Eur Biophys J (2010) 39:769–780
DOI 10.1007/s00249-009-0549-3

ORIGINAL PAPER

Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering

Pau Bernadó

Received: 11 June 2009 / Revised: 10 September 2009 / Accepted: 24 September 2009 / Published online: 21 October 2009

Abstract Multidomain proteins in which consecutive globular regions are connected by linkers are prevalent in nature (Levitt in Proc Natl Acad Sci USA 106:11079–11084, 2009). Some members of this family have largely resisted structural characterization as a result of challenges associated with their inherent flexibility. Small-angle scattering (SAS) is often the method of choice for their structural study. An extensive set of simulated data for both flexible and rigid multidomain systems was analyzed and modeled using standard protocols. This study clearly shows that SAXS profiles obtained from highly flexible proteins can be wrongly interpreted as arising from a rigid structure. In this context, it would be important to identify features from the SAXS data or from the derived structural models that indicate interdomain motions to differentiate between these two scenarios. Features of SAXS data that identify flexible proteins are: (1) general attenuation of fine structure in the scattering profiles, which becomes more dramatic in Kratky representations, and (2) a reduced number of interdomain correlation peaks in $p(r)$ functions that also present large D_{max} values and a smooth decrease to 0. When modeling this dynamically averaged SAXS data, the structures obtained present characteristic trends: (1) ab initio models display a decrease in resolution, and (2) rigid-body models present highly extended conformations with a lack of interdomain contacts. The ensemble optimization method represents an excellent strategy to identify interdomain motions unambiguously. This study provides information that should help researchers to select the best modeling strategy for the structural interpretation of SAS experiments of multidomain proteins.

Keywords Small-angle scattering (SAS) · Multidomain proteins · Protein structure · Protein dynamics

Introduction

Small-angle scattering of X-rays (SAXS) and neutrons (SANS) are established techniques in the study of biological macromolecules in solution (Svergun and Koch 2002, 2003; Petoukhov and Svergun 2007; Putnam et al. 2007). Traditional approaches for low-resolution structure determination based on small-angle scattering (SAS) experiments rely on the optimization of a global shape or a conformation able to reproduce the experimental profile faithfully. Therefore, these methods use a single-molecule representation for the description of the data. Approaches in which SAS profiles are fitted as a weighted average of a definite number of different species in solution have been used to describe conformational changes and oligomerization processes (Konarev et al. 2003). Recently, a novel approach, the so-called ensemble optimization method (EOM), has been presented to describe SAS profiles in terms of an ensemble of available conformations in solution (Bernadó et al. 2007). This strategy has proven appropriate for the structural description of highly flexible systems.

Modular or multidomain proteins, comprising two or more folded domains tethered by linkers, are common in nature (Lim 2002; Ekman et al. 2005; Levitt 2009). Due to

Electronic supplementary material The online version of this article (doi:10.1007/s00249-009-0549-3) contains supplementary material, which is available to authorized users.

P. Bernadó (✉)
Laboratory of Biomolecular NMR, Institute for Research in Biomedicine, Parc Científic de Barcelona, Baldiri Reixac, 10-12, 08028 Barcelona, Spain
e-mail: pau.bernado@irbbarcelona.org

their special architecture, multidomain proteins can often adopt several conformations in solution. In some cases the rigid (globular) domains move among distinct arrangements in a concerted way, facilitated by the flexibility of the linkers (Hammel et al. 2004, 2005; Fetler et al. 2007; Bernadó et al. 2008; Bron et al. 2008; Taraban et al. 2008). A typical example of this situation is the collapse from an extended conformation of the calcium binding protein calmodulin (Trewhella et al. 1990). In other situations, the inherent flexibility of the linker induces the protein to sample an astronomical number of conformations (Braddock et al. 2002; Jacobs et al. 2003; Mulder et al. 2004). Several biological advantages of this arrangement can be envisioned with respect to the situation where the domains are presented separately, such as adaptability to different targets or simultaneous recognition of multiple binding sites (Dunker et al. 2002).

For highly flexible proteins, traditional structure determination by X-ray crystallography is more complicated and often impossible because of difficulties in the crystallization process. Nuclear magnetic resonance (NMR) is a more suitable technique to address structural studies of flexible multidomain proteins (Pickford and Campbell 2004). Although the relative orientation of the domains can be easily obtained by NMR, no methods are available to interpret structural restraints in terms of large ensembles of conformations with very different overall shapes.

SAS techniques provide an easy entry for the structural characterization of multidomain proteins, and there are a multitude of examples in which structural models of these proteins have been obtained using SAXS, SANS or a combination of both (Receveur et al. 2002; Zou et al. 2003; Márquez et al. 2003; Nagar et al. 2006; Petoukhov et al. 2006; Grela et al. 2008; von Castelmur et al. 2008). In addition, the molecular size and shape properties derived from SAS experiments can be combined with NMR short-range restraints in order to validate (Improta et al. 1998; Marino et al. 2006) or derive more accurate structural models (Grishaev et al. 2005, 2008; Mareuil et al. 2007; Gabel et al. 2008). However, SAS is not exempt from problems when studying flexible proteins. In this case, the sample must be considered as polydisperse, and the scattering profile measured is a weighted average over all accessible structures (Svergun and Koch 2003). Therefore, the above-mentioned single-conformation approaches are no longer valid and ensemble approaches must be used. The influence of the coexistence of a small and well-defined number of conformations on SAXS data and on the models derived has been studied and difficulties in distinguishing between static and dynamic scenarios have been reported (Heller 2005).

Here we use synthetic data calculated for distinct, highly flexible, modular proteins to address two questions: (1) testing the effects of this flexibility on the structural models derived when single-conformation and ensemble approaches are used to describe the data, and (2) characterization of the effect of dynamics on the scattering data and derived models in order to identify features useful for the detection of large-scale motions. The results of our simulations highlight that SAS profiles and $p(r)$ functions are smoothed by dynamics, and that structures can be severely aliased when dynamics is not specifically accounted for in the modeling process. These aspects can be used for the diagnosis of interdomain dynamics in modular proteins and appropriate selection of the best SAS modeling strategies.

Methods

Synthetic SAXS profile calculation

Synthetic SAXS profiles were computed for four polyubiquitin chains encompassing two, three, four, and five ubiquitin domains connected by 20-residue-long linkers. Ensembles of 5,000 random conformations were calculated for each case using the program Pre_Bunch (Bernadó et al. 2007), which treats the folded domains as rigid entities connected by self-avoiding linkers that observe a conformational sampling defined by an empirical C_α–C_α quasi-Ramachandran potential (Kleywegt 1997). The 1–72 fragment of the ubiquitin X-ray structure, 1ubq (Vijay-Kumar et al. 1987), was used for the calculation. The theoretical scattering profile for each conformation of the ensemble was computed with the program CRYSOL (Svergun et al. 1995) using default parameters and averaged to produce the synthetic curves for the four test cases studied.

Synthetic noise was introduced into the curves by taking experimental data of a 4.66 mg ml^{-1} sample of a 52-kDa protein as a reference (Bernadó et al. 2008). The synthetic curves were scaled with respect to the experimental one, and the noise was added to each point as a Gaussian distribution centered on the theoretical value and with a standard deviation corresponding to the experimental error measured for that point. A momentum transfer range covering $0.008\ \text{Å}^{-1} < s < 0.47\ \text{Å}^{-1}$ was simulated in all cases. A second data set was computed in the same way but with a threefold increase in the level of noise (referred to herein as "3X").

A random structure from each of the ensembles of the different polyubiquitin chains was selected with the only restriction of having an R_g similar to the $\langle R_g \rangle$ of the ensembles. These four rigid conformations were analyzed in the same manner as the dynamically averaged ones. In another test, the *asymmetric* chain was identical to that previously described but with four distinct globular

domains: SH3 (1awx, 55 aa), ubiquitin (1ubq, 72 aa), MAD2A (1s2h, 205 aa), and MBP (1jvx, 371 aa).

Size and shape parameters, R_g and maximum dimension (D_{max}), defined by the dynamically averaged curves were derived with the program PRIMUS (Konarev et al. 2003) and GNOM (Svergun 1992), respectively. The distance distribution function, $p(r)$, for the complete momentum transfer range was calculated with the program GNOM (Svergun 1992).

The individual radius of gyration for each conformer obtained with CRYSOL were averaged to obtain the ensemble-averaged R_g, $\langle R_g \rangle$, in the following way:

$$\left\langle R_g^2 \right\rangle = {}^1\!/_N \sum_{i=1,N} R_{g,i}^2, \quad (1)$$

where N is the number of conformations simulated.

Single-conformation structural modeling of SAXS data

The resulting synthetic curves were submitted to structural modeling using DAMMIN (Svergun 1999), an ab initio low-resolution shape reconstruction software, and BUNCH (Petoukhov and Svergun 2005), a program that performs a rigid-body optimization of the globular entities encompassing the modular protein. Default parameters were used for both types of calculation. Two independent sets of ab initio reconstructions were performed using $p(r)$ functions derived from the complete momentum transfer ($s < 0.47$ Å^{-1}), and a shortened one ($s < 0.30$ Å^{-1}). The complete momentum transfer range was used for rigid-body calculations. For each case, ten independent runs were performed using both programs.

The structural divergence between the resulting models was monitored with the normalized spatial discrepancy (NSD) obtained with the program SUPCOMB (Kozin and Svergun 2001). The NSD between two structure sets S_1 and S_2 containing three-dimensional (3D) coordinates of points (e.g., atoms, residues or beads) is defined as follows. For every point s_{1i} from the set $S_1 = \{s_{1i},\ i = 1,\ldots,N_1\}$, the minimum value among the distances between s_{1i} and all points in the set $S_2 = \{s_{2i}, i = 1,\ldots,N_2\}$ is denoted as $\rho(s_{1i}, S_2)$. The NSD is a normalized average

$$\mathrm{NSD}(S_1, S_2) = \left[\frac{1}{2}\left(\frac{1}{N_1 d_2^2}\sum_{i=1}^{N_1}\rho^2(s_{1i}, S_2) + \frac{1}{N_2 d_1^2}\sum_{i=1}^{N_2}\rho^2(s_{2i}, S_1)\right)\right]^{1/2} \quad (2)$$

where N_i is the number of points in S_i and the fineness d_i equals the average distance between the neighboring points in S_i. For ideally superimposed objects NSD tends to 0; when it significantly exceeds 1, the objects systematically differ from one another. When comparing the structures from an ensemble, the NSD between all pairs of conformers were calculated with the program SUPCOMB (Kozin and Svergun 2001) and the average value, $\langle \mathrm{NSD} \rangle$, is reported.

The ensemble optimization method (EOM)

The previously described pools composed of 5,000 structures for each polyubiquitin chain were used as potentially selectable conformations/curves in the EOM protocol. In the EOM approach the scattering profile, $I_{EOM}(s)$, is represented by the average of N individual curves

$$I_{EOM}(s) = \frac{1}{N}\sum_{n=1}^{N} I_n(s), \quad (3)$$

where $I_n(s)$ is the scattering from the n-th conformer. The N curves (chromosome) are selected from the pool in a way that, when averaged, minimizes its discrepancy χ^2 from the experimental scattering profile, $I_{exp}(s)$

$$\chi^2 = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\mu I_{EOM}(s_j) - I_{exp}(s_j)}{\sigma(s_j)}\right]^2, \quad (4)$$

where K is the number of experimental points, $\sigma(s)$ are standard deviations, and μ is a scaling factor. This optimization is performed using a genetic algorithm (Bernadó et al. 2007). In this case ensembles of $N = 50$ structures were used, and they were allowed to evolve for 500 generations. The whole process was repeated 200 times starting from a different set of chromosomes to produce a selected ensemble of 10,000 (50×200) conformations. The R_g distribution of the selected conformations was monitored to assess the differences between the dynamic and the static scenarios.

Results

Overall parameters derived from the averaged profiles

The scattering curves derived for all simulated polyubiquitin modular systems are displayed in Fig. 1a. The R_g values derived from the SAXS profiles and from the ensembles of conformations are shown in Table 1. As expected, a systematic increase in R_g was observed when the size of the simulated system increased. The profile-derived R_g values lay within the broad range of the ensemble, but were smaller than the averaged ones computed from the different conformations of the ensemble. All $p(r)$ functions computed with the program GNOM presented a common peak around 18 Å that corresponds to the intrasubunit distances (Fig. 1b). Interestingly, in all cases $p(r)$ functions presented a second peak that corresponds to the distance correlation between adjacent folded

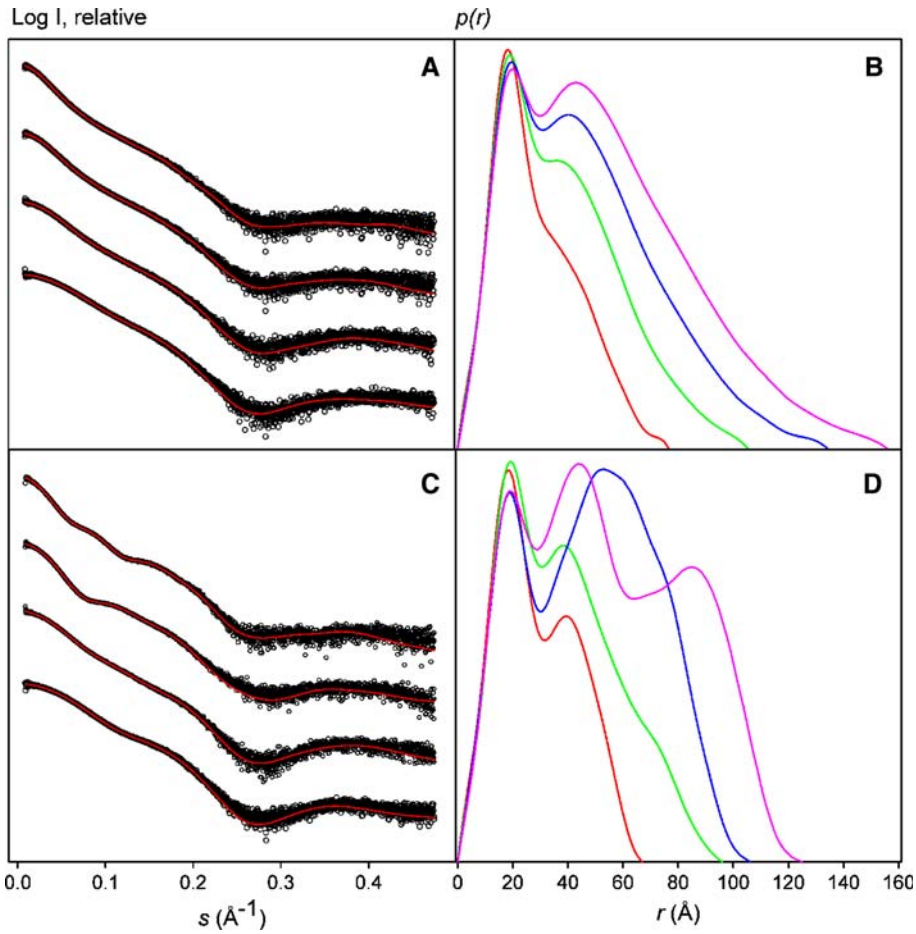

Fig. 1 Scattering profiles and distance distribution functions for dynamic (**a**, **b**) and static (**c**, **d**) scenarios of the four systems simulated. **a** and **c** Scattering intensity (in logarithmic scale) as a function of the momentum transfer ($s = 4\pi\sin(\theta)/\lambda$) for the averaged curves derived from the **a** ensembles and **c** the static structures of the di-, tri-, tetra-, and penta-ubiquitin (from *bottom* to *top*). The curves are appropriately displaced along the axis for better visualization. *Solid red lines* represent profiles obtained for the best BUNCH model for each curve. **b** and **d** $p(r)$ functions calculated with GNOM using the complete scattering profile for the di- (*red*), tri- (*green*), tetra- (*blue*), and penta-ubiquitin (*pink*) simulated systems in the **b** dynamic and the **d** static scenarios

Table 1 Overall parameters calculated from the conformational ensembles and from the derived SAXS curves

	R_g SAXS (Å)[a]	$\langle R_g \rangle$ Ensem (Å)[b]	R_g Range (Å)[c]	D_{max} SAXS (Å)[d]	D_{max} Ensem (Å)[e]
Di-Ubi	22.2 ± 0.04	23.47	16.03–37.99	77 ± 2	97.5
Tri-Ubi	29.5 ± 0.07	31.53	19.42–49.02	106 ± 2	141.1
Tetra-Ubi	35.6 ± 0.09	38.28	22.25–59.48	135 ± 2	185.1
Penta-Ubi	41.2 ± 0.13	44.56	24.43–74.26	157 ± 3	222.7
Asymm.	43.4 ± 0.10	46.31	29.93–67.63	168 ± 3	226.8

[a] Derived using Guinier's approach with the program PRIMUS (Konarev et al. 2003)

[b] Average R_g computed from the ensemble, see the "Methods" section

[c] Minimum and maximum R_g obtained from the ensemble of 5,000 conformations

[d] Maximum distance derived from the averaged curve with the program GNOM (Svergun 1992)

[e] Maximum distance among the 5,000 conformers of the ensemble

domains. This peak was slightly shifted towards larger distances when the size of the simulated protein increased. Additional correlation peaks for tri-, tetra-, and penta-ubiquitin systems were expected (Svergun and Koch 2003), although they appeared to be averaged out by dynamics. Dynamic interdomain motions induced $p(r)$ functions to decrease smoothly to $p(D_{max}) = 0$. The maximum distance values, D_{max}, obtained in the GNOM analysis are shown in Table 1 along with the largest D_{max} computed for a single conformation in the ensemble. As expected, D_{max} increased with the length of the chain; however, it remained notably below the maximum distance present in the simulated ensemble. Importantly, ~15% of the conformations computed had D_{max} values larger than the one derived from the SAXS data. This result is in agreement with previous observations suggesting that the D_{max} derived from the

indirect transformation from SAS data does not reproduce the longest dimension in an ensemble of conformations (Heller 2005).

Impact of interdomain dynamics on SAXS data

To monitor the impact of interdomain dynamics on SAXS data, synthetic scattering profiles were obtained from static conformations. A random structure from each of the ensembles of the different polyubiquitin chains was selected with the only restriction of having an R_g similar to the $\langle R_g \rangle$ of the ensembles (shown in Table 1). The addition of random noise and the $p(r)$ computation were performed equivalently as previously described, and compared with the dynamically averaged data.

The SAXS curves computed for the four randomly selected static polyubiquitin chains were less smooth and presented more fine structure in the whole momentum transfer range than their dynamic counterparts (Fig. 1c). The effect of interdomain dynamics in the $p(r)$ functions was also evaluated (Fig. 1d). The distance distributions of static conformations presented much more fine structure than those calculated from the dynamically averaged profiles. Some of the $p(r)$ maxima relating to the interdomain distances, which were missing in the dynamic averaged functions, were now identified. In addition, a more abrupt ending of the $p(r)$ function was observed in the static regime.

These differences were enhanced in the Kratky representation of the two datasets (Fig. 2). Kratky plots are a very useful representation of the scattering intensity. These plots qualitatively assess the compactness of the molecule and allow differentiation between folded and unfolded proteins (Doniach 2001). For unstructured chains, Kratky plots present a continuous increase in the $s^2I(s)$ value, whereas a pronounced peak is observed for globular proteins. Despite the highly flexible nature of the systems simulated, Kratky plots presented the appearance expected for globular proteins (Fig. 2a). Conversely, Kratky plots of rigid conformations of polyubiquitin chains (Fig. 2b) displayed more complex features with several small peaks and shoulders. According to these observations, the presence of a single peak in the Kratky plot in dynamically averaged data seems to indicate disorder and a spatial decorrelation between the different globular domains.

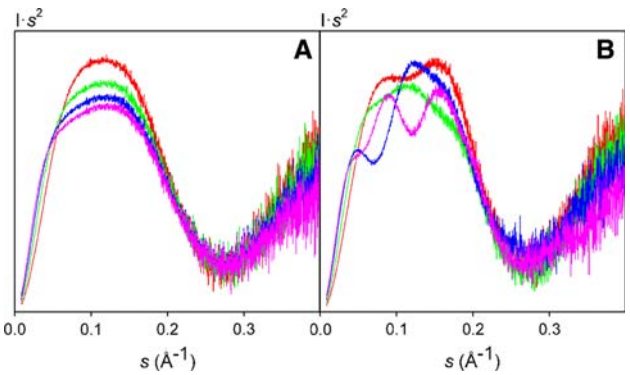

Fig. 2 Kratky plots for the four polyubiquitin systems simulated in a highly flexible (**a**) and rigid (**b**) scenarios. The di- (*red*), tri- (*green*), tetra- (*blue*), and penta-ubiquitin (*pink*) systems have the same color code in both panels

These observations show that interdomain dynamics has dramatic effects on the SAS data. As a general effect, motion smoothes scattering profiles, Kratky plots, and $p(r)$ functions.

Structural modeling of the averaged SAXS profiles

To assess the capacity of structure modeling programs to describe dynamically averaged data, these profiles obtained from the four ensembles were submitted to standard structural modeling analysis. The artifacts induced by the single-particle analysis of SAXS curves of modular systems that sample a large number of conformations were also evaluated.

Ab initio shape reconstructions with DAMMIN

Ten independent ab initio reconstructions were performed with the program DAMMIN (Svergun 1999) for each of the four systems simulated using the $p(r)$ functions shown in Fig. 1b. The structure of the independent solutions was very similar, as shown by the average NSD values (Table 2). Independently of the size of the chain, $\langle \text{NSD} \rangle$ remained much below 1.0. Therefore, the dynamically averaged scattering profiles yielded a well-defined solution despite the highly flexible nature of the proteins simulated.

The average shape reconstructions obtained with the DAMAVER (Volkov and Svergun 2003) suite of programs are shown in Fig. 3. The resulting envelopes presented an elongated shape in order to accommodate the relatively large values of R_g and D_{max} shown by the dynamically averaged curves. The individual domains could not be distinguished as the density was homogeneously thick along the envelope. Therefore, the exact positions of the linker could not be assessed, thereby indicating that the interdomain dynamics induces extra density in the tethering regions.

Rigid-body modeling with BUNCH

Ten independent rigid-body structure determinations were performed from the four dynamically averaged SAXS curves with the program BUNCH (Petoukhov and Svergun

Table 2 Summary of the modeling parameters

	DAMMIN			BUNCH		
	χ^2 Range[a]	⟨NSD⟩[b]	⟨NSD⟩ 3X[c]	χ^2 Range[a]	⟨NSD⟩[b]	⟨NSD⟩ 3X[c]
Di-Ubi	0.99–1.02	0.58 ± 0.01	0.54 ± 0.03	1.32–1.44	0.78 ± 0.01	0.82 ± 0.02
Tri-Ubi	1.15–1.21	0.57 ± 0.03	0.56 ± 0.03	1.35–1.51	0.89 ± 0.04	0.93 ± 0.02
Tetra-Ubi	1.11–1.38	0.64 ± 0.03	0.65 ± 0.01	1.24–1.42	1.17 ± 0.07	1.15 ± 0.06
Penta-Ubi	1.15–1.48	0.72 ± 0.02	0.76 ± 0.03	1.22–1.38	1.31 ± 0.04	1.46 ± 0.08
Asymm.	1.64–1.74	0.67 ± 0.03	N.A.	1.12–1.41	1.58 ± 0.09	N.A.

[a] Range of figure of merit, χ^2, from the ten independent calculations using synthetic curve with standard level of noise

[b] Average NSD among the ten independent structure calculations

[c] Average NSD among the ten independent structure calculations using the synthetic curve with a threefold increase of the level of noise

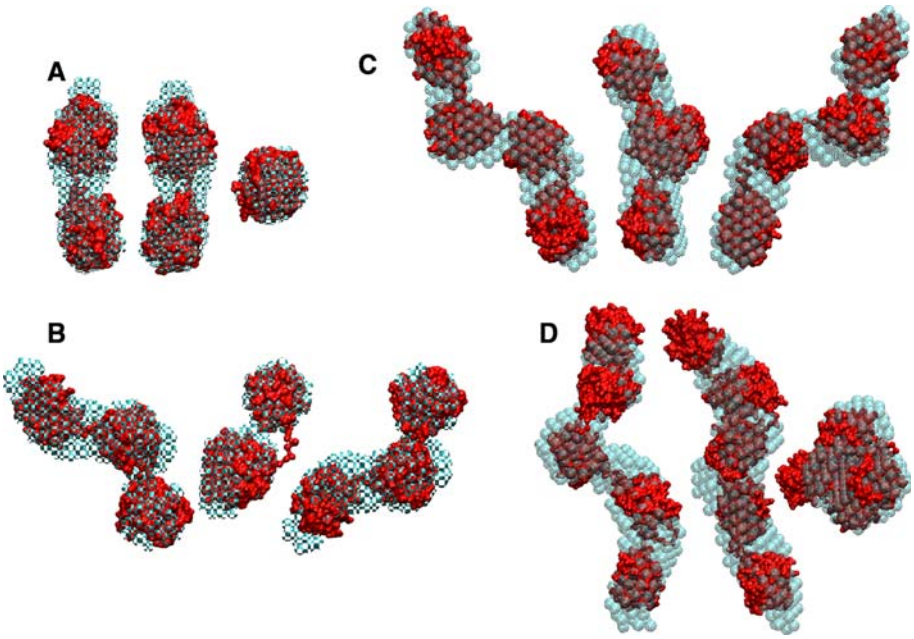

Fig. 3 Structural modeling of the four highly flexible polyubiquitin modular systems simulated. Orthogonal views of the superpositions of the average ab initio reconstructions (*blue envelopes*) and most representative rigid-body modeling solutions with BUNCH (*red envelopes*) for the **a** di-, **b** tri-, **c** tetra-, and **d** penta-ubiquitin proteins

2005). The ubiquitin structure was used as a rigid entity, and the conformation of the linker was optimized in order to fit the synthetic data. Excellent agreement with the SAXS curves was obtained (Fig. 1a). The ranges of χ^2 values obtained monitoring the quality of the fit are presented in Table 2. Therefore, in all cases a single conformation was found which had the capacity to describe the scattering profiles, even though they originated from completely flexible proteins.

The most representative solution among the ten independent runs is shown in Fig. 3 overlapped with the average ab initio shape reconstruction. In all cases the linkers presented relatively extended conformations. In addition, ubiquitin domains appeared in all cases as independent units and no interdomain contacts were observed. As a consequence of these two observations, notably elongated conformations were derived. Highly similar structural models were obtained between the two modeling techniques, as observed in Fig. 3. The structural divergence among the independent solutions was monitored with the average NSD value presented in Table 2. Although ⟨NSD⟩ remained relatively low, BUNCH solutions presented systematically larger divergence than the ab initio ones. In addition, ⟨NSD⟩ values clearly increased with the protein size and the number of degrees of freedom, reaching ⟨NSD⟩ = 1.31 for the penta-ubiquitin system. Close inspection of the two most divergent BUNCH solutions for penta-ubiquitin, NSD = 1.67, showed that the two most external domains had coincident positions, and the three internal domains presented translations within the general envelope.

The effect of instrumental noise on the structure determination by rigid-body methods was studied by using a synthetic curve with a threefold noise level (3X; see the "Methods" section). The quality of the fits was not impaired (data not shown). Despite the lower quality of the

data, the uniqueness of the optimal solution was maintained for all systems, with ⟨NSD⟩ values very similar to those obtained for the less noisy SAXS curves (Table 2). Not surprisingly, the resulting structures were very similar to those obtained for the standard data (results not shown).

Modeling modular proteins with domains of distinct sizes: the *asymmetric* case

A synthetic modular protein comprising four domains of different sizes was also simulated and modeled equivalently to the polyubiquitin chains (see the "Methods" section). The overall size descriptors of the simulated protein are shown in Table 1 (labeled "asymmetric"). Despite the presence of domains of different sizes, a single conformation fitted the scattering profile in both structural modeling strategies. However, slightly worse agreement with the experimental curve was obtained for the ab initio reconstructions (see the χ^2 ranges in Table 2). This observation can be partially attributed to the larger overall size of the system.

The averaged shapes obtained with DAMMIN displayed a very homogeneous density. Consequently, the linker regions and the different domains could not be unambiguously identified within the density (Figure S1 in Supplementary Information). This situation was already observed for flexible polyubiquitin chains and was not ameliorated for the nonsymmetrical modular protein case.

Impact of interdomain dynamics on structure determination

Shape reconstruction and rigid-body structural modeling was performed on the scattering data derived from the previously described static structures. Figure 4 shows the shape reconstructions for the static structures overlapped with the conformations used to calculate the synthetic data. The resulting envelopes correctly described the size and shape of the original structures. In all cases, thicker densities were found along the structure, which were easily assigned to the folded domains. This was confirmed when overlapping the original structure, and excellent coincidence with the ubiquitin domains was found. This observation is exemplified in the di-ubiquitin reconstruction (Fig. 4a), a dumbbell shape where the two lobes were connected through a thin line of density that correspond to the rigid linker. This shape was compared with that derived from the dynamically averaged ensemble that appeared as a homogeneously thick density encompassing both folded domains and the linker (Fig. 3a). In the tetra-ubiquitin case, two high-density lobes were identified and assigned to two di-ubiquitin units. Although less clearly, these lobes did not have a homogeneous density but they did have a depression in the middle that allowed the identification of the individual ubiquitin domains. Ab initio reconstructions were performed using the same procedure but using a shorter momentum transfer range ($s < 0.3$ Å^{-1}); for the $p(r)$ calculation, see Figure S2 in Supplementary Information. The same trends of enhanced resolution with respect to these obtained from the dynamically averaged scattering profiles were also observed.

Ten independent rigid-body structures were also obtained in the static scenario with BUNCH. An excellent fit to the SAXS curves was obtained, as observed in Fig. 1c. The resulting structural models present were compared with the original structures, showing ⟨NSD⟩ of 0.85, 0.99, 1.32, and 1.53 for the di-, tri-, tetra-, and penta-ubiquitin systems, respectively. These values of ⟨NSD⟩ indicate high similarity to the original structure used to generate the data. These results emphasize the reliability of

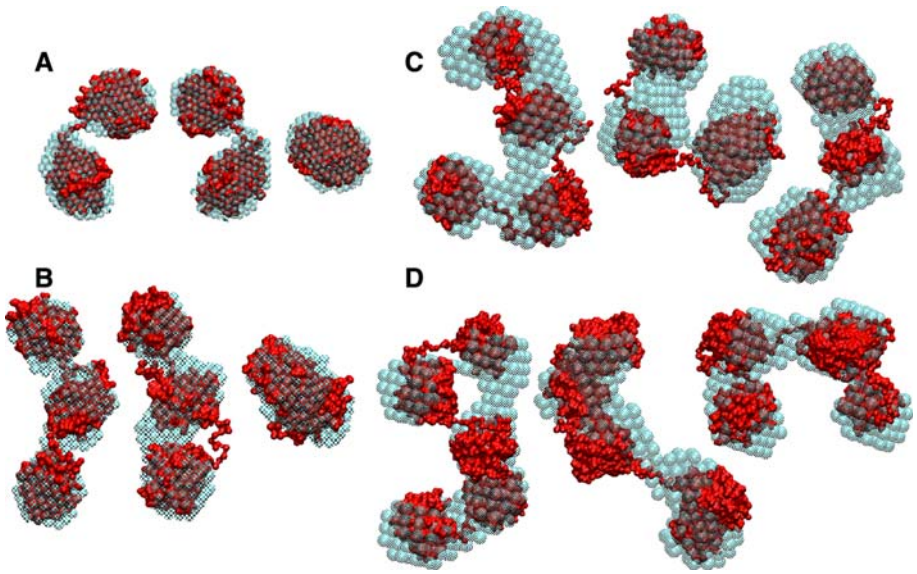

Fig. 4 Ab initio shape reconstructions (*blue envelopes*) calculated from synthetic scattering profiles derived from rigid conformations (*red envelopes*) of **a** di-, **b** tri-, **c** tetra-, and **d** penta-ubiquitin models that have R_g of 22.82, 30.85, 37.30, and 43.20 Å respectively

SAXS for the determination of low-resolution structural models of static modular proteins when the information of the folded regions is available from other sources.

Apparent interdomain distances in rigid-body calculations

Although originating from highly flexible systems probing a large variety of conformations, rigid-body calculations can find a single solution able to describe the SAXS curves perfectly. The question as to whether this rigid optimal conformation is of physical relevance remains open. To evaluate this point, the interdomain distance distributions derived from the ensembles were calculated and compared with those found in the rigid-body modeling (Fig. 5). For all the cases tested, broad distributions were found, exemplifying the complete conformational freedom of the proteins. Interestingly, the optimized structures of the rigid body presented interdomain distances that lay on the most populated regions present in the ensemble, close to maxima in the distributions. This observation was further analyzed by comparing the optimized interdomain distances with the average ones found in the ensembles (straight lines in Fig. 5), where excellent agreement was found. Of relevance, this result was observed for all systems regardless of the size of the chain and the separation between the domains. These results suggest that, for dynamic multidomain proteins, the low-resolution static structures obtained from SAXS data contain valuable structural information as they reflect, in a first approximation, the average domain separations.

Application of the ensemble optimization method (EOM)

A new methodology, the so-called EOM, has recently been developed in order to study potentially flexible systems by

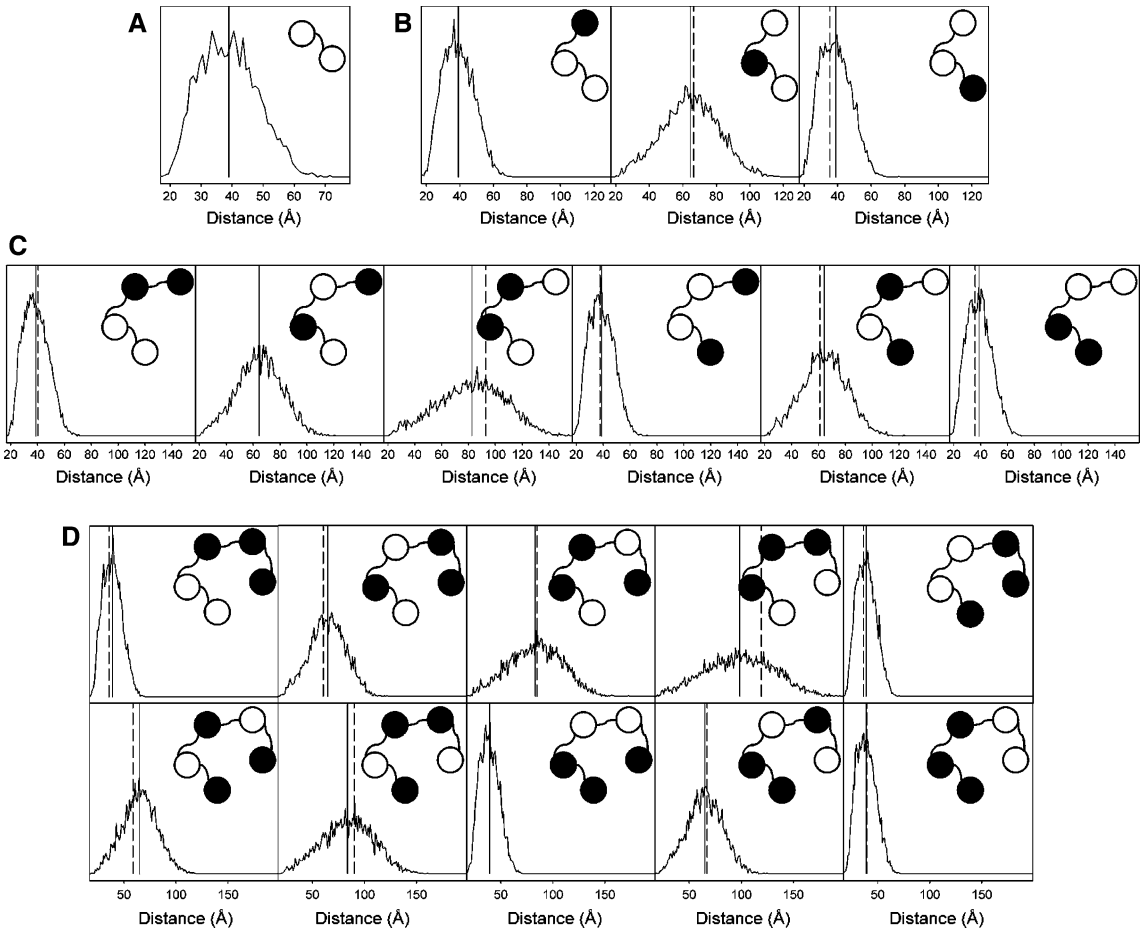

Fig. 5 Domain–domain distances calculated from their respective centers of mass. Interdomain distance distributions (*solid line*) computed from the ensemble of 5,000 structures for the **a** di-, **b** tri-, **c** tetra-, and **d** penta-ubiquitin systems. The two domains used for the calculation are represented as *empty circles* in the cartoon of the molecule. *Solid straight line* is the average interdomain distance calculated from the ensemble. *Dashed straight line* is the average interdomain distance from the ten independent BUNCH calculations for each modular protein simulated. In some of the plots both average distances are coincident and only one *straight line* is observed. The ranges plotted differ for the four simulated systems

SAXS (Bernadó et al. 2007). Using this method, the experimental SAXS curve is assumed to arise from a certain number of coexisting conformational states. The ensemble of selected structures is analyzed in terms of its R_g distribution and compared with that of the random model in order to derive structural information.

The scattering profiles of dynamic and the rigid polyubiquitin chains were submitted to EOM using the previously generated 5,000 random structures as a pool for the genetic algorithm selection, see the "Methods" section. The resulting R_g distributions are displayed in Fig. 6 and compared with these of the different pools. The R_g distributions of the ensembles selected for the flexible multidomain proteins are virtually equivalent to these of the pools, indicating the presence of a broad distribution of protein sizes and the capability of the EOM to detect the dynamic origin of the SAXS curve. Conversely, the R_g distributions arising from the SAXS profiles of the static structures present much narrower distributions and sharp peaks centered around the R_g values of the structures used to simulate the data. These results demonstrate a distinct behavior of SAXS profiles from static and disordered proteins in the EOM analysis, indicating that this strategy is useful for the diagnosis of interdomain motions.

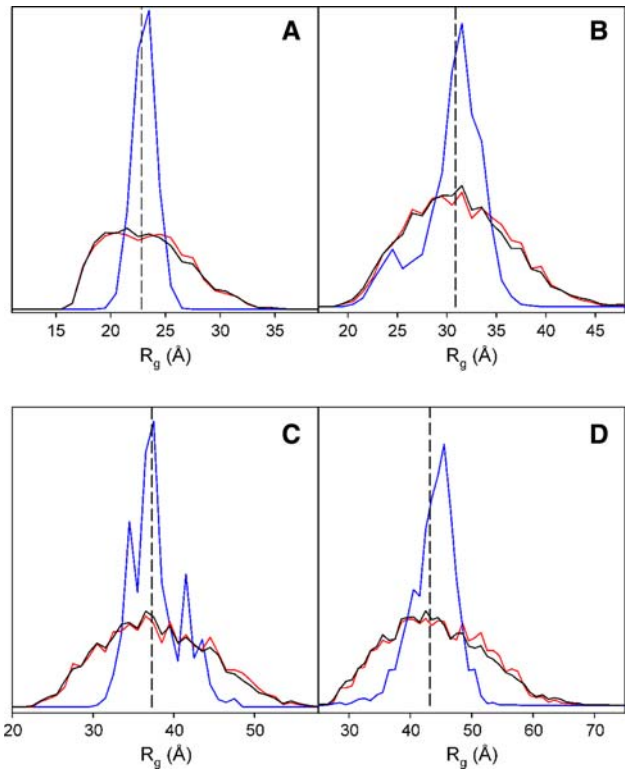

Fig. 6 R_g distributions derived from the EOM analysis of the synthetic SAXS profiles computed for the dynamically averaged ensemble of conformations (*red*), and rigid structures (*blue*) displayed in Fig. 5 for **a** di-, **b** tri-, **c** tetra-, and **d** penta-ubiquitin chains. The *vertical dashed lines* represent the R_g of the conformation used to calculate the synthetic curves

Discussion

The synthetic SAXS profiles computed from large ensembles of polyubiquitin chains provided with complete conformational freedom displayed differential features when compared with those arising from individual static conformers. Dynamic averaging induced a general smoothing of the scattering profiles, and the fine structure expected from the distance correlation of the globular domains was severely weakened. These differential features were enhanced using Kratky plots that present a single maximum in the dynamic scenario, whereas much more complex features with several maxima are observed in the static one. Kratky plots have traditionally been very useful for the detection of unstructured proteins. Flexible multidomain proteins exhibit a dual nature in which globular domains are connected through unstructured regions. In the systems modeled in this study, the feature of globular proteins, a prominent peak, was clearly dominant. However, it is expected that, in systems with much longer linkers, Kratky plots could present an enhanced contribution from the unstructured regions, a continuous increase of $I(s)s^2$, yielding a dual representation (Bernadó et al. 2005; Tsutakawa et al. 2007). The smoothness of the data was also apparent in distance distribution functions, $p(r)$, derived from the Fourier transformation of the scattering profiles. In addition to the intramolecular correlation peak, only a further interdomain distance correlation were observed in flexible systems, and other expected correlations between domains that neatly define the interdomain distances were averaged out.

Low-resolution ab initio and rigid-body structural models were derived from the dynamically averaged scattering profiles. For both methods in all cases a single conformation was found that was able to fully describe the synthetic scattering profile. Although several independent runs were performed, both methods converged to very similar structures, with very low NSD among them. Therefore, obtaining a unique solution is not indicative of a rigid protein.

From the simulations, an apparent loss of resolution was observed in ab initio reconstructions, and long and homogeneous densities were obtained in which the positions of the individual domains could not be assessed. This observation is related to the above-mentioned reduction of fine structure and the loss of interdomain correlation information in the SAXS profiles caused by the dynamic averaging. Therefore large-amplitude dynamics severely aliases shape reconstructions derived from SAXS data.

Rigid-body modeling with BUNCH of dynamically averaged data yielded highly elongated conformations where the globular regions appeared isolated and no interdomain contacts were observed. In order to fulfill the spatial restrictions coded in the SAXS data, rigid-body modeling strategies must place folded domains in concrete positions. Interestingly, globular domains are placed at the average distance of the dynamic ensemble that originated the data. This is not an obvious result, as averaging of scattering profiles is a different process from averaging the position of structural elements in a highly flexible molecule.

The use of the EOM strategy to fit the data provides a clear indication of interdomain motions. The derived R_g distributions for the rigid and flexible polyubiquitin chains display very distinctive features. Whereas the dynamic scenarios present broad R_g distributions, SAXS data from rigid conformations yield much narrower ones that are centered around the theoretical R_g. Presence of large particles detected by EOM could be misinterpreted as arising from oligomeric species coexisting in solution. The analysis of the forward scattering, $I(0)$, and/or the R_g at different concentrations which is routinely performed in most beamlines would discount this possibility. In addition, numerical simulations indicate that SAXS curves containing small amounts of oligomers were not appropriately fitted by EOM when the pool of conformations only contains monomeric molecules (Bernadó et al. 2008).

Present observations can be analyzed in the context of previous SAXS studies of multidomain proteins. SAXS data for three different constructs of Bruton's tyrosine kinase (Btk), a four-domain protein, were modeled, yielding a highly elongated and very homogeneous density when using ab initio methods, and presenting the four domains in remote positions when applying rigid-body modeling (Márquez et al. 2003). According to the results mentioned above, their results would suggest the presence of flexibility. This was recently demonstrated when the same data were modeled using the EOM strategy, and Btk clearly emerged as a highly flexible protein (Bernadó et al. 2007).

In another example, the structural characterization of two-domain full-length metalloproteinases 1 and 12 (MMP-1 and MMP-12) was performed by combining NMR and SAXS (Bertini et al. 2008, 2009). Spin relaxation experiments demonstrate that, in both proteins, the domains have distinct hydrodynamic properties and therefore present interdomain dynamics despite the different length of the linker. EOM analyses for both MMP proteins display broad R_g distributions substantiating the NMR results, and demonstrating the use of EOM for detecting flexibility. Broad R_g distributions from the EOM analysis were also obtained for the BRCA1-associated RING domain protein 1 (BARD1) two-domain fragment, indicating the presence of interdomain mobility (Edwards et al. 2008). This protein also displays a smooth $p(r)$ function with a large D_{max} value, and a single-peak Kratky plot. A similar situation has been observed for the Protein kinase R (PKR) protein that, in addition to the broad R_g distribution obtained from the EOM analysis, presents smooth scattering profiles and $p(r)$ functions with large D_{max} values (VanOudenhove et al. 2009). In this case rigid-body modeling was performed with BUNCH. In agreement with the observations of this study, elongated conformations were obtained for the final BUNCH solutions where the three domains appeared isolated, and no interdomain contacts were observed in this single-conformation analysis (VanOudenhove et al. 2009).

Conversely, for the Z1Z2 tandem from titin protein the absence of pervasive dynamics was demonstrated by spin relaxation and residual dipolar couplings, probably due to the short length of the linker (Marino et al. 2006). In this case, the SAXS profile and the $p(r)$ function present noticeable features. Unfortunately, no EOM analysis is available for this protein. Another example of rigid multidomain protein is c-Src kinase in its inactive state, a three-domain protein whose SAXS profile is completely featureless (Bernadó et al. 2008). According to the present results, this could potentially indicate interdomain dynamics. However, when analyzing the $p(r)$ function, the structural models derived, and the EOM R_g distribution, a rigid and highly compact structure emerges, in agreement with a crystal structure.

In summary, this study clearly shows that SAS profiles obtained from highly flexible proteins can be wrongly interpreted as arising from a rigid structure. As a consequence, features from the SAS data or from the derived structural models that indicate interdomain motions have been identified. Features of SAS data are: (1) general attenuation of fine structure in the scattering profiles that is enhanced in Kratky representations, and (2) a reduction of interdomain correlation peaks in $p(r)$ functions that also display large D_{max} values and a smooth decrease to 0. Structural models derived from these data present characteristic trends: (1) a systematic decrease in the resolution of ab initio models, and (2) rigid-body models yield highly extended conformations with a lack of interdomain contacts. EOM represents an excellent strategy to identify interdomain motions unambiguously when high-resolution models of the globular domains are available. This study provides information that should help researchers to select the best modeling strategy for the structural interpretation of SAS experiments of multidomain proteins.

Acknowledgments The author thanks Prof. Miquel Pons and Jascha Blobel (Institute for Research in Biomedicine, Barcelona), and Dr. Dmitri I. Svergun (EMBL-Hamburg) for insightful discussions. Tanya Yates is acknowledged for careful reading of the manuscript. Financial support form the Spanish Ministry of Education-FEDER (BIO2007-63458) is gratefully acknowledged. The author holds a Ramón y Cajal contract that is partially financed by the Spanish Ministry of Education and by funds provided to the Institute for Research in Biomedicine by the Generalitat de Catalunya.

References

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102:17002–17007

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

Bernadó P, Pérez Y, Svergun DI, Pons M (2008) Structural characterization of the active and inactive states of Src kinase in solution by small-angle X-ray scattering. J Mol Biol 376:492–505

Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun DI (2008) Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. J Am Chem Soc 130:7011–7021

Bertini I, Fragai M, Luchinat C, Melikian M, Mylonas E, Sarti Svergun DI (2009) Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1). J Biol Chem 284:12821–12828

Braddock DT, Louis JM, Baber JL, Levens D, Clore GM (2002) Structure and dynamics of KH domains from FBP bound to single-stranded DNA. Nature 415:1051–1056

Bron P, Giudice E, Rolland JP, Buey RM, Barbier P, Díaz JF, Peyrot V, Thomas D, Garnier C (2008) Apo-Hsp90 coexists in two open conformational states in solution. Biol Cell 100:413–425

Doniach S (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev 101:1763–1778

Dunker AK, Bown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. Biochemistry 41:6573–6582

Edwards RA, Lee MS, Tsutakawa SE, Williams RS, Tainer JA, Glover JNM (2008) The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50. Biochemistry 47:11446–11456

Ekman D, Björklund ÅK, Frey-Skött J, Elofsson A (2005) Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. J Mol Biol 348:231–243

Fetler L, Kantrowitz ER, Vachette P (2007) Direct observation in solution of a preexisting structural equilibrium for a mutant of the allosteric aspartate transcarbamoylase. Proc Natl Acad Sci USA 104:495–500

Gabel F, Simon B, Nilges M, Petoukhov MV, Svergun DI, Sattler M (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. J Biomol NMR 41:199–208

Grela P, Bernadó P, Svergun DI, Kwiatowski J, Abramczyk D, Grankowski N, Tchórzewski M (2008) Structural relationships among the ribosomal stalk proteins from the three domains of life. J Mol Evol 67:154–167

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127:16621–16628

Grishaev A, Tugarinov V, Kay LE, Trewhella J, Bax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. J Biomol NMR 40:95–106

Hammel M, Fierobe HP, Czjzek M, Finet S, Receveur-Brechot V (2004) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. J Biol Chem 279:55985–55994

Hammel M, Fierobe HP, Czjzek M, Kurkal V, Smith JC, Bayer EA, Finet S, Receveur-Brechot V (2005) Structural basis of cellulosome efficiency explored by small angle X-ray scattering. J Biol Chem 280:38562–38568

Heller WT (2005) Influence of multiple well defined conformations on small-angle scattering of proteins in solution. Acta Crystallogr D61:33–44

Improta S, Krueger JK, Gautel M, Atkinson RA, Lefèvre JF, Moulton S, Trewhella J, Pastore A (1998) The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. J Mol Biol 284:761–777

Jacobs DM, Saxena K, Vogtherr M, Bernadó P, Pons M, Fiebig KM (2003) Peptide binding induces large scale changes in interdomain mobility in human Pin1. J Biol Chem 278:26174–26182

Kleywegt GJ (1997) Validation of protein models from Calpha coordinates alone. J Mol Biol 273:371–376

Konarev PV, Volkov VV, Sokolova A, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 36:1277–1282

Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. J Appl Crystallogr 34:33–41

Levitt M (2009) Nature of the protein universe. Proc Natl Acad Sci USA 106:11079–11084

Lim WA (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. Curr Opin Struct Biol 12:61–68

Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand JY, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data. Eur Biophys J 37:95–104

Marino M, Zou P, Svergun DI, Garcia P, Edlich C, Simon B, Wilmanns M, Muhle-Goll C, Mayans O (2006) The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. Structure 14:1437–1447

Márquez JA, Smith CIE, Petoukhov MV, Lo Surdo P, Mattson PT, Knekt M, Westlund A, Scheffzek K, Saraste M, Svergun DI (2003) Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. EMBO J 22:4616–4624

Mulder FAA, Bouakaz L, Lundell A, Venkataramana M, Liljas A, Akke M, Sanyal S (2004) Conformation and dynamics of ribosomal stalk protein L12 in solution and on the ribosome. Biochemistry 43:5930–5936

Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G, Kuriyan J (2006) Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. Mol Cell 21:787–798

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 (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr Opin Struct Biol 17:562–571

Petoukhov MV, Monie TP, Allain FHT, Matthews S, Curry S, Svergun DI (2006) Conformation of polypyrimidine tract binding protein in solution. Structure 14:1021–1027

Pickford AR, Campbell ID (2004) NMR studies of modular protein structures and their interactions. Chem Rev 104:3557–3565

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285

Receveur V, Czjzek M, Schülein M, Panine P, Henrissat B (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. J Biol Chem 277:40887–40892

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 28:768–773

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

Svergun DI, Koch MHJ (2002) Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol 12:654–660

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

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

Taraban M, Zhan H, Whitten AE, Langley DB, Mattews KS, Swint-Kruse L, Trewhella J (2008) Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. J Mol Biol 376:466–481

Trewhella J, Blumenthal DK, Rokop SE, Seeger PA (1990) Small-angle scattering studies show distinct conformations of calmodulin in its complexes with two peptides based on the regulatory domain of the catalytic subunit of phosphorylase kinase. Biochemistry 29:9316–9324

Tsutakawa SE, Hura GL, Frankel KA, Cooper PK, Tainer JA (2007) Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. J Struct Biol 158:214–223

VanOudenhove J, Anderson E, Krueger S, Cole JL (2009) Analysis of PKR structure by small-angle scattering. J Mol Biol 387: 910–920

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol 194:531–544

Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr 36:860–864

von Castelmur E, Marino M, Svergun DI, Kreplak L, Ucurum-Fotiadis Z, Konarev PV, Urzhumtsev A, Labeit D, Mayans O (2008) A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. Proc Natl Acad Sci USA 105:1186–1191

Zou P, Gautel M, Geerlof A, Wilmanns M, Koch MHJ, Svergun DI (2003) Solution scattering suggests cross-linking function of telehonin in the complex with titin. J Biol Chem 278:2636–2644