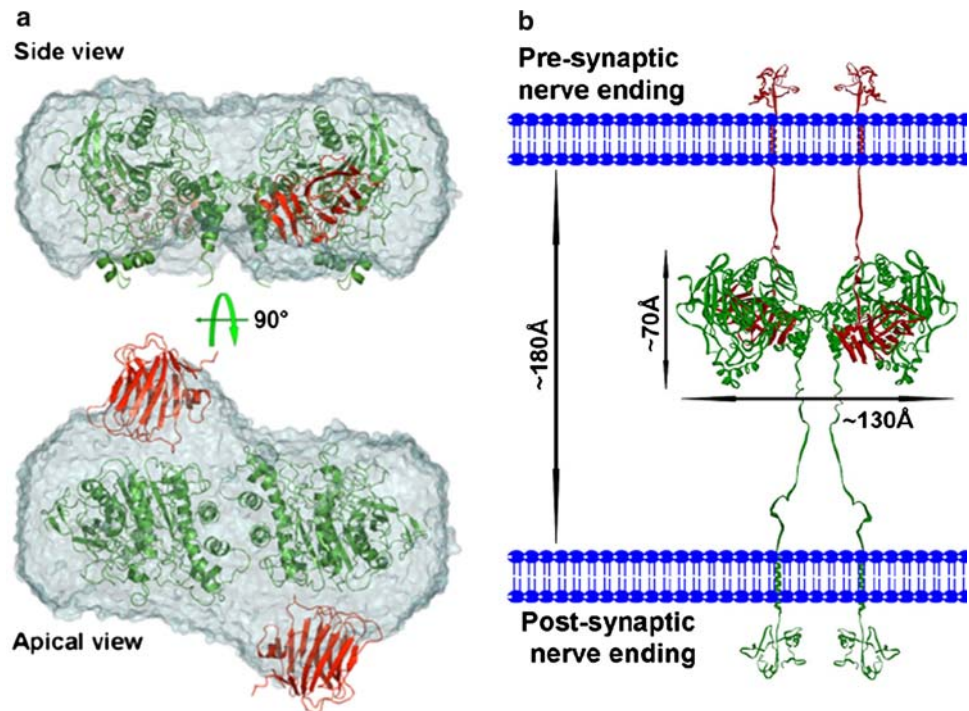
The first approach to SAS data analysis is to fit known high-resolution structures into a model that can describe the experimental data. In its simplest form this is merely confirming that a known crystal structure is consistent with the solution scattering profile. More often the aim is to define the relationships between components of a system

for which the structure of the components is either known or can be modelled based on high resolution structures of related molecules. In its general form rigid body modelling involves preparing a large number of possible model structures and comparing the predicted scattering from the model to experimental data. The models can either be directly refined against experimental data or can be prepared independently and then the best model structure selected. These approaches make it possible to generate high-resolution models from SAS data. However as they do not allow for detailed motion of model components, i.e. the separate parts of the model are rigid, the details of small conformational changes are lost.

The work of Comoletti et al. provides a good example of the overall approach that can be taken here. The study here concerns the structure of the neuroligins and their complex with neurexin (Fig. 1). Neurexin is a transmembrane protein found in the presynaptic membrane with a single transmembrane helix and a long stalk linked to the globular domain that binds neuroligin. Neuroligin has a similar architecture with the globular neurexin binding domain dimerising and is bound to the postsynaptic membrane (Lisé and El-Husseini 2006; Yamagata et al. 2003). These proteins and their complex have an important role in forming and/or maintaining the connections across synapses (Chih et al. 2005; Lisé and El-Husseini 2006; Yamagata et al. 2003). A high-resolution structure of the neurexin under study was available but no structure was available of a neuroligin. However, the neuroligins have significant homology to acetylcholine esterase making it possible to build a homology model (Hoffman et al. 2004). The use of the homology model was validated by showing the scattering from neuroligin constructs was similar to that previously obtained from acetylcholinesterase.

The first application of rigid body modelling was to refine the model structure of the neuroligin dimer itself using the CNS program (Brunger et al. 1998) supplemented with an X-ray scattering data fitting module (Grishaev et al. 2005). An initial model structure was prepared and refined by simulated annealing against experimental SAXS data to generate a model in which there was a 20° rotation of one monomer with respect to the starting model. Two sets of data were used for the modelling of the neuroligin–neurexin complex. SAXS data showed that two neurexin molecules were binding to the periphery of the complex but could not define the positions and orientation. To tackle this, the neurexin was prepared in deuterated form and SANS data on the complex with hydrogenated neuroligin determined in 42% D₂O where the neuroligin is contrast matched. Analysis of the distance distribution function shows a distinctive two-humped shape making it possible to define the distance between the centres of mass of the two-neurexin molecules. The final model was consistent

Fig. 1 Structure of the neuroligin-neurexin complex. **a** The refined model of the complex shown from two directions. Neuroligin is shown in green and neurexin in red. The grey surface is the model obtained from ab initio structure reconstruction based on SAXS data. **b** A model of the neuroligin-neurexin complex in its biological context. Extended linkers connect the complex of the two globular domains to the pre- and post-synaptic membranes. Figures adapted from Comoletti et al. (2007), reproduced with permission of Cell Press



with biochemical data on the basis of interaction and placed the C-termini of the two proteins facing in opposite directions. As the stalks linking each of the two proteins to the appropriate membrane are the correct length to bridge the synaptic gap this orientation acts as additional conformation of the validity of the structure.

A different approach to rigid body modelling used by Bonner et al. (2007) is that of ‘Constrained Modelling’ that has been developed over many years in the group of Stephen Perkins at University College London [to be reviewed in Perkins and Bonner (2008)]. This has been applied to the resolution of the solution structure of a wide range of immunoglobulins (Perkins et al. 2002). These molecules are generally made up of multiple domains connected with flexible linkers. In the case of the human secretory component (SC) studied here there are five immunoglobulin like domains termed D1–D5 (Mostov et al. 1984). A crystal structure is available of an isolated D1 subunit (Hamburger et al. 2004) and the homology between the subunits allows this to be used as a model for all five subunits. However, the flexibility, size, and glycosylation state of the complete SC makes it very unlikely the complete structure can be crystallised. The size of the complex lies outside the range accessible to conventional NMR techniques making it an attractive target for small angle scattering techniques.

The constrained modelling approach is somewhat different to that used in the other papers considered here as it does not involve refinement of a model to fit the data. Instead a large number of candidate model structures are built and filtered based on their agreement with parameters

that can be simply obtained from experimental data such as radius of gyration or cross section as well as an overall fit of the simulated scattering from the model to the data. The models are converted to spheres and the scattering calculated from these models (or a model that has been hydrated for the X-ray scattering data) (Fig. 2a). The best fitting model is then adopted as the candidate structure. Bonner et al., obtain both X-ray and neutron small angle scattering data but in contrast to the other papers considered here these are not used for identifying the contribution to scattering from different parts of the system. Instead the neutron data is used as a control for radiation damage as well as an independent data set, which can be compared to the X-ray data.

They studied two proteolytic fragments of SC as well as the complete five-domain protein. In each case homology models were built for each subunit using the D1 subunit crystal structure. The subunits were joined together with designed linkers obtained from conformational libraries and the linkers therefore defined the relationship between the domains. The 42 c-terminal residues of SC was predicted to be unstructured and was therefore modelled as another unconstrained structure. Models were first built of the two fragments D1–D3 and D4–D5. The D1–D3 model determination was complicated by the sample containing further cleavage products. However it was possible from the data to obtain models of the D1–D3 fragment consistent with the data. The full SC structure was modelled by linking the best-fit models for each fragment together as allowing free linkages between all domains led to models

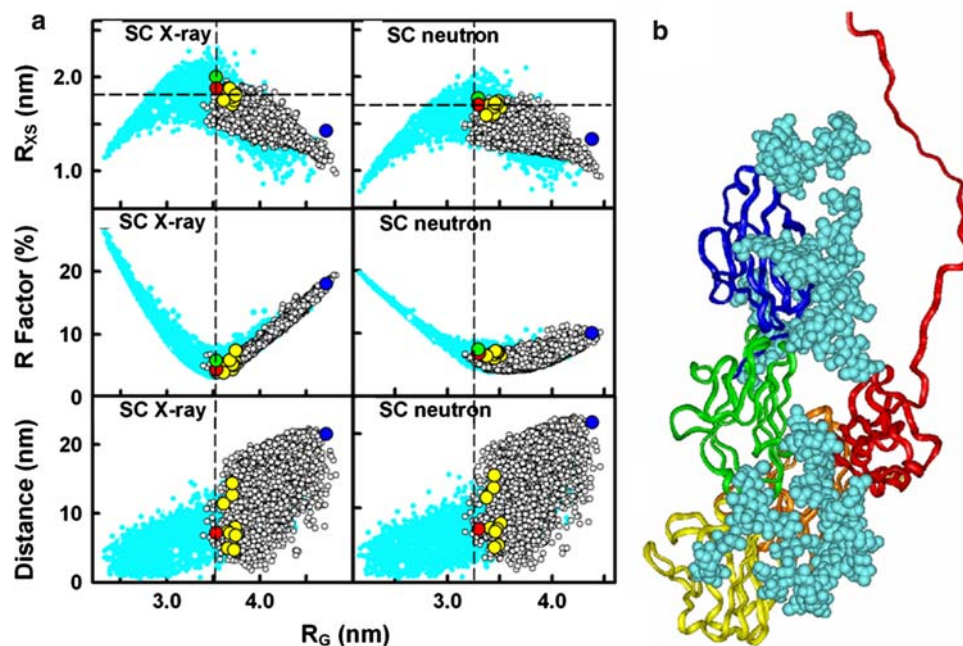


Fig. 2 Constrained modelling of the secretory complex. **a** The 10,000 candidate model structures are plotted with the radius of gyration versus other model derived parameters. *Top panels* show radius of gyration versus cross sectional radius. *Middle panel* shows R_g versus R-factor for the scattering calculated from the model compared to the experimental data. The bottom panel compares the distance from N to C terminus in the model with R_g . The 5,000 SC models represented by blue circles show the unrestrained linker

conformations. The final SC analyses are represented by 5,000 unfilled circles. In each analysis, the best-fit model is shown in red, related best-fit models in yellow, models with compact oligosaccharides in green, and selected poor fit models in blue. **b** Best fit model of the solution structure of the secretory complex. The domains D1 to D5 are coloured from blue (N-terminus) to red (C-terminus) and oligosaccharide chains are shown in blue. Figures adapted from Bonner et al. (2007), reproduced with permission of ASBMB

with too many steric overlaps. SC was shown to adopt a surprisingly compact J-shaped conformation that places the oligosaccharide chains on one side of the protein leaving the CDR regions exposed in a position to make SC's biological interaction with polymeric IgA to form secreted IgA (Fig. 2). The structure has been deposited in the protein data bank as 2OCW.

Ab initio modelling

Obviously rigid body modelling requires the availability of high-resolution structures with sufficient similarity. Where such structures are not available or where their quality is in question it is necessary to take a different approach. A variety of analysis tools have been developed with the aim of reconstructing the overall shape of the target system directly from the scattering data. The most popular of these are the reconstruction tools within the ATSAS suite developed by the group of Dmitri Svergun (Petoukhov and Svergun). DAMMIN and MONSA (Svergun 1999) both attempt to reconstruct the overall shape of systems using dummy atom models and a reverse Monte Carlo (RMC) minimisation approach. The dummy atoms are placed on a lattice, which determines the resolution of the model, and

through the minimisation switched between 'solvent' and 'sample' (or one of a number of different contrast phases in MONSA).

As noted above it is necessary to impose additional constraints to obtain reasonable and stable solutions to the reconstruction problem. DAMMIN and MONSA work by imposing constraints of connectivity (proteins are chains) and compactness (proteins are folded) to reduce the possible range of solutions. Additionally both programmes can include symmetry constraints within the fitting process. While the resolution of the structures obtained from DAMMIN and MONSA is in principle determined by the size of spheres used, this approach is based on the assumption that the internal structure of the molecule being studied is homogeneous. The instructions for DAMMIN recommend removing 'high' resolution data ($Q > 0.3 \text{ \AA}^{-1}$) and manipulates the data to force a Q^{-4} dependency at the highest angles remaining, effectively imposing the assumption of an internally homogeneous structure. In practise the resolution is therefore limited by the length scale on which this assumption of an internally homogeneous structure is plausible, generally around 8–12 Å.

Callow et al. determined the overall subunit organisation and structure of the methyltransferase form of the restriction–modification (R-M) system AdhI (M.AdhI). These are

DNA modification and cleavage enzymes that form part of the bacterial defence system against foreign DNA by labelling specific sequences within the genome with methyl groups and then cleaving (or restricting) unmethylated, and therefore foreign, DNA of the same sequence (Bickle and Kruger 1993). The Type I R-M systems are composed of three individual subunits responsible for specificity (S), DNA methylation (M), and DNA cleavage (R). These can form a methyltransferase enzyme with M_2S stoichiometry or an endonuclease with R_2M_2S stoichiometry. In this case there were no high-resolution structures available of the components of the system. Although there are structures of putative S and M subunits from other organisms (Kim et al. 2005; Calisto et al. 2005; Rajashankar et al. 2005) these were of insufficient homology or quality to be directly useful in model building.

Callow et al. prepared deuterated S subunits and reconstituted these into the complete M_2S methyltransferase. SANS data were collected for fully hydrogenated M_2S in 100% D_2O and for the complex containing deuterated S subunits in both 40% D_2O (M subunits matched) and in 100% D_2O (S subunits matched). This is a classical contrast matching experiment where specific components are matched as nearly as possible and the structure of the remaining components determined. The paper gives an excellent description of how DAMMIN was applied to the re-construction of the structure (Svergun 1999). Briefly the overall structure was reconstructed from the data for hydrogenated M_2S in D_2O . The structure of the individual components was then determined from the appropriate contrast matched data set. The combination of the two individual structures was then fitted onto the overall structure. This gave excellent overall agreement (Fig. 3).

The area of worst agreement is from the extended tips of the M-subunits with the overall structure. As discussed above the DAMMIN reconstruction tool depends on constraints to generate unique solutions and one of the most useful constraints is compactness. It is therefore often challenging to reconstruct extended segments such as that seen in the M-subunit. Allowing for the limitations of the high-resolution structures there is also excellent agreement between these and the ab initio structural model. The combination of the overall structure determined from SANS and the high-resolution models can provide some molecular detail on the likely position of specific important residues but this is limited.

Combining rigid body modelling approaches with contrast variation data analysis

The contrast matching approach used by Callow et al. is the common method for defining the contribution to the overall

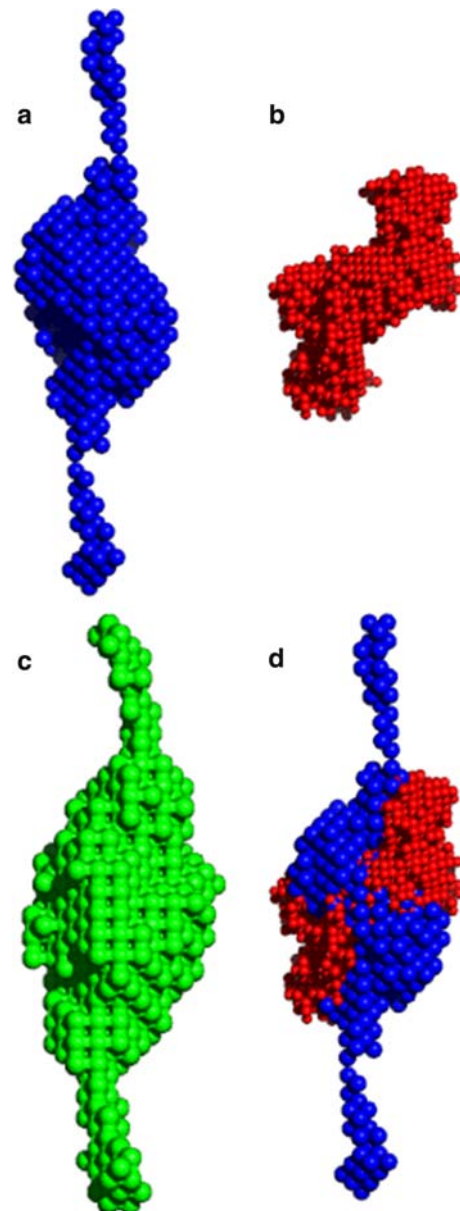


Fig. 3 Ab initio structure reconstruction of the M.AdhI methyltransferase. The panels show the final optimised structures for **a** the M subunit structure obtained from the complex deuterated S contrast matched in 100% D_2O , **b** the S subunit structure obtained from the complex with the hydrogenated M subunit contrast matched in 40% D_2O , **c** the overall structure obtained from the fully hydrogenated complex in D_2O , and **d** the two subunit structures aligned onto the structure of the full complex. Figure adapted from Callow et al. (2007), reproduced with permission of Elsevier Press

structure from different components of a system by SANS. It does however suffer from potential problems with internal inhomogeneity in scattering length density as well as the practical problem of components not being precisely matched. A more sophisticated approach is to use multiple different contrasts to decompose the contribution of two components into the scattering from each component and a

cross term. The scattering from a system with two components with different scattering length densities (e.g. hydrogenated and deuterated) can be written as;

$$I(Q, \Delta\rho_1, \Delta\rho_2) = \Delta\rho_1^2 I_1(Q) + \Delta\rho_1 \Delta\rho_2 I_{1,2}(Q) + \Delta\rho_2^2 I_2(Q)$$

where $\Delta\rho_x$ is the difference between the solvent and subunit SLD (contrast) for each component and $I_x(Q)$ is the contributed intensity from each component and the cross term respectively. As $\Delta\rho_x$ depends on the solvent and can be calculated from the volume and atomic composition of each component the contributions of the subunit scattering and the cross term can be extracted. This can be used very effectively, as shown by Whitten et al. in building and evaluating potential models.

The system under study was the interaction between the histidine kinase KinA that is involved in sporulation of *B. subtilis* and the DNA damage checkpoint inhibitor Sda. Sda binds to KinA preventing autophosphorylation and ultimately preventing the action of a downstream transcription factor (Burkholder et al. 2001). A crystal structure of KinA from *Thermotoga maritima* shows the catalytic domains well separated from the substrate histidines in an extended conformation that would require significant movement for autophosphorylation activity (Marina et al. 2005). Two molecules of Sda, a small protein of ~5 kDa the structure of which has been solved by NMR (Rowland et al. 2004), bind to one KinA dimer.

Scattering of the separate components showed that both KinA and Sda were dimers in solution and confirmed the extended conformation of KinA seen in the crystal structure. On binding of Sda the radius of gyration did not increase from that seen for the KinA dimer and inverse Fourier transform gave a reduction of the overall maximum dimension suggesting significant compaction of the overall structure. Using deuterated Sda and the contrast variation approach described above it was possible to define in some detail plausible structural models of the complex (Fig. 4a).

Rigid body modelling was then used to refine these models using the SASREF7 (Petoukhov and Svergun 2005). This provides a rigid body modelling approach that can take account of X-ray and neutron scattering data with multiple contrast data sets. The catalytic domain was allowed to rotate around two peptide bonds in the linker region connecting it to the stalk and the Sda molecules were lightly constrained to maintain the appropriate binding face of Sda in the correct orientation. The system was constrained by a two-fold symmetry axis along the centre of the stalk and the simulated annealing carried out 14 times. The CA domains and Sda were consistently placed at opposite ends of the stalk. This showed that the mechanism of inhibition by Sda could not be through directly preventing the CA domain reaching the target histidine and

that the mechanism of regulation must be through an allosteric mechanism (Fig. 4b).

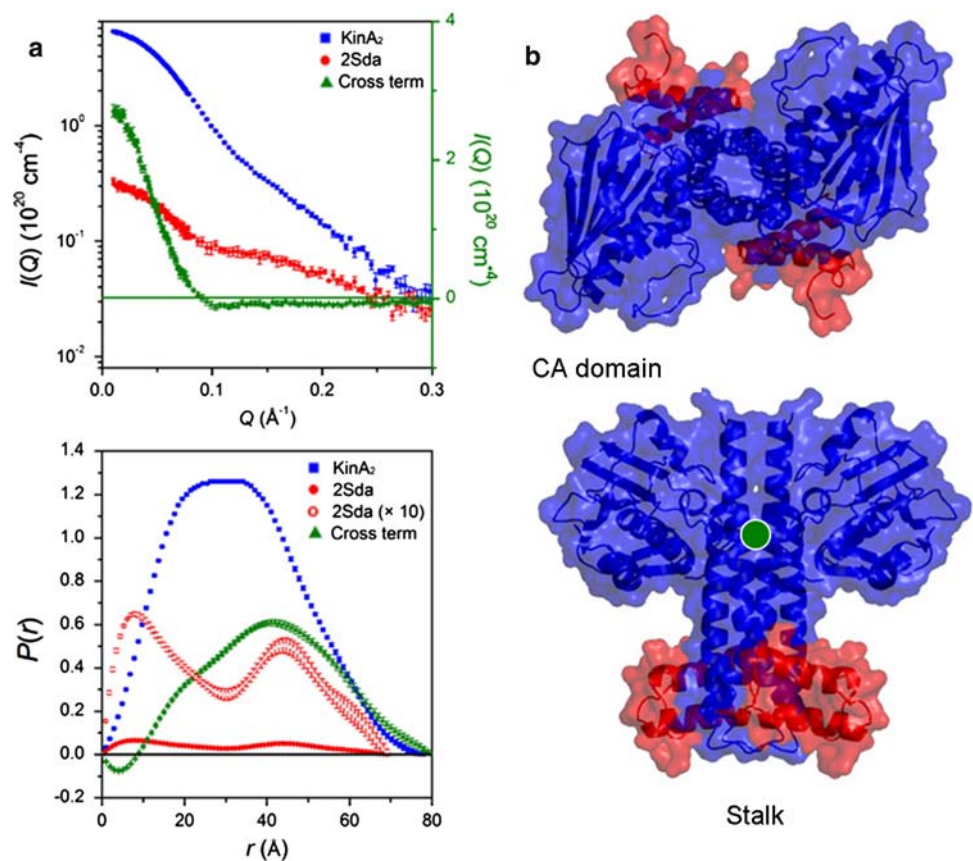
This study was helped a great deal by the availability of useful high-resolution structural data. It is valuable to contrast this with the study of Callow et al., on M.AdhI. A simple contrast match experiment with KinA–Sda would most likely have been unsuccessful due to the small size of the Sda molecule (~5.6 kDa monomer) compared to KinA (54 kDa dimer). The scattering extracted from the two bound Sda molecules by the contrast variation and the scattering from the complex in 40% D₂O (in which KinA is close to contrast match) show relatively weak scattering. In the case of KinA–Sda an ab initio data analysis approach provided structures that were consistent with that obtained from the rigid body modelling. However, the fit of the ab initio models to the data was not as good. The combination of a good high-resolution structure for Sda, multiple contrast data sets, and careful analysis by multiple approaches was critical for developing a good structural model for this system. However, for M.AdhI the available high resolution structures were not of sufficient similarity or quality to use directly in structural modelling. This limits the available approaches to structural determination making ab initio reconstruction the best available technique.

Importance of complementary data

As discussed above the nature of small angle scattering data and the process of reconstructing or modelling three dimensional structure from one dimensional data leads to an inherent ambiguity and the potential for degenerate models. This in combination with the length scales that are probed makes the use of complementary techniques to confirm the validity of models particularly valuable. These may take the form of other structural techniques such as analytical ultracentrifugation or light scattering or biochemical techniques including mutagenesis and binding studies by fluorescence, calorimetry or other methods. Bonner et al. use both sedimentation equilibrium and sedimentation velocity to provide additional information on the structure of secretory complex. These provide details of monodispersity, elongation, and aggregation independently of the small angle scattering data and provide a valuable confirmation of both models and reliability of the scattering data.

Comoletti et al. have a range of biochemical data that provides independent details of the interface between neuroligin and neuroligin (Boucard et al. 2005; Comoletti et al. 2003; Comoletti et al. 2006). This type of data, often based on mutagenesis of one of the binding partners, is frequently available for protein–protein and other

Fig. 4 Structural analysis of the KinA–Sda complex. **a** Deconvolution of the contributions from KinA and Sda via a contrast variation experiment. The extracted scattering profiles and the distance distribution obtained from these profiles is shown. The data extracted can be compared to that from obtained in the experiment where the two components are approximately matched (see Fig. 3 in Whitten et al.). **b** The optimised structure of the complex obtained from rigid body modelling. KinA is shown in blue and Sda in red. The approximate position of the target histidine for phosphorylation is shown as a green dot. Figures adapted from Whitten et al. (2007), reproduced with permission of Elsevier Press



protein–macromolecule interactions. Where it provides a clear indication that specific regions and residues are involved in a binding interface it is extremely valuable. More care needs to be taken with mutations that are associated with functional changes or disease states as the link between the effect of mutation and the binding interaction can be tenuous. Whitten et al., discuss two mutations of KinA that lead to reduced binding of Sda without effecting kinase activity (Whitten et al. 2007). In their final model these residues are unlikely to have a direct interaction with Sda and these mutations provide further support for the biological mechanism of action being mediated through a conformational change on binding. However, the use of this data in model building could have lead to misleading results. The most useful mutational or similar data will be that which can be linked directly through biophysical experiment to binding interactions.

Callow et al., carry out a detailed analysis of their structures within the context of high-resolution structures of related molecules. While these structures are too incomplete and of insufficient homology to use in rigid body modelling they are still useful as a means of testing the model. Comparison and alignment of the reconstructed structure shows can define which regions of the structure

must be different between the two homologs. Alignment of the relevant protein sequences would be expected to show poor alignment for those regions where the structure is expected differ. In the case of the comparison of the M subunit of M.AdhI reconstructed from SANS and the EcoKI M structure obtained by crystallography the central regions of the two structures align well whereas the periphery of the protein which corresponds to the areas which do not readily align within the sequence and do not correspond between the two structures.

Any technique that provides information at the level of single residues on solvent exposure, protein–target interactions, or small-scale distances provides a valuable complement to small angle scattering data. Secondary structural data can also be valuable if it is consistent with the model structures and can aid in the building of more detailed model structures. The inherently low resolution of structural models based on small angle scattering data means that precision data on small length scales is tremendously valuable for building and validating model structures. The integration of these techniques into data analysis for small angle scattering is still at an early stage but future developments can be expected to provide significantly enhanced capabilities in the future.

Is SAS providing biologically relevant information?

The examples discussed here demonstrate the power of small angle scattering using X-rays and neutrons in the structural resolution of large, flexible, and multicomponent systems. It is highly unlikely in any of the cases presented here that the system could be crystallised due to the flexibility and glycosylation of these systems. These are however relatively low-resolution structures providing information on the overall shape the system and the disposition of subunits but not the detailed positioning of specific residues. A criticism that is often levelled at small angle scattering in particular, and other low resolution techniques such as electron microscopy more generally, is that low resolution structures cannot provide the details of structure at the atomic scale that are required to define the chemical basis of biological action. It is reasonable therefore to ask whether the examples here provide useful biological information.

Callow et al., have determined the first structure of an intact Type I R-M system methylase enzyme. The structure cannot provide direct insight into either the mechanism of catalysis or of the specific determinants of recognition of the DNA target. The current structure does not contain the DNA substrate and this is an obvious next step. However, the extended structure does provide an explanation of the large extent of protection of DNA substrates in footprinting experiments and supports a binding mechanism where the extended flexible regions of the protein collapse around the DNA. Such binding mechanisms have been proposed for a number of protein-DNA binding systems but are difficult to prove by crystallography as the non-liganded proteins will not generally crystallise. These can only be tackled at the structural level by determining solution structures of the free and bound proteins.

In the study of the KinA–Sda structure determined by Whitten et al., the structure provides clear evidence that one possible mechanism of inhibition is not viable. It had been proposed that Sda blocked autophosphorylation by physically preventing the catalytic domain swinging into position (Rowland et al. 2004; Varughese et al. 2006). However, Sda is not bound near the hinge region and is therefore not in a position to act as such a ‘molecular barricade’. The mechanism of inhibition must therefore be mediated through a change in conformation of KinA on Sda binding. However Sda appears to be positioned to block the binding of the next target of phosphorylation, the sporulation factor Spo0F (Varughese et al. 2006). This is inconsistent with the biochemical data, which show that Sda blocks autophosphorylation of KinA but not phosphotransfer to Spo0F. This may be due to the fact that the position of Spo0F on KinA is modelled on the basis of a homologous structure which has known structural

differences, or it points to subtleties in the interactions of the components of this system and its regulation that will require further structural and biochemical study. Thus, like any good structure, the KinA–Sda complex both resolves a central issue of mechanism while providing the basis for further biochemical studies to resolve the details of this mechanism. The structure provides a solid framework within which to analyse the data from these experiments.

The folded back structure of the secretory complex determined by Bonner et al. was unexpected based on the extended structure found in two related proteins made up of multiple immunoglobulin domains (Boehm and Perkins 2000; Hu et al. 2005). However the folded structure explains the relative sizes of the linker regions, particularly the D3–D4 linkage of 10 amino acids, as well as the protease susceptibility of this linker. The overall organisation of the domains places the positions for carbohydrate linkage on one side of the protein, leaving the CDR-like regions of the D1 domain and Cys502 in D5. These are known to be required for interaction with the SC binding partner, polymeric IgA. Again the structure raises issues of the details of binding interactions between SC and other molecules, providing a framework within which more detailed biochemical and structural work can be placed.

The study of the neurexin–neuroligin complex structure by Comoletti et al. provides a significant number of important biological insights. In the first case the relative disposition of the two molecules in the neuroligin dimer is established providing a clear view of the dimer interactions. This structure also shows the two stalk regions, that connect the globular domain to the transmembrane domain, leave the globular domain on the same side. The structure of the complex also shows the stalks of the two neurexins leaving their associated globular domains on the same side, and on the opposite side of the complex to that of the neuroligins. This explains how the complex can span the synaptic gap, something that is critical for function but was not clear from the available crystal structures (Fig. 1b). A model based on structural data of the complex is also valuable as it can provide a stronger basis for further investigation. The proposed structure is for instance inconsistent with another model structure proposed on the basis of homology modelling and a neurexin crystal structure (Dean and Dresbach 2006). The structure also provides the basis for beginning to probe the role of neuroligin mutations that have been identified as being associated with autism. These mutations are on the opposite face of neuroligin from the neurexin binding site but cluster in a small region near the symmetry axis. Comoletti et al. suggest that these mutations are therefore less likely to act through effecting the neurexin and neuroligin interaction.

It is clear that the structures obtained from small angle scattering provide different information to that available from high-resolution structural techniques. High-resolution structures provide information on precise atomic positions, chemical mechanism, and detailed molecular recognition. The insights provided by structures obtained from small angle scattering are pointers to allosteric mechanisms, conformational changes, and the relationship between complex shape and the local structural environment. The structure of the neuroligin–neurexin complex shows how this complex spans the synaptic gap. Crystallography will never provide a view of this interaction between molecules, complexes, and the large-scale environment in which they act. Small angle scattering, along with other large scale structural techniques such as analytical ultracentrifugation and electron and now high-resolution optical microscopy have an important role to play in bridging the gap between molecular, cellular, and organismal structures. Small angle neutron scattering in particular has an important role to play in the battery of large-scale structural techniques by providing a unique ability to dissect the structures of macromolecular complexes through contrast variation techniques.

Conclusion

The papers discussed here report high quality structural determination and show the critical importance of maintaining high standards of sample quality, experimental design, and data analysis. The importance of sample characterisation and optimisation in achieving results of the quality described here cannot be overemphasised. The data analysis techniques applied are both challenging to implement and unfortunately easy to misapply. Critical analysis of sample quality, experimental design, and data analysis procedures is therefore highly recommended. Minimum standards for both data reporting and data quality for publication may have a role to play here. Complementary data, whether structural or biochemical, can also play a very valuable role in validating model structures, and should always play a role in modelling or validating structures.

Looking to the future the availability of brighter sources, improved instrumental design, more sophisticated data analysis tools and higher powered computing resources promises significant further advances in the application of SANS and SAXS to structural biology. More dilute samples and faster data acquisition will open the doors to the study of more complex and more challenging biological systems. Many of the most difficult areas for structural biology including membrane proteins and natively unstructured proteins are appealing targets for future

development providing a wide range of potential targets suited to small angle scattering studies. The papers discussed here show a good cross section of what is possible today. The potential is there to be exploited and with the advances expected over the next 5–10 years small angle scattering should see a growing role in structural biology as attention shifts to larger systems more representative of the functional molecular machines that make biology work.

Acknowledgments The author would like to thank Professor Stephen Perkins, Professor Jill Trewhella, and Dr Andrew Whitten for comments on the manuscript.

References

- Ashton AW, Boehm MK, Gallimore JR, Pepys MB, Perkins SJ (1997) Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J Molec Biol* 272:408–422
- Bickle TA, Kruger DH (1993) Biology of DNA restriction. *Microbiol Rev* 57:434–450
- Boehm MK, Perkins SJ (2000) Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. *FEBS Lett* 475:11–16
- Bonner A, Perrier C, Corthesy B, Perkins SJ (2007) Solution structure of human secretory component and implications for biological function. *J Biol Chem* 282:16969–16980
- Boucard AA, Chubykin AA, Comoletti D, Taylor P, Sudhof TC (2005) A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α - and β -neurexins. *Neuron* 48:229–236
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst D Biol Crystallogr* 54:905–921
- Burkholder WF, Kurtser I, Grossman AD (2001) Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. *Cell* 104:269–279
- Calisto BM, Pich OQ, Pinol J, Fita I, Querol E, Carpena X (2005) Crystal structure of a putative type I restriction-modification subunit from *Mycoplasma genitalium*. *J Molec Biol* 351:749–762
- Callow P, Sukhodub A, Taylor JE, Kneale GG (2007) Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering. *J Molec Biol* 369:177–185
- Chih B, Engelman H, Scheiffele P (2005) Control of excitatory and inhibitory synapse formation by neuroligins. *Science* 307:1324–1328
- Comoletti D, Flynn R, Jennings LL, Chubykin A, Matsumura T, Hasegawa H, Sudhof TC, Taylor P (2003) Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1 β . *J Biol Chem* 278:50497–50505
- Comoletti D, Flynn RE, Boucard AA, Demeler B, Schirf V, Shi J, Jennings LL, Newlin HR, Sudhof TC, Taylor P (2006) Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for β -neurexins. *Biochemistry* 45:12816–12827
- Comoletti D, Grishaev A, Whitten AE, Tsigelny I, Taylor P, Trewhella J (2007) Synaptic arrangement of the neuroligin/ β -neurexin complex revealed by X-ray and neutron scattering. *Structure* 15:693–705

- Dean C, Dresbach T (2006) Neuroligins and neuroligins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci* 29:21–29
- Gill SC, Hippel PHv (1989) Calculation of protein extinction coefficients from amino-acid sequence data. *Anal Biochem* 182:319–326
- Grishaev A, Wu J, Trewella 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* 127:16621–16628
- Hamburger AE, West AP, Bjorkman PJ (2004) Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. *Structure* 12:1925–1935
- Hoffman RC, Jennings LL, Tsigelny I, Comoletti D, Flynn RE, Sudhof TC, Taylor P (2004) Structural characterization of recombinant soluble rat neuroligin 1: mapping of secondary structure and glycosylation by mass spectrometry. *Biochemistry* 43:1496–1506
- Hu Y, Sun Z, Eaton JT, Bouloux PMG, Perkins SJ (2005) Extended and flexible domain solution structure of the extracellular matrix protein anosmin-1 by X-ray scattering, analytical ultracentrifugation and constrained modelling. *J Molec Biol* 350:553–570
- Kim J-S, DeGiovanni A, Jancarik J, Adams PD, Yokota H, Kim R, Kim S-H (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc Natl Acad Sci USA* 102:3248–3253
- Lisé MF, El-Husseini A (2006) The neuroligin and neuroligin families: from structure to function at the synapse. *Cellular Molec Life Sci (CMLS)* 63:1833–1849
- Marina A, Waldburger CD, Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *Embo J* 24:4247–4259
- Mostov KE, Friedlander M, Blobel G (1984) The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature* 308:37–43
- Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411–2423
- Perkins SJ (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur J Biochem* 157:169–180
- Perkins SJ, Bonner A (2008) Structure determinations of human and chimaeric antibodies by solution scattering and constrained molecular modelling. *Biochem Soc Trans* (in press)
- Perkins SJ, Gilbert HE, Aslam M, Hannan J, Holers VM, Goodship TH (2002) Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem Soc Trans* 30:996–1006
- Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89:1237–1250
- Petoukhov MV, Svergun DI (2006) Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *Eur Biophys J* 35:567–576
- Petoukhov MV, Svergun DI (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr Opin Struct Biol* 17:562–571
- Rajashankar K, Kniewel R, Lima C, Burley S (2005) Crystal structure of Type I restriction enzyme EcoKI M protein (EC 2.1.1.72) (M.EcoKI). Protein Data Bank
- Rowland SL, Burkholder WF, Cunningham KA, Maciejewski MW, Grossman AD, King GF (2004) Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in *Bacillus subtilis*. *Molec Cell* 13:689–701
- Stuhrmann HB (2004) Unique aspects of neutron scattering for the study of biological systems. *Rep Prog Phys* 67:1073–1115
- Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886
- Trewella J (2006) Neutrons reveal how nature uses structural themes and variation in biological regulation. *Physica B* 385–386:825–830
- Varughese KI, Tsigelny I, Zhao H (2006) The crystal structure of beryll fluoride SpoOF in complex with the phosphotransferase SpoOB represents a phosphotransfer pretransition state. *J Bacteriol* 188:4970–4977
- Whitten AE, Jacques DA, Hammouda B, Hanley T, King GF, Guss JM, Trewella J, Langley DB (2007) The structure of the KinA–Sda complex suggests an allosteric mechanism of histidine kinase inhibition. *J Molec Biol* 368:407–420
- Yamagata M, Sanes JR, Weiner JA (2003) Synaptic adhesion molecules. *Curr Opin Cell Biol* 15:621–632