Journal of
Applied Crystallography
ISSN 0021-8898

Received 29 June 2007
Accepted 31 October 2007

MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies

Andrew E. Whitten,[a,b] Shuzhi Cai[b] and Jill Trewhella[b]*

[a]Bragg Institute, Australian Nuclear Science and Technology Organisation, Lucas Heights, NSW 2234, Australia, and [b]School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia. Correspondence e-mail: jtrewhella@usyd.edu.au

Small-angle neutron scattering with contrast variation can fill important gaps in our understanding of biomolecular assemblies, providing constraints that can aid in the construction of molecular models and in subsequent model refinements. This paper describes the implementation of simple tools for analysing neutron contrast variation data, accessible *via* a user-friendly web-based interface (http://www.mmb.usyd.edu.au/NCVWeb/). There are three modules accessible from the website to analyse neutron contrast variation data from bimolecular complexes. The first module, *Contrast*, computes neutron contrasts of each component of the complex required by the other two modules; the second module, R_g, analyses the contrast dependence of the radii of gyration to yield information relating to the size and disposition of each component in the complex; and the third, *Compost*, decomposes the contrast variation series into composite scattering functions, which contain information regarding the shape of each component of the complex, and their orientation with respect to each other.

1. Introduction

The small-angle scattering of X-rays or neutrons from biological molecules in solution yields low-resolution structural information that can provide useful insights into their functions (Svergun & Koch, 2003; Wall *et al.*, 2000). The complementary nature of small-angle scattering to high-resolution techniques such as crystallography and NMR, and the ever increasing desire to understand more complex biological systems, has brought about a recent surge in interest in the technique. This surge has been greatly facilitated by not only developments in radiation sources and instrumentation allowing fast collection of accurate data, but also the rapid evolution of computing power. The power of modern desktop PCs allows intensive three-dimensional modelling algorithms to be run routinely in relatively short periods of time (Chacon *et al.*, 1998; Petoukhov & Svergun, 2005; Svergun, 1999).

Determination of low-resolution three-dimensional structures from solution scattering data is an appealing prospect. However, model structures may not be uniquely determined by a single scattering profile. This ambiguity can be resolved, in part, through the inclusion of neutron contrast variation data in the modelling process. The neutron contrast variation experiment involves systematic variation (*via* manipulation of the ratio of ^{1}H to ^{2}H) of the neutron scattering length density of the solvent surrounding a scattering particle that possesses two components of distinctly different scattering density (*e.g.* deuterated and non-deuterated protein, or protein and DNA). Variation of the scattering length density of the solvent alters the contribution of each component to the scattering, and the changes in the measured scattering profiles can be related to the structure of each individual component. There is a wealth of information relating to the shapes and dispositions of the components of a complex that can be extracted directly from the contrast variation data using simple analysis techniques. These analyses are also an important step towards model refinement as they can aid in model building and provide an indication of the actual information content of the neutron contrast variation data. The information content will have an impact on the reliability of models of a complex obtained using either *ab initio* or rigid-body modelling techniques.

Programs that calculate atomic model-independent structural parameters have historically been written on an *ad hoc* basis for specific applications. To help make them more generally available to the structural biology community, we have developed a set of modular, web-accessible programs to analyse neutron contrast variation data. An outline of the implementation of the programs and some initial testing is presented in the following sections, along with guidelines for their use in planning and analysing data from neutron contrast variation experiments. There are three modules: the *Contrast* module is used to calculate the contrast of each component of a complex for both experimental planning and analysis of contrast variation data, the R_g module is used to extract radii of gyration, R_g, values of the component structures and provide information on their dispositions, and the *Compost* module is used to extract composite scattering profiles from the contrast variation data.

2. Estimating the contrast of the complex and its components

Key to planning and analysing contrast variation data is reliably estimating the contrast of the scattering particle and its components at each contrast point. Contrast is a fundamental concept in small-angle scattering, which is shown by the expression for the radiation scattered by N non-interacting particles in solution

$$I(q) = N(\Delta\tilde{\rho}V)^2 P(q). \quad (1)$$

In equation (1), the form factor, $P(q)$, describes the intensity variation as a function of momentum transfer, q, and is related to the shape of the particle; V is the volume of the particle; and the contrast, $\Delta\bar{\rho} = \bar{\rho} - \rho_S$, is the difference between the mean scattering length densities of that molecule and the solvent. The neutron contrast is dependent on the isotopic composition of the particle, the solvent, and the degree to which protons and deuterons are exchanged between the particle and the solvent. The *Contrast* module calculates the contrast of a molecule in solution, taking as input: a protein, RNA or DNA sequence, or molecular formula; deuteration level (f_D); volume of the molecule (V); fraction of the exchangeable hydrogen atoms that are accessible by the solvent ($f_{accessH}$); and the deuterium content of the solvent (f_{D_2O}). Given all these parameters, the contrast of the molecule is then evaluated using

$$\Delta\bar{\rho} = \frac{\sum_i^{atoms} b_i + (n_H - n_{XchH}) f_D (b_D - b_H) + n_{XchH} f_{accessH} f_{D_2O} (b_D - b_H)}{V} - \frac{2b_H + b_O + 2f_{D_2O}(b_D - b_H)}{V_{H_2O}}. \quad (2)$$

The number of hydrogen positions (n_H) and the number of exchangeable hydrogen positions (n_{XchH}) can be inferred from the sequence; the deuteration level can be determined from mass spectrometry data; volumes are estimated from the volume of the constituent residues (Nadassy *et al.*, 2001; Tsai *et al.*, 1999; Voss & Gerstein, 2005) by default; however, the user can manually optimize the volumes such that the calculated match point of the complex matches the experimental value determined from a plot of $[I(0)]^{1/2}/c$ *versus* f_{D_2O}; and while $f_{accessH}$ is in practice difficult to determine, it is generally between 0.9 and 1.0. The program also possesses the functionality to correct the scattering-length density of the solvent for the effects of salts and other additives, such as glycerol.

3. Extracting component R_g values and dispositions from the contrast dependence of the total radius of gyration

Stuhrmann and co-workers (Ibel & Stuhrmann, 1975) showed that the radius of gyration of a particle is related in a quadratic manner to the inverse contrast of the particle,

$$R_{obs}^2 = R_m^2 + \frac{\alpha}{\Delta\bar{\rho}} - \frac{\beta}{\Delta\bar{\rho}^2}. \quad (3)$$

The coefficients of the quadratic expression are related to: the radius of gyration of the object with a homogenous distribution of scattering density, R_m; the second moment of the density fluctuations α; and the square of the first moment of the density fluctuations β. For two-component systems, where the difference in scattering length density between the two components is large, these coefficients can be related to the radii of gyration of each subunit, and the separation between the two (Moore, 1982; Olah *et al.*, 1994). These parameters can also be obtained straightforwardly using the parallel-axis theorem,

$$R_{obs}^2 = \frac{\Delta\bar{\rho}_1 V_1}{\Delta\bar{\rho} V} R_1^2 + \frac{\Delta\bar{\rho}_2 V_2}{\Delta\bar{\rho} V} R_2^2 + \frac{\Delta\bar{\rho}_1 V_1}{\Delta\bar{\rho} V} \frac{\Delta\bar{\rho}_2 V_2}{\Delta\bar{\rho} V} D^2. \quad (4)$$

Equation (4) relates the measured radius of gyration, R_{obs}, to the distance between the centres of scattering density of each subunit, D (which will approximate the geometric centres in most circumstances) and the radius of gyration of each of the components, R_1 and R_2. The parallel-axis theorem was first applied to X-ray scattering experiments on an enzyme, its inhibitor, and the complex between the two, allowing the separation distance of the two components in the complex to be determined (Damaschun *et al.*, 1968). It is also possible to extract the radii of gyration of each component and their separations by measuring the radius of gyration of the particle at various contrasts. The contrast variation method was applied firstly to a combination of light, X-ray and neutron scattering (Serdyuk & Fedorov, 1973) data, and subsequently to neutron contrast variation data (Moore *et al.*, 1974).

Both the Stuhrmann analysis and the parallel-axis theorem have been implemented into the module R_g. This module takes as input: the contrast of each subunit at each contrast point (calculated using the *Contrast* module); the radii of gyration and associated estimated standard deviations at each contrast point; and the volume fraction of one subunit (from which the volume fraction of the second can be calculated). Details of the implementation of these expressions can be found in the supplementary material.[1]

4. Extracting composite scattering profiles

A reasonable approximation to the total scattering, $I(q)$, from an object composed of two regions with different contrasts is (Goodisman & Brumberger, 1971; Heller *et al.*, 2002; Kuzmanovic *et al.*, 2006; Olah *et al.*, 1994; Svergun & Koch, 2003; Zhao *et al.*, 1998)

$$I(q) = N\left[\Delta\bar{\rho}_1^2 V_1^2 P_{11}(q) + \Delta\bar{\rho}_2^2 V_2^2 P_{22}(q) + 2\Delta\bar{\rho}_1 V_1 \Delta\bar{\rho}_2 V_2 P_{12}(q)\right] = \Delta\bar{\rho}_1^2 I_{11}(q) + \Delta\bar{\rho}_2^2 I_{22}(q) + \Delta\bar{\rho}_1 \Delta\bar{\rho}_2 I_{12}(q). \quad (5)$$

Conceptually, $I_{11}(q)$ and $I_{22}(q)$ represent the scattering profiles of each component of the object, and $I_{12}(q)$ is the scattering signal due to interference between scattering elements in different components of the object. Hence, $I_{11}(q)$ and $I_{22}(q)$ are related to the shapes of each of the components in the complex and $I_{12}(q)$ is related to their relative dispositions. Using equation (5), the module *Compost* ('composite scattering functions') decomposes a contrast variation series into these three intuitive scattering profiles. While it is possible to generalize equation (5) to any multi-component system, the number of additional variables rises very quickly; hence the program has been written to analyse two-component systems only. Of note, because $P_{11}(0) = P_{22}(0) = 1$, then from equation (5) it is seen that $I_{11}(0) = NV_1^2$ and $I_{22}(0) = NV_2^2$, revealing that the zero-angle scattering intensities of the composite scattering functions are related to the number of particles in solution and the volumes of the components. If the extraction is performed using contrast values with units of 10^{10} cm^{-2} on data that are on an absolute scale, and the concentration of the particles in solution is known, the volume of each component can be calculated using

$$V_X = \left[\frac{I_{XX}(0)\ (\mathrm{cm}^{-1})}{N_A C\ (\mathrm{mol\,L}^{-1})} \times 10^{31}\right]^{1/2} \text{Å}^3. \quad (6)$$

Details regarding the implementation of equation (5) can be found in the supplementary information.

5. Testing

Recently, we completed a combined small-angle X-ray scattering and neutron contrast variation study to determine the low-resolution structure of a histidine kinase dimer (KinA, 50.7 kDa) complexed with a pair of small inhibitor proteins (Sda, 2 × 5.6 kDa) (Whitten *et al.*, 2007). The analysis tools presented here were used in that study to

[1] Supplementary data discussed in this paper are available from the IUCr electronic archives (Reference: ce5024). Services for accessing these data are described at the back of the journal.

Table 1
Comparison of the radii of gyration and separation distances for the KinA–Sda complex obtained from various methods.

Standard uncertainties are given in parentheses.

	R_H (Å)	R_D (Å)	D (Å)†	R_m (Å)
Actual values‡	25.74–26.88§	20.54–21.37	29.37–32.30	27.54
Match point	–	22.38	–	–
Parallel axis‡	25.77 (<1)	22.37 (<1)	28.59 (1)	–
Stuhrmann‡	–	–	–	27.37 (<1)
Composite SF‡	25.70 (<1)	22.29 (1)	–	–

† The distance between the geometric centres is 29.43 Å. ‡ Determined using data at 0, 10, 20, 30, 40, 70, 80, 90, 100% ^2H_2O. § R_g value at 40% ^2H_2O was not included in range, as it is an imaginary number ($R_g^2 = -1289$ Å^2).

build a starting model of the complex in solution, which was subsequently refined against all the measured scattering data. The refined model (supplementary Fig. 1) of the complex from that work is used here as the test case as it represents an extremely challenging example due to the disparity between the sizes of the two components.

All model scattering profiles (for the complex, and each component in the same conformation as that observed in the complex) were calculated using the program *CRYSON* (Svergun *et al.*, 1998), setting the deuteration level of the Sda molecules at 85% (supplementary Fig. 2). Consideration of the solvation layer is complicated when a molecule is in a complex because large parts of its surface interact with its binding partner, and is neglected in all cases to simplify the interpretation of the test results.

5.1. Radii of gyration

The implementation of the Stuhrmann and parallel-axis analyses in the module R_g were tested using radii of gyration determined from model contrast variation data on the complex. The 'actual' radii of gyration shown in Table 1 are taken from the model contrast variation data for each component of the complex alone, in the same conformation as in the complex. The comparison between the actual radii of gyration with those obtained from the parallel-axis theorem show excellent agreement for KinA, but the radius of gyration of the pair of Sda molecules, and the separation distance deviate by ∼1 Å from the actual results, which is much larger than the precision of these values. The actual radii of gyration for KinA and Sda are contrast dependent, partly because ^1H and ^2H exchange between the solvent and the protein, and also because the scattering density in each subunit is heterogeneous (Witz, 1983). The parallel-axis theorem can be applied (without approximation) at any contrast point; however, fitting a single value of R_H, R_D and D, to a contrast variation series will be accurate only when the distribution of scattering density within each subunit is homogeneous. As particles are not homogenous, R_H, R_D and D will always deviate from the real values, and the degree of deviation will depend on the details of size and distribution of internal density fluctuations in a given component. The better agreement for the radii of gyration of the KinA component with real values is due mostly to the disparity in size of the two components.

5.2. Composite scattering functions

Decomposition of the scattering data into composite scattering functions was performed using contrast points at 0, 10, 20, 30, 40, 70, 80, 90 and 100% ^2H_2O (Fig. 1). It would be expected that the obtained scattering functions, $I_{11}(q)$ and $I_{22}(q)$, would, at best, resemble scattering data obtained from contrast-matching experiments. For this example, the scattering length density of Sda is greater than that of pure ^2H_2O; hence it is only possible to solvent-match the KinA. Theoretical scattering data at the match point of the KinA molecule (40.45% ^2H_2O) were generated, and compared with the composite scattering function $I_{22}(q)$. The percentage difference between the two is plotted in Fig. 2, and shows that the differences between them are extremely small, being less than 0.016% of the intensity between q = 0.0 and 0.4 Å^{-1}. This is evidence that the composite scattering functions are an excellent approximation to the data obtained from a solvent-matching experiment.

The probable distribution of interatomic distances, $P(r)$, determined from $I_{11}(q)$, $I_{22}(q)$ and $I_{12}(q)$ using *GNOM* (Svergun, 1992) is shown in Fig. 3, along with those for the isolated KinA and Sda molecules in the same conformation as observed in the structure of the complex.[2] As Fig. 2 shows that the extracted and solvent-matched scattering profiles are the same, the differences between $I_{22}(q)$ and the reference profiles must be due to density fluctuations (Fig. 3*a*) of

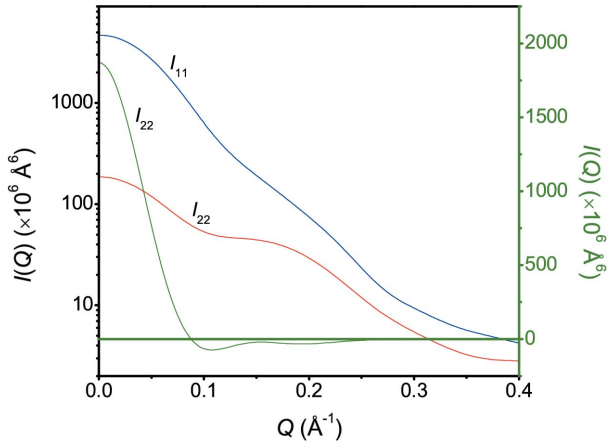

Figure 1
Composite scattering functions for the $KinA_2$–2Sda complex.

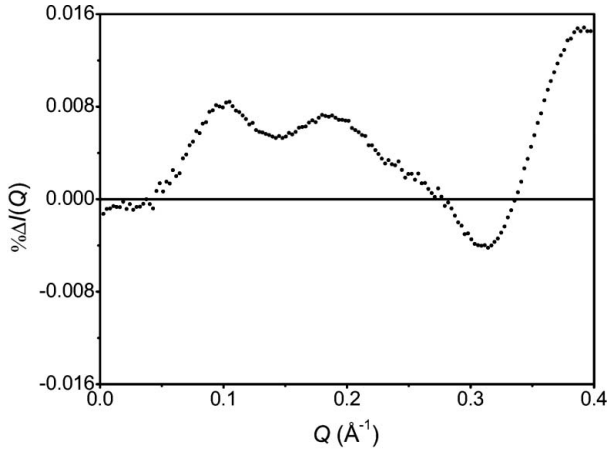

Figure 2
Percentage difference between the scattering profile of the complex at the match point of KinA and the composite scattering profile corresponding to Sda.

[2] The reference $P(r)$ profiles for 2Sda were determined for two Sda molecules with 40.45% ^2H_2O; $KinA_2$ was determined using a single KinA dimer with 100.0% ^2H_2O. The cross-term was determined from $I_{complex}(Q) - I_{KinA}(Q) - I_{Sda}(Q)$ with 0.0% ^2H_2O.

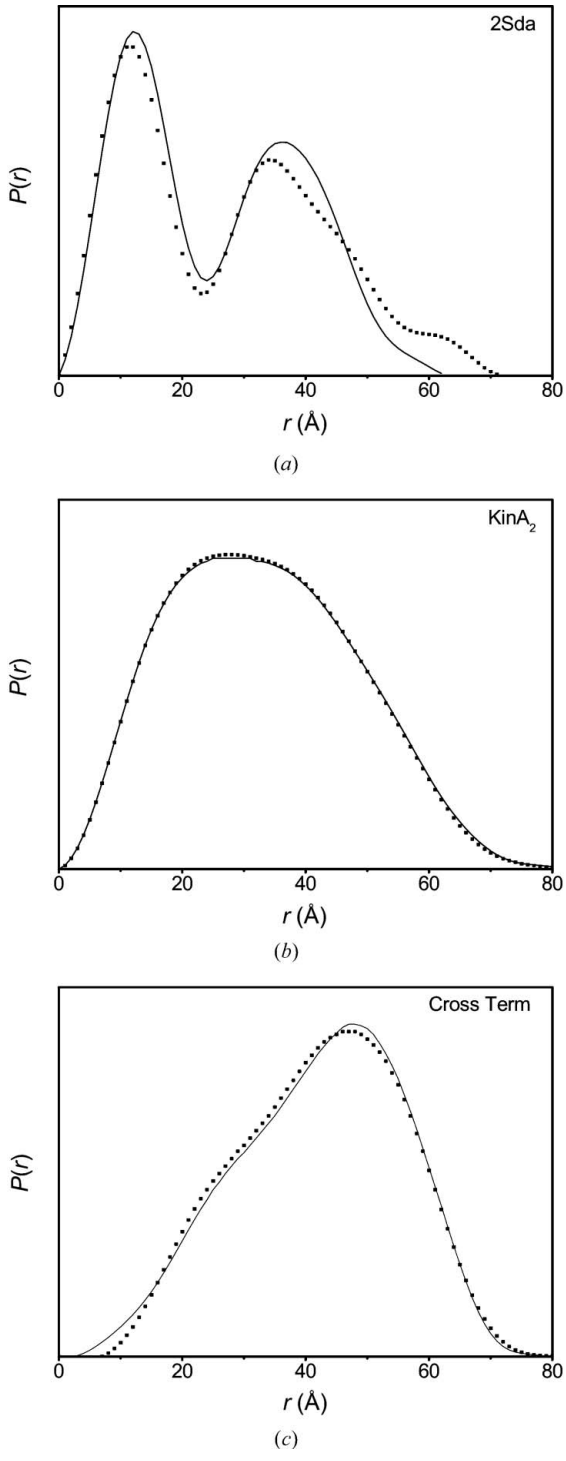

Figure 3
Comparison between actual (solid line) and composite $P(r)$ (dotted line) profiles for: Sda (top); KinA (middle); cross-term (bottom).

the larger KinA dimer at its match point. An analogous plot for the KinA (Fig. 3*b*) shows essentially perfect agreement, indicating that the density fluctuations from the small Sda molecules do not affect the KinA results significantly. The distribution of interatomic distances between KinA and Sda molecules show very good agreement with the reference profile (Fig. 3*c*). This profile encodes information regarding the relationship between the KinA and Sda molecule, and the experimental work on this complex related the shape of this profile to the orientation of the Sda molecules to the catalytic domains of the KinA molecule.

6. Evaluation of the effects of noise in the data and number of contrast points

High-quality data collected at poorly chosen contrast points will not necessarily provide a great deal of information regarding the components of a complex. Hence, there is some interplay between experimental uncertainties and collection strategies. The relationship between experimental uncertainties and collection strategies was explored by applying varying levels of noise to the model profiles and using the various analysis tools on subsets of the contrast variation data (see supplementary data). It was found that the size and shape of the KinA component was retrieved accurately for all situations tested; however, the size and shape of the Sda component was sensitive to data quality and the subset of contrast points used. It was found that accurate retrieval of size and shape information of the Sda component was possible only when the analyses used data of high precision (similar to the experimental work on the high-concentration samples) and at least five well spaced contrast points (0, 20, 40, 80 and 100%).

Based on our testing of the effects of noise on our model data, and also on our experience with our published KinA–Sda experiment (which included a very noisy data set and a much higher quality set), the collection strategy recommended includes a minimum of two contrast points on either side of the average match point of the entire complex. For a protein complex, the deuteration level must therefore be tuned to keep the match point of the entire complex between 50 and 70% ^2H_2O. In general a contrast variation series would include data at: 0 and 100% ^2H_2O, where the contrast of each component is maximized; 20 and 80% ^2H_2O; and one at 40% ^2H_2O, which approximates the match point of an unlabelled protein. In special circumstances, an additional 10 and 90% ^2H_2O measurement may be useful. Such a well spread range of contrast points will give good accuracy and precision for the various analyses. Collection times and concentrations for all samples should be guided by the requirement for the 40% ^2H_2O sample (or the lowest contrast point) to have adequate signal. If the statistical quality of the neutron scattering data is poor, X-ray scattering data may be brought into the analyses. X-ray data are easily incorporated into the extraction of composite scattering functions and parallel-axis analysis, but their relevance to Stuhrmann analysis is not as obvious. As neutron and X-ray scattering experiments measure different things, both should only be included when there are insurmountable limitations to obtaining additional and/or higher quality neutron scattering data. Of course, it is of paramount importance that each neutron data set be free of the effects of sample aggregation and interparticle interference for the accurate interpretation of structural information.

7. Conclusions

The example used here to test the implementation of the various analyses serves well for evaluating the accuracy of the various analyses as it simultaneously represents a best case and worst case scenario. The testing showed that the composite scattering functions and radii of gyration determined from the various analyses agree with what is expected for the KinA dimer component. While the agreement of the composite scattering function for the pair of Sda molecules did not agree as well, the deviations are small and there is

excellent agreement with the scattering profile simulating a solvent-matching experiment, the common alternative approach to composite scattering function extraction. Both the radius of gyration of the pair of Sda molecules determined from the parallel-axis theorem and the separation distance between the two components shows a small variance (∼1 Å) with the expected values. These deviations are due to a combination of the approximations inherent in the analysis that are accentuated by the small size of the Sda component. The extracted cross-term is found to be realistic.

For our test model, the contrast is known and so the effects of incorrectly estimating contrast on the various analyses have not been quantified here. The experimentally determined match point of the entire molecule can be used as a guide to whether the estimated contrasts of the components are sensible, and careful sample characterization of (including mass spectrometry) will aid in ensuring that the estimates are reasonable. We advocate evaluating the sensitivity of any derived parameters and models for each experimental case.

The analyses described here are useful for evaluating the information content in a neutron contrast variation data set and in providing constraints for the construction of initial models. They can also be used to gauge the extent of conformational rearrangement a component of the structure undergoes upon complexation, or to indicate whether a given homology structure appropriately represents part of the structure. While it is possible to refine models of each component against the composite scattering functions, due to the approximations in the extraction algorithm models should be refined directly against the measured scattering data. There are excellent programs available for carrying out such refinements (Petoukhov & Svergun, 2006). These programs also can properly account for the X-ray and neutron scattering data simultaneously. Nonetheless, we believe that the analyses presented here are extremely useful in helping to understand the nature of the complexes being studied and providing useful structural constraints.

The authors wish to acknowledge Doug Chappell for his help with administration of the website, and David Langley and David Jacques for testing the programs and providing helpful suggestions regarding their functionality. JT wishes to acknowledge that the work was carried out under Federation Fellowship FF0457488.

References

Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**, 2760–2775.
Damaschun, G., Fichtner, P., Purschel, H. V. & Reich, J. G. (1968). *Acta Biol. Med. Germanica*, **21**, 309–316.
Goodisman, J. & Brumberger, H. (1971). *J. Appl. Cryst.* **4**, 347–351.
Heller, W. T., Abusamhadneh, E., Finley, N., Rosevear, P. R. & Trewhella, J. (2002). *Biochemistry*, **41**, 15654–15663.
Ibel, K. & Stuhrmann, H. B. (1975). *J. Mol. Biol.* **93**, 255–265.
Kuzmanovic, D. A., Elashvili, I., Wick, C., O'Connell, C. & Krueger, S. (2006). *J. Mol. Biol.* **355**, 1095–1111.
Moore, P. B. (1982). *Methods Exp. Phys.* **20**, 337–390.
Moore, P. B., Engelman, D. M. & Schoenborn, B. P. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 172–176.
Nadassy, K., Tomas-Oliveira, I., Alberts, I., Janin, J. & Wodak, S. J. (2001). *Nucleic Acids Res.* **29**, 3362–3376.
Olah, G. A., Rokop, S. E., Wang, C. L. A., Blechner, S. L. & Trewhella, J. (1994). *Biochemistry*, **33**, 8233–8239.
Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.
Petoukhov, M. V. & Svergun, D. I. (2006). *Eur. Biophys. J. Biophys. Lett.* **35**, 567–576.
Serdyuk, I. N. & Fedorov, B. A. (1973). *J. Polym. Sci. C Polym. Lett.* **11**, 645–649.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D. I. & Koch, M. H. J. (2003). *Rep. Progr. Phys.* **66**, 1735–1782.
Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.
Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). *J. Mol. Biol.* **290**, 253–266.
Voss, N. R. & Gerstein, M. (2005). *J. Mol. Biol.* **346**, 477–492.
Wall, M. E., Gallagher, S. C. & Trewhella, J. (2000). *Annu. Rev. Phys. Chem.* **51**, 355–380.
Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trewhella, J. & Langley, D. B. (2007). *J. Mol. Biol.* **368**, 407–420.
Witz, J. (1983). *Acta Cryst.* A**39**, 706–711.
Zhao, J. K., Hoye, E., Boylan, S., Walsh, D. A. & Trewhella, J. (1998). *J. Biol. Chem.* **273**, 30448–30459.