Chapter 11
Protein shape and assembly studied with X ray solution scattering: Fundaments and practice

Rubén M Buey, Pablo Chacón, José Manuel Andreu and José Fernando Díaz

Abstract Small Angle X ray Scattering (SAXS) is a widely used technique to study non-crystalline systems such as protein solutions. The size and shape of proteins can be determined from the scattering profile at low-medium resolution using computer simulations. The generated structural information perfectly complements high resolution data resulting from other structural biology methodologies such as X ray crystallography or Nuclear Magnetic Resonance (NMR). In this manuscript, the methods employed in a common protein SAXS experiment are briefly reviewed, covering diverse aspects ranging from data collection and analysis to computer modeling and docking procedures.

11.1 Introduction

Small Angle X ray Solution Scattering (SAXS) is becoming a widely used technique in the study of large non-crystalline biological macromolecular systems, i.e. proteins in solution and biological fibers, in a resolution range (in the order of a few nanometers). It allows the study of the conformation of molecules under physiological conditions, and it is also suitable for time-resolved studies. Thus, with SAXS one can follow the response of a biological system to a perturbation in the physical or chemical environment, e.g. rapid dilution, pH-jump, etc. In principle, there are

Rubén M Buey
Paul Scherrer Institute. Villigen PSI. Switzerland. e-mail: ruben.martinez-buey@psi.ch

Pablo Chacón
Centro de Investigaciones Biologicas. CSIC. Madrid. Spain. e-mail: pablo@cib.csic.es

José Manuel Andreu
Centro de Investigaciones Biologicas. CSIC. Madrid. Spain. e-mail: j.m.andreu@cib.csic.es

José Fernando Díaz
Centro de Investigaciones Biologicas. CSIC. Madrid. Spain. e-mail: fer@cib.csic.es

no limitations in the protein size, a unique advantage for studying molecules too small for Electron Microscopy (EM) and too large for Nuclear Magnetic Resonance (NMR). No sample crystals are needed and the systems to be studied may also own substantial flexibility, both of which represent an additional advantage in comparison to X ray crystallography. However, the information content in scattering curves is relatively modest compared with other methods (SAXS resolution is limited to 5 to 1 *nm*, depending on the sample). In summary, it represents an excellent complementary tool to other structural biology techniques, especially, in combination with atomic structures which might provide more complete models of protein conformations and complexes in solution. In this review, we will focus on the methods and applications of SAXS for structural studies of protein solutions, covering aspects of the most common applications, ranging from time-resolved monitoring of protein polymerization to protein 3D *ab initio* modeling . We will not present here all of the numerous applications of SAXS to protein structure described in the literature but rather focus on a practical description of the data analysis, stressing those issues successfully applied in our laboratory. More complete reviews on SAXS applications can be found in the literature [1, 2]. This review does not intend either to exhaustively review the theoretical basis of SAXS but to present some important statements that should be considered when performing a protein SAXS experiment. Several excellent reviews on SAXS theory, methodology and instrumentation can be found in the literature for interested readers [2, 3].

11.2 Protein SAXS basics

This section is organized following the typical steps carried out in a standard protein SAXS experiment.

11.2.1 Data collection

In a SAXS experiment, a solution of proteins is exposed to an X-ray beam, typically with a wavelength of around 0.15 nm, and the scattered intensity is recorded as a function of the scattering vector s ($s = 2\sin\theta/\lambda$, where 2θ is the angle between the incident X ray beam and the scattered radiation), (note that other authors use the momentum transfer $q = 4\pi\sin\theta/\lambda$ instead of s). Since proteins are poor scatters at very low angles, powerful X ray sources are required, such as those found in synchrotrons.

Buffer subtraction. A basic SAXS experiment consists of recording the elastically scattered X-ray photons by a sample of randomly oriented molecules in solution. The measured signal represents the difference in the average electron density of the protein molecules and the solvent. Thus SAXS can be considered a contrast methodology. In an experiment involving a protein solution, it is therefore, neces-

sary to measure the scattering from the solution and the solvent. It is assumed that the solvent has a constant scattering intensity, and this has to be precisely subtracted from the scattering profile of the protein solution. This is achieved by subtracting the necessary amount of buffer to superimpose the protein and the buffer curves at high angle, where the contribution of both should be the same. When possible, it is strongly recommended to use the same buffer in which the sample is (the dialysis buffer or the buffer used to gel filtrate the protein).

Camera length. The scattering function is described by a Fourier transformation by that converts real space coordinates (distances in the sample; *nm*) to reciprocal space coordinates of the scattering vectors $(s; nm^{-1})$. The range of the scattering vector depends on the angle, i.e. on the distance from the sample to the detector, which must be adjusted depending on the distance range needed to cover for each individual experimental set up. To cover a wide range of of scattering vectors, it is recommended to perform the experiments at two different camera lengths (usually in the range between 1 and 6 meters) and merge the spectra afterwards to get the final scattering profile. Care should be taken since the scattering intensity falls off rapidly at high angles. In this context higher concentrations, or longer time exposures, should be used in order to improve signal-to-noise ratio.

Protein concentration. The protein should be dissolved in an appropriate buffer. Most of the commonly used buffers in biochemistry are suitable for SAXS measurements. The protein in the selected buffer should be monodisperse . This can be checked by dynamic light scattering or gel filtration chromatography.

In dilute monodisperse samples, the total scattered intensity will be a continuous isotropic function proportional to the scattering from a single particle averaged over all orientations. However, in semi-dilute conditions (non ideal solutions) an additional contribution to the scattered intensity, due to inter-particle interactions could be detected. The inter-particle contribution might be separated from the particle scattering by performing experiments at different concentrations and/or solvent conditions. It is therefore recommended to measure the protein sample at different concentrations (typically between 1 and 15 mg/mL, although this will depend on the protein). Differences of the scaled scattering curves at multiple concentrations observed at low angles (typically $s > 0.15\ nm^{-1}$) indicates contributions of attractive or repulsive forces. An increase in the scattering intensity at low angle with the protein concentration indicates repulsive inter-particle interactions, whereas an increase in the scattering intensity with concentration indicates attractive inter-particle interactions. The later could be due to either protein aggregation or oligomerization. This inter-particle effects can be minimized by diluting the samples, changing the buffer conditions or extrapolating the data at low angle to zero concentration.

Monodispersity. If the solution is not monodisperse, comprising particles that differ in size and/or shape, the scattering intensity will be determined by the weighted average of the scattering intensities from the different types of particles. Several programs are available that deal with these kind of situations, although a priori information about the shapes of the different species in the mixture is required [4]. These could be especially useful for time resolved SAXS monitoring of processes

involving inter-conversion of species, such as protein polymerization, as it will be described in the section 11.3.2.

Radiation damage. During data collection, the sensitivity of the measured protein to X-ray irradiation must be checked. This can be easily performed by comparison of the SAXS profiles between different time frames in the same measurement. Time frames showing radiation damage (different from the initial one, usually showing higher R_g) should be discarded before averaging. If the protein of interest has a high sensitivity to radiation damage, it is recommended to cool down the sample during measurements, the use of flow cells (if the available amount of protein allows it) or the use of cells that move up and down during the measurement [5]. Besides, even the most stable X ray sources will have significant variations in the beam intensity with time, which must be corrected for proper further data analysis. Therefore, another important issue when processing SAXS data is the normalization by the beam intensity along the time frames of the experiment.

An excellent set of programs has been specifically designed for SAXS data manipulation and analysis in Svergun's group:
(http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html). These programs, collected in the software PRIMUS (table 11.1), greatly facilitate most of the tasks described above in a graphical interface based and user-friendly environment [4].

11.2.2 Data analysis

For dilute monodisperse protein solutions, the random orientation results in a spherical average of the scattering intensity of a single particle in each of the possible orientations. The scattering intensity, represented as a function of the scattering vector, is recorded at very low angle (typically 0.1-10^o) and contains information about the global structure of the proteins in the sample, allowing the estimation of parameters such as the radius of gyration (R_g), the molecular weight (MW) or the maximum intra-molecular distance (D_{max}).

Radius of gyration. The R_g of a protein is defined as the root mean square distance of all the protein atoms from the center of gravity of the molecule. At very low angles (namely in the range of $s < 0.21/R_g$), the relation of $\ln I(s)$ versus s^2, should be linear [6]. This representation is known as Guinier plot and constitutes a very important tool for the SAXS analysis of isotropic monodisperse samples. The interception of this line with the y-axis yields $I(0)$ (also called forward scattering intensity), while the slope yields the R_g (alternatively $p(r)$ (see below) can be used to calculate a more stable computation of R_g). Linearity of the Guinier plot is also used as a test of sample aggregation and homogeneity since R_g should not vary with concentration for well-behaved samples with no inter-particle interference or aggregation.

Molecular weight. In an ideal solution of identical particles, $I(0)$, or the forward scattering intensity, is directly proportional to the number of particles in a sample.

Table 11.1 Computer programs described/used in this review.

	Data collection analysis	
PRIMUS	Primary data analysis and manipulations including buffer subtraction, zero concentration extrapolation, merging different camera length experiments, time frames averaging and Guinier plots. It is also a graphical interface to data analysis programs.	[4]
GNOM	Data processing using the regularisation technique. It evaluates the pair distribution function and calculates the maximun distance and the radius of gyration.	[7]
	ab initio shape determination	
SASHA	Simple shape reconstruction applying spherical harmonics. Symmetry restrictions can be applied.	[10]
DALAI_GA	Shape reconstruction using a genetic algorithm and dummy atoms. No bead connectivity penalty.	[11, 12]
DAMMIN	Shape reconstruction using simulated annealing and dummy atoms. Symmetry restrictions can be applied.	[13]
SAXS_3D	Shape reconstruction using Montecarlo simulations. Uses a ”give and take” algorithm and there is no limitation of search space.	[14]
GASBOR	Shape reconstruction using simulated annealing (similar to DAMMIN) with chain-compatible dummy atom models. It exists a variant that fits the pair distribution function instead of the scattering intensity.	[18]
ELLSTAT	Shape reconstruction using a combination of single shapes, ellipsoids, cylinders, spheres, etc.	[9]
	Docking	
SITUS	Modeling of atomic resolution structures into low-resolution density maps. Converts bead models into standard density map formats and dock atomic structures into them.	[29]
DAMAVER	A set of programs to align *ab initio* models, select the most typical one and build an averaged model. It includes the programs SUPCOMB, DAMSEL, DAMSUP, DAMAVER and DAMFILT.	[19, 20]
MASSHA	Modeling of atomic structures and shape analysis. Interactive and automated rigid body refinement against experimental data.	[30–33]
	Flexibility analysis	
CONCOORD	Generation of protein conformations around a known structure based on geometric restriction	[28]
	Hydrodynamic parameters estimation	
HYDRO	Solution properties of bead models for rigid molecules, including Stokes′ radius, radius of gyration and rotational diffusion as well as sedimentation coefficients.	[21]

After scaling to zero concentration, the forward scattering intensity is directly related to the MW although, contrary to R_g or D_{max} this is just an approximate value. In the practice, the most usual methods to estimate the MW consist in the comparison of the sample to a series of reference proteins, such as bovine serum albumine (BSA) or lysozyme. Another useful and straightforward way of calculating the protein MW will be discussed in the section 11.2.4.

Pair-distribution function and maximum distance. From the scattering intensity profile, the intra-molecular distance probability function, also called pair-distribution function, pair-density distribution function or $P(r)$, can be derived. This function, which in principle contains the same information as the scattering intensity, provides direct information about the distances between electrons in the scattering proteins in the sample, i.e. the probability of finding a certain distance in the studied particle. The $P(r)$ function could, therefore, be calculated directly from the electron density. This function is represented in real space and is much more intuitive for shape estimation by simple visual inspection than the scattering intensity (in reciprocal space). Theoretically, the $P(r)$ function is equal to zero at $r = 0$ and at $r \geq D_{max}$, which is the maximum linear dimension of the scattering protein. In practice, for processing real data, only scattering data at angles (s values) $\geq 1/D_{max}$ should be taken into account. Usually, D_{max} estimation involves iteratively choosing multiple D_{max} values and visual evaluation of the resulting $P(r)$ functions. The $P(r)$ function can be used to calculate R_g and $I(0)$, taking into account all the collected data and not only the small region at very low angles used for the Guinier plot approximation. This is very convenient for samples in which small amounts of aggregates are detected, since these will mainly affect the low angle region. In addition, important information can be extracted by simple visual inspection of the $P(r)$ function. For example, the $P(r)$ function of a sphere of radius x is represented by a Gaussian curve with the maximum placed at x, while the $P(r)$ for a cylinder of radius y will be characterized by a peak with the maximum at y and a long tail.

A computer program, called GNOM [7] (table 11.1), was designed to calculate in an easy and user-friendly way the $I(0)$, $P(r)$ function and R_g, given a maximum distance which should be introduced by the user (either by previous knowledge or iteratively, after evaluating the resulting $P(r)$ functions). Alternatively, the program PRIMUS [4] estimates the R_g from the Guinier plot (table 11.1).

Whereas the small angle portion of a scattering curve contains information about R_g, MW and D_{max}, the higher angle section contains information about the molecular shape. At very high angles, details about the surface and atomic structure will contribute significantly. If the protein is a homogeneous scatterer, it will follow Porods law $(I(s) \propto s^{-4})$ [8], which like the Guinier plot, is only applicable for a portion of the scattering curve (most internal features can be ignored at resolutions lower than $s = 0.3$ nm^{-1}). At very high angles, however, the protein internal heterogeneities contribute significantly and the scattering profiles do not follow Porods law any more.

11.2.3 Ab initio protein shape reconstruction

If the measured sample is monodisperse, then ab initio shape determination can be used to model the low resolution envelope of the protein. Several programs have been created for calculating ab initio shapes from SAXS experiments in the absence of additional external information (table 11.1). Although initial attempts were done to restrict the conformational search to envelopes defined either by spheres, oblates or prolates (ELLSTAT; [9]) or by spherical harmonics (SASHA; [10]), the most recent approximations use beads, also called "dummy atoms", to generate volumes whose scattering profiles fit the scattering data (DALAI_GA; [11, 12]; DAMMIN [13]; SAXS3D; [14]). These programs have been evaluated by different authors and found to reproduce the shape of selected test proteins with high reliability and uniqueness [15–17]. A different implementation also based on dummy atom models is called GASBOR [18], in which the volume is represented by an ensemble of beads forming a chain-compatible model. The spatial positions of these dummy residues aim to approximately correspond to those of the alpha-carbon atoms in the protein structure, thus the number of dummy atoms is the same to the number of amino acids in the studied protein. The program DALAI_GA, developed in our lab, will be described in more detail in section 11.2.4. Successful models generate low resolution envelopes very similar to the electron microscopy reconstructions, although these programs assume that the shape is a continuous object to reduce the computationally expensive conformational space search. Some of these programs also allow imposing symmetry restrictions, which further constrain the solution and can also improve model resolution. Usually, several independent runs (from 8 to 14) are performed with each of these programs and the resulting models are further superimposed in pairs. The most divergent models are then discarded and the most likely protein structure (that one with the lowest average spatial discrepancy, as described in [19]) is determined from the remaining ones. The selected aligned structures are then averaged and filtered. An excellent suite of programs specifically designed for this purpose has been written in the Svergun's group, including SUPCOMB, DAMSEL, DAMSUP, DAMAVER and DAMFILT [19, 20]. The final SAXS-derived low resolution shapes can be either compared to other low/medium resolution structural information available (coming from EM, for example) or high resolution structures can be docked into them (see section 2.4). In the absence of any other structural information, the SAXS models can be checked by estimating their theoretical hydrodynamic parameters and comparing them with experimental analytical ultracentrifugation data. The program HYDRO [21, 22] takes as input file the Cartesian coordinates of the beads representing the SAXS-derived protein envelopes and theoretically estimates some of their hydrodynamic parameters (table I), such as the sedimentation coefficient, the radius of gyration or the volume of the model [21].

11.2.4 Genetic algorithms for SAXS ab initio modeling

As described above, low resolution structures of proteins can be modeled *ab initio* from SAXS data, without other external information. Several algorithms have been implemented to perform this task (table 11.1). In this section, we will describe the implementation of a genetic algorithm in the program DALAI_GA [11, 12], developed at our lab, to solve the inverse scattering problem from protein solutions. In principle, any protein structure can be approximated at any resolution by a set of spheres of small enough diameter, and the solution scattering pattern of such a model structure can be calculated using the Debye formula [23]. This means that the so called *direct scattering problem* (this is, calculation of the scattering profile from the structure) has an analytical solution for any protein in solution. Several CPU-efficient algorithms for the fast computation of the scattering profile of protein structures have been developed in the last decade [24, 25]. In contrast to the direct scattering problem, the *inverse scattering problem* which consists of retrieving the structure of a protein in solution from its x-ray scattering profile, cannot be solved analytically (figure 11.1).

Moreover, the inverse scattering problem has no unique solution, that is, different models can show identical profiles at a given resolution. In 1998, the first CPU-efficient algorithm dealing with this problem was published. It consisted of a genetic algorithm (GA) implemented to exhaustively search for a model compatible with the scattering profile in a predefined region of space discretized in finite particles, i.e. beads of a certain radius. The procedure consists mainly of two steps (figure 11.2): first, fitting the scattering profile computed from sphere packet models of the molecule, using the Debye formula; and second, these bead models are optimized by means of a GA that searches the huge space of possible mass distributions and evolves convergent models. The use of random choice in guiding the search is an advantage of the GA approach ensuring objectivity. In addition, the stochastic character of the GA allows the resolution of ambiguities because each run represents an independent sampling of the configurational space.

A GA consists basically of evolving a population of elements under selection pressure which replicate following genetic rules [26]. First of all, an appropriate initial object must be defined. A certain volume is filled with spheres of a certain radius. The presence of the sphere at a certain position inside the volume is described by the binary value 1, while its absence is defined by the value 0. The volume is now described by an array of binary data, which is called chromosome by analogy to the function of cellular chromosomes (figure 11.2).

An initial chromosome population is randomly generated, transformed into the corresponding models and the scattering profile of each model is calculated and fitted to the experimental scattering profile. Some of the chromosomes (models) with the best fitting are selected at random and recombined by operators simulating genetic mechanisms: a) uniform crossover takes two individuals, so-called parents, and mixes their components (genes) randomly and b) random mutation changes genes of offspring from 0 to 1 or from 1 to 0. Then, the theoretical SAXS curves are calculated for the new set of offspring (figure 11.2), after discarding identical

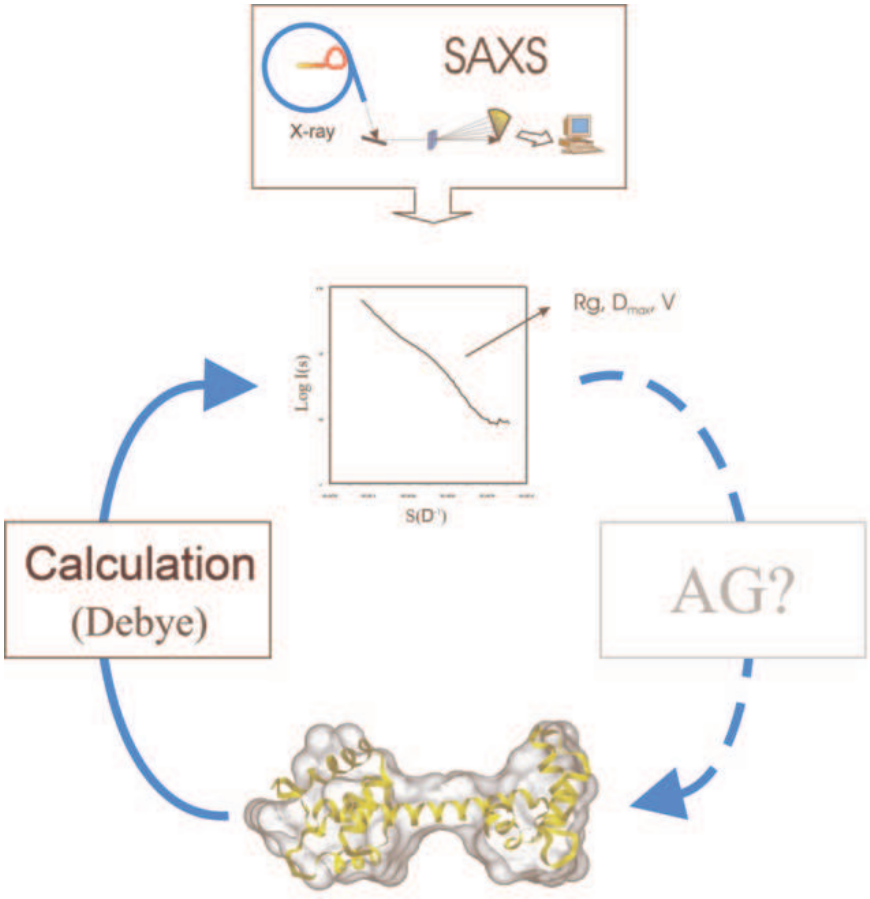

Fig. 11.1 Schematic representation of the direct and inverse scattering problems.

twins. A new ranking order is established and offspring that fit the target profile best are promoted into the breeding stock while unfit ancestors are discarded. The repeated application of genetic operators to the fittest chromosomes increases the average fitness of the population and thus, generates better models. This process is repeated until the system converges, or a good enough solution is found, i.e. a model with a theoretical scattering profile very close to the experimental one. Then, a mask is generated around the best model, constituting the new search space, with a smaller bead radius. The procedure is repeated until reaching a search space with 0.3 *nm* bead radius. In this way, the model resolution is automatically increased by the algorithm. To ensure the model diversity, the procedure should be repeated with at least ten independent parallel executions and the final models compared and merged as commented in section 11.2.3. At the end of the process the shape and the size of the protein is successfully characterized by a single envelop or pseudo

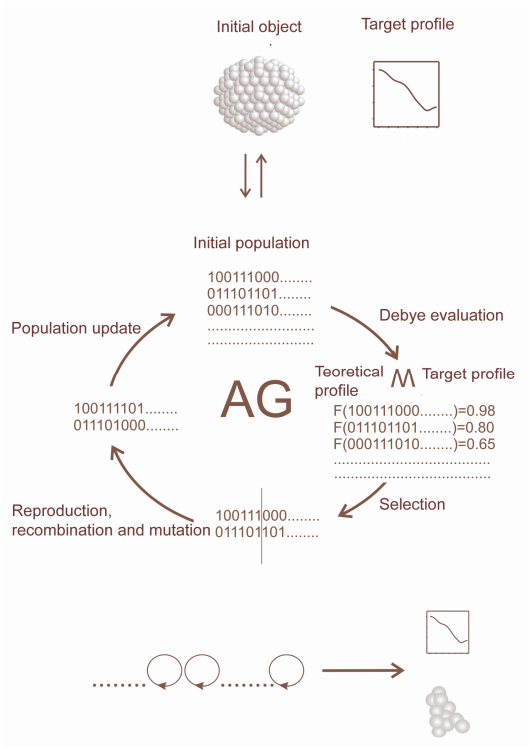

Fig. 11.2 Scheme of the implementation of the genetic algorithm method of SAXS simulation to numerically solve the inverse scattering problem.

density map although in some cases other alternative envelops must be considered. In this degenerate case, additional experimental information is needed to select the correct one.

In addition to the robust shape reconstruction, an interesting observation is the existence of a very good correlation between the number of beads of the SAXS-derived *ab initio* models and the protein *MW*. Thus, the *MW* of a given protein can be estimated from the inverse correlation (figure 11.3). The limiting error in *MW* calculations inherent for this bead modeling method is around 1 *kDa*, estimated using 5 different proteins [12].

11.2.5 SAXS-based 3D modeling and docking

An important challenge in structural biology is the structure determination of protein complexes. When atomic resolution structures of the individual components are available, SAXS, like EM, becomes a crucial technique to calculate the orientation

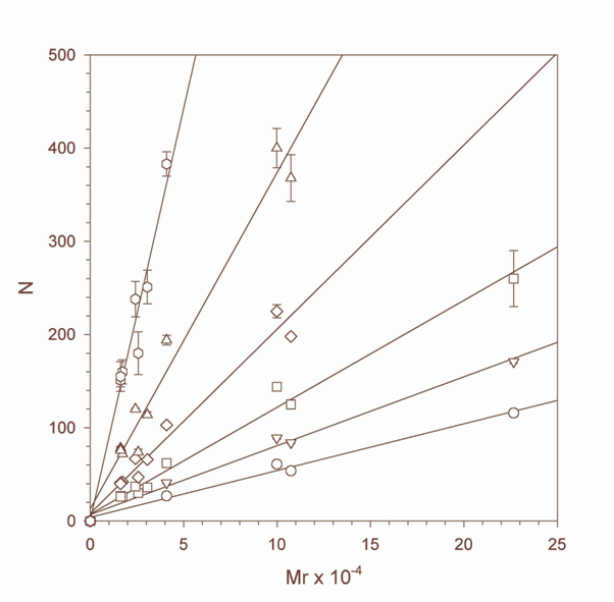

Fig. 11.3 Size of solution scattering models. The number of beads constituting the models is plotted against the molecular mass of each protein sequence. This is made clockwise with bead radius 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8 nm. The correlation coefficients are above 0.96 for all the lineal regressions. Taken from [12].

and arrangement of these components in the complex. Additionally, it is not unusual to find crystal structures that differ in the arrangement of modules and/or domains in comparison to the solution structure. Moreover, these differences are often important to decipher the biological function of the protein, which often involves domain rearrangements and flexibility [27].

Several computational approaches can be used to generate an ensemble of possible conformations around a single conformational state (for example, from a crystal structure), accounting for the potential flexibility of the protein in solution. The generated conformations might be further used for docking into the experimental SAXS-derived envelopes. These computational methods include normal mode analysis, molecular dynamics simulations and non-dynamical methods based on geometric restrictions implemented in the program CONCOORD [28].

There are two major ways of combining high resolution structures with SAXS data. The structures might be either directly docked into the SAXS-derived protein envelopes or used to generate models of the complex by best fitting the theoretical and experimental scattering profiles. These modeling methodologies are independent and can be performed in parallel; when similar results are obtained, the reliability of the proposed models increases.

In principle, any SAXS envelope can be treated as an EM model, so most of the software available for the later is also applicable to SAXS envelopes, after appro-

priate file formatting. The software package SITUS [29] includes SAXS-specific software where rigid-body and flexible docking of structures to SAXS bead models can be performed. Alternatively, other programs (table 11.1) based on graphical interfaces have been specifically created to couple rendering of SAXS low resolution envelopes to computational module that allow interactive and automated rigid body refinement of the available individual high resolution protein structures relative to the solution scattering data, such as MASSHA [30–32].

11.3 Applications of SAXS for structural analysis of proteins

11.3.1 Structures of tubulin polymers revealed by SAXS

Microtubules are dynamic components of the cytoskeleton essential in cellular organization and main constituents of the mitotic spindle. They are formed by reversible assembly of the $\alpha\beta-$tubulin heterodimer into long hollow cylinders, typically made of 13 protofilaments, to whose outer surface microtubule-associated proteins and cytoplasmic motors bind. In addition to these structures, tubulin is also able to form other kinds of polymers in solution, depending on the nucleotide bound and the environmental conditions, such as temperature and/or presence of divalent ions. SAXS studies of tubulin and its polymers have been described in a number of papers in our lab [34–39].

Tubulin heterodimers. Homogeneous tubulin heterodimers studied by SAXS show a radius of gyration, 3.1 ± 0.1 *nm* [34,40]. Sample aggregation can be detected by incubating the protein for 2-5 hours at room temperature, where the R_g increases to a value of 3.7- 3.8 nm. A very simple model of the tubulin heterodimer derived from these studies consists of two adjoining spheres of 22 nm radius each [34].

Microtubules. Under the appropriate *in vitro* conditions, isolated tubulin is able to polymerize into microtubules. The typical SAXS profile of a microtubule at 3 *nm* resolution (s maximum approximately 0.3 nm^{-1}) is made up of six main peaks (figure 11.4; the numbers above the discontinuous line show the peak positions). This profile had been analyzed by reference to their fiber diffraction diagrams, interpreted by the helical diffraction theory [41]. The three first peaks (at lowest angles, J_{01}, J_{02} and J_{03}) correspond to the subsidiary maxima of the J_0 Bessel function of order *n* and represent the low resolution transform of the excess electron density of the cylindrical structures with respect to the solvent. These maxima correspond to the reciprocal radial space, i.e. microtubule diameter. Thus, from the positions of these maxima, the average protofilament number of microtubules can be extracted [35]. The three peaks (J_{12}, J_3 and J_9) at higher angle are sensitive to the microtubule surface lattice and to the shape of the monomer [35]. They correspond to the equatorial J_n Bessel function and the helical J_3 and J_{n-3} Bessel functions, arising, respectively, from the electron density periodicity around an $n-protofilaments$ microtubule and

from the features in the direction of the three-start and the $n-3-start$ helices in the microtubule lattice.

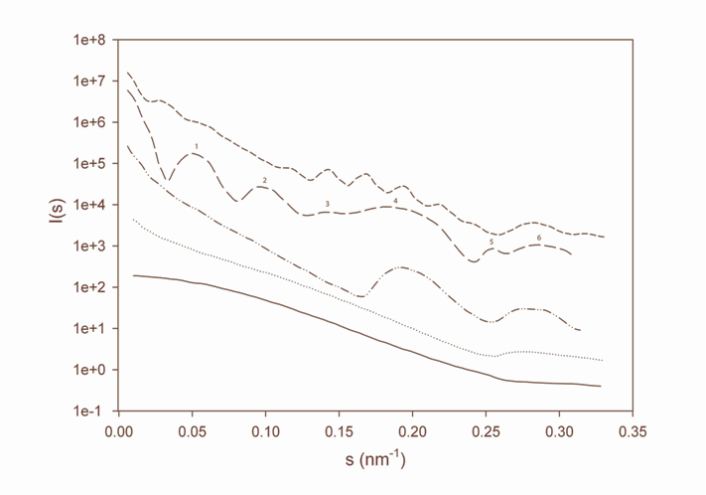

Fig. 11.4 Computed X-ray profiles of model tubulin assemblies. From up to down: tubulin double rings, microtubules, bidimensional sheets, oligomers, and dimers. Numbers above the microtubule scattering curve indicate the positions of the maxima for the six main peaks in this profile, as described in the text.

Comparisons of the positions of these maxima have been used to study the structure of microtubules assembled from purified tubulin in presence or absence of microtubule stabilizing ligands, such as taxol or taxotere. The structure of microtubules polymerized in the absence of ligands, is indistinguishable from the ligand-induced microtubules at high scattering angles, but the position of the three maxima at low angles is displaced toward higher s values in the case of taxol-induced microtubules (figure 11.5), indicating a lower average radio. In fact, taxol-induced microtubules consist of 12 instead of 13 protofilaments, which is the most abundant population in docetaxel-induced microtubules as well as in the absence of drugs [36]. These data were additionally corroborated by EM.

Tubulin rings. GDP liganded tubulin, which is inactive for microtubule assembly, polymerizes into rings more readily than the GTP liganded protein, this process being facilitated by divalent cations. The SAXS profile of these tubulin rings shows characteristic maxima of a closely packed double ring of 38 *nm* diameter with a 5.5 *nm* mean spacing between the rings, and a 4.2 *nm* centre-to-centre spacing between tubulin monomers within both rings [38]. This means that the external ring has 32 monomers, while the internal ring has 24 (figure 11.6). These numbers exactly correspond to the number of tubulin monomers found in a recent study showing cryo-EM reconstructions, at 1.2 nm resolution, of double layered tubes closely resembling the tubulin rings [42, 43].

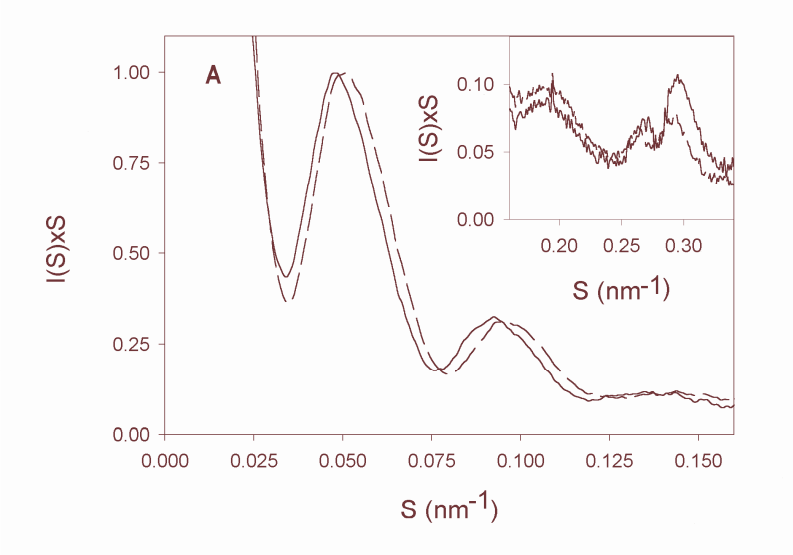

Fig. 11.5 Taxoid induced changes in microtubule X-ray scattering. The corrected intensity profiles ($I(s) * s$) of microtubules assembled from pure tubulin in presence (dashed line) or absence (solid line) of taxol. Taken from [39].

11.3.2 Time-resolved SAXS studies of tubulin polymerization

Tubulin self-assembly can also be examined by time-resolved X-ray scattering. The addition of microtubule stabilizing drugs, such as taxol, and the increase of tem-

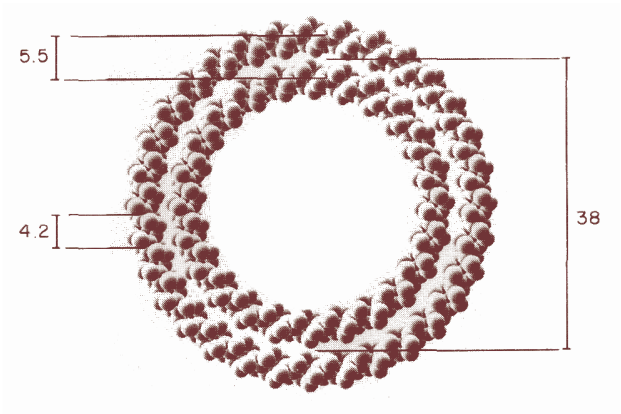

Fig. 11.6 Computer model of tubulin double rings. The measured mean diameter, the centre-to-centre spacing between monomers within each ring, and the spacing between both rings are indicated in nm units. Taken from [38].

perature from 2 to 37^oC induce microtubule polymerization from purified tubulin, providing a rigorously defined model system for the study of microtubule assembly. In this system, polymerization can be followed by monitoring the increase in the central scattering as well as the appearance of the maxima corresponding to the J_0 Bessel function (figure 11.7), corresponding to the low resolution transforms of the hollow cylindrical structures [35]. Moreover, the low resolution structures of the tubulin dimer, tubulin double rings and microtubules (see above), together with the computed scattering profiles of model tubulin assemblies (sheets, oligomers, etc; see scattering profiles in figure 11.4) can be compared to the experimental ones detected at the different time frames during the course of tubulin polymerization, allowing the determination of the structural intermediates and the pathways of self-association of purified tubulin into microtubules and/or tubulin rings [37].

In this way, it seems that taxoid induced microtubule polymerization proceeds via the initial formation of open microtubule sheets (monitored by the central scattering and the maximum corresponding to the J_n Bessel function) which slowly close into microtubules (monitored by the appearance of their characteristic J_0, J_3 and J_{n-3} Bessel function maxima [37]. These data provided evidence for a bidimensional assembly of microtubule polymers in solution (against the helical growth theory). Interestingly, similar results were obtained using cryo-EM: tubulin sheets curving outwards at the end of growing microtubules followed by cylindrical closure [42, 44].

11.3.3 Ab initio SAXS modeling of oligomeric structures of choline binding proteins

Phosphocholine moieties, a common component of the pneumococcal cell wall, are recognized by a set or surface proteins, so called choline binding proteins (CBPs). Choline recognition attaches these proteins to the bacterial cell surface and is essential for their function and may also be a determinant of their quaternary structure. Members of the CBP family involve pneumococcal cell wall murein hydrolases, involved in cellular recognition and some bacteriophage encoded enzymes [45].

All CBPs known to date are modular proteins, composed of at least two separate modules; the functional or enzymatic module, usually placed at the N-terminus and the recognition module or choline binding module (CBM) which targets selectively the protein to the cell wall. Despite the importance of these enzymes in infectious diseases, only a few high resolution structures are available up to date: C-LytA's CBM dimer and full length monomeric Cpl-1 and Pce [46–49]. To our knowledge, no crystallization of full length dimeric proteins has been achieved to date. Since the dimeric MW of Cpl-1 protein is around 80 kDa, it is very difficult to analyze its structure either by NMR (too large) or by EM (too small). In this cases, SAXS provided a unique opportunity to study the quaternary structure of these proteins, as we will describe in this section.

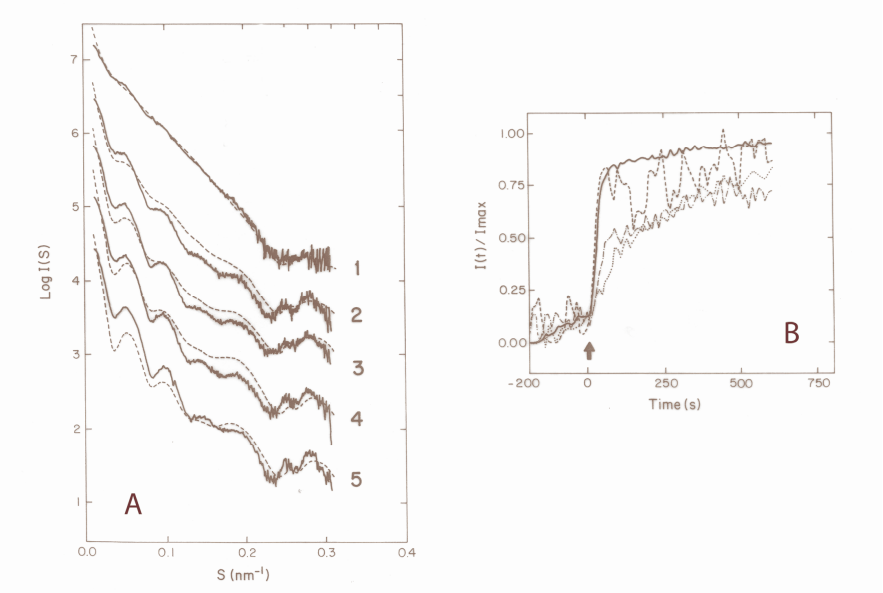

Fig. 11.7 X-ray scattering profiles during taxol-induced tubulin assembly. A. Curve 1, at 2^oC: curves 2 to 5, average scattering profiles between seconds 0-200, 200-400, 400-600 and 1400-2000, after warming the solution to 37^oC. The dashed lines are model mixtures of single linear oligomers and microtubules. All possible mixtures of the two species in 5% steps were calculated, and those giving the least-squares deviation are shown. B. Solid, dash, point and dash-point lines, respectively: time curse of the intensities of the central scattering, J_{12}, J_{01} and J_{02} peaks in the experiment of A. Taken from [37].

In addition, the possibility to directly compare the theoretical SAXS profile of the X-ray crystallographic atomic models to the experimental curves has been demonstrated to be a very useful tool for discerning whether the crystal lattice influences the observed structure. It is generally assumed that a crystal structure is in the lowest energy state under the crystallization conditions and inside the crystal lattice, but this might be not indicative of the lowest energy conformation for the protein in solution. As described next, SAXS is a very useful technique to address this issue.

The SAXS-derived ab initio 3D model of Cpl-1 monomer, although at low resolution, showed a different organization of the two protein modules in solution compared to the crystal structure. This suggests that Cpl-1 structure exhibits some inherent flexibility that allows certain reorientation of the modules. This was corroborated in two different ways: by comparing of the theoretical SAXS profile of the crystal structure to the experimental one and by *ab initio* modeling and docking. On the basis on these results, it was proposed that the intrinsic flexibility of the Cpl-1 protein, facilitated by the presence of a long hydrophilic linker connecting both modules, might be relevant for the protein function by allowing the simultaneous interaction of the enzyme with the substrate (the glycanic chains to be hydrolyzed)

and the choline residues (of (lipo)teichoic acids) used as a docking station on the cell wall [27].

The intrinsic flexibility of the Cpl-1 protein was then computationally studied by using the software CONCOORD [28]. The structures generated by CONCOORD around Cpl-1 crystal structure (root mean square deviations on backbone atoms between 0.8 and 3.5 Å) could not be directly fitted into the SAXS envelope. However, essential dynamics analysis of the CONCOORD simulation showed that the relative orientation of the protein modules might be reorganized while maintaining the hydrophobic interface between the CBM and the catalytic modules. Automated rigid body modeling, considering the catalytic and CMB as independent rigid bodies, using the software MASSHA [30, 31] yielded structures that perfectly fitted the experimental SAXS profile but that disrupted the interface between modules, which should not occur in solution, according to the CONCOORD computational flexibility studies. Taking into account these restrictions, a model was built by manually modifying the torsion angles of the residues in the N-terminal part of the CBM. This model perfectly fits the experimental SAXS data and mostly preserves the interface between the two modules as shown by the crystal structure, in agreement with the flexibility studies (figure 11.8).

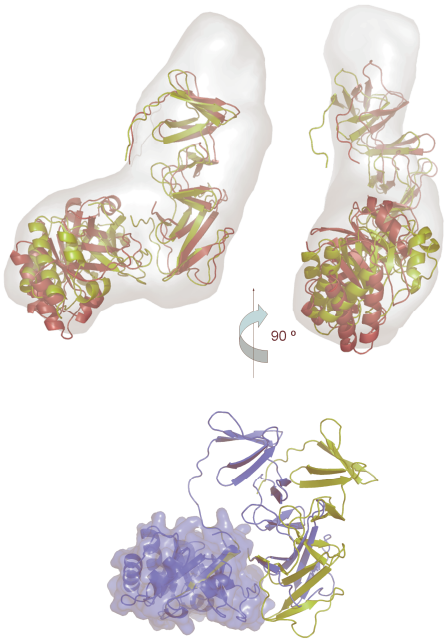

Fig. 11.8 Low resolution envelope of the monomer of Cpl-1 in solution without choline. Upper panel: docking of the manually generated (in green) and MASSHA generated (in red) structures into the *ab initio* SAXS model. Lower panel: superimposition of the crystal structure (PDB accession number 1h09, in blue) to the manually-generated model (in green), showing the different orientation of the modules. Taken from [27].

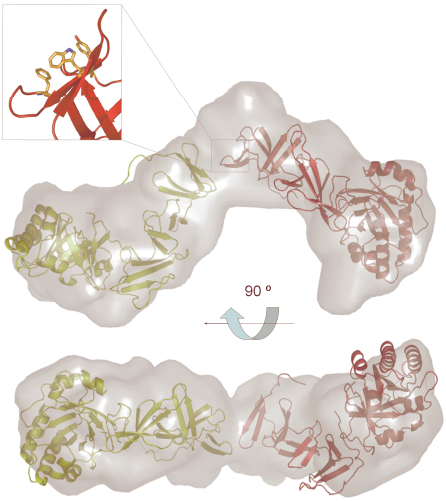

Fig. 11.9 Low resolution envelope for the dimmer of Cpl-1 in the presence of choline and docking of the solution dimeric model. Boxed picture. Detail of the dimerization region showing the side-chain of the aromatic residues forming the non-canonical choline binding site. Taken from [27].

Upon choline binding, Cpl-1 self-associates into a dimer, as demonstrated by both SAXS and analytical ultracentrifugation experiments [27]. Despite the importance of the dimer formation for Cpl-1 function regulation [50], no crystal structure of the model is available. At the SAXS resolution, the tertiary structure of the monomeric and dimeric protein is essentially identical. The SAXS *ab initio* models allowed to map the dimerization interface unequivocally at the N-terminal end of the CMB (figure 11.9), where a putative non-canonical choline binding site was found (figure 11.9 inset), explaining why choline induces dimerization of this protein. Additionally, the SAXS model for the Cpl-1 dimer was corroborated by comparison of the theoretical hydrodynamic parameters to the experimental ones, obtained in similar conditions to the SAXS experiments, by using the software HYDRO [21], which further confirmed the reliability of the model.

SAXS technique emerges as a powerful low/mid-resolution approach, complementary to molecular microscopy and X-ray crystallography. It is applicable to monodisperse randomly oriented species, from relatively small to large proteins. As we show with the illustrative examples of the oligomeric structures of choline binding proteins and tubulin polymerization, SAXS modeling permitted the shape and the size *ab initio* determination in solution of specimens in different conditions. It also provides time resolved information and makes possible to characterize flexibility aspects of these systems and to correlate them with the available atomic structures.

We hope this mini practical review will boost the interest in the undemanding (at least less than others) and relatively new SAXS approaches.

Aknowledgments

We thank Dr. John H. Missimer and Dr. Michel Steinmetz for critical reading of the manuscript. Work in the authors laboratory was supported by grants BIO2007-61336 from Ministerio de Educación y Ciencia to JFD, BIPPED-CM from Comunidad de Madrid to JMA, PC and JFD and BFU2005-00505/BMC from Ministerio de Educación y Ciencia to JMA."

References

1. Lipfert, J., Doniach, S. Annu Rev Biophys Biomol Struct 36, 307-327(2007)
2. Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A. Q Rev Biophys 40, 191-285(2007)
3. Koch, M.H., Vachette, P., Svergun, D.I. Q Rev Biophys 36, 147-227(2003)
4. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I. J App Cryst 36 1277-1282(2003)
5. Kuwamoto, S., Akiyama, S., Fujisawa, T. Journal of Synchrotron Radiation 11, 462-468(2004)
6. Guinier, A., Fournet, F.: Small angle scattering of X-rays. Wiley Interscience. New York (1955)
7. Svergun, D. J App Cryst 25, 495-503(1992)
8. Porod, G. Kolloid Zeitschrift 124, 83-114(1951)
9. Heller, W. J App Cryst 39, 671-675(2006)
10. Svergun, D.I., Volkov, V.V., Kozin, M.B., Stuhrmann, H.B. Acta Cryst Sec A 52 419-426(1996)
11. Chacon, P., Moran, F., Diaz, J.F., Pantos, E., Andreu, J.M. Biophys J 74, 2760-2775(1998)
12. Chacon, P., Diaz, J.F., Moran, F., Andreu, J.M. J Mol Biol 299, 1289-1302(2000)
13. Svergun, D.I. Biophys. J. 76, 2879-2886(1999)
14. Walther, D., Cohen, F.E., Doniach, S. J App Cryst 33, 350-363(2000)
15. Zipper, P., Durchschlag, H. J App Cryst 36, 509-514(2003)
16. Durchschlag, H., Zipper, P., Krebs, A. Journal of Applied Crystallography 40, 1123-1134(2007)
17. Takahashi, Y., Nishikawa, Y., Fujisawa, T. J App Cryst 36, 549-552(2003)
18. Svergun, D.I., Petoukhov, M.V., Koch, M.H.J. Biophys. J. 80, 2946-2953(2001)
19. Kozin, M.B., Svergun, D. J App Cryst 34, 33-41(2001)
20. Volkov, V.V., Svergun, D. J App Cryst 36,860-864 (2003)
21. Garcia de la Torre, J., Navarro, S., Lopez Martinez, M.C., Diaz, F.G., Lopez Cascales, J.J. Biophys. J. 67, 530-531(1994)
22. Garcia de la Torre, J., del Rio Echenique, G., Ortega, A. J. Phys. Chem. B 111, 955-961(2007)
23. Debye, P.: Zerstreeung von röntgenstrahlen. Ann Phys 46,809-823 (1915)
24. Pantos, E., van Garderen, H.A., Hilbers, P.A.J., Beelen, T.P.M., van Santen, R.A. J. Mol. Struct. 383, 303-308(1996)
25. Svergun, D.I., Barberato, C., Koch, M.H.J. J App Cryst 28, 768-773(1995)
26. Davis, L.: Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
27. Buey, R.M., Monterroso, B., Menendez, M., Diakun, G., Chacon, P., Hermoso, J.A., Diaz, J.F.: J Mol Biol 365, 411-424(2007)
28. de Groot, B.L., van Aalten, D.M., Scheek, R.M., Amadei, A., Vriend, G., Berendsen, H.J.: Proteins 29, 240-251(1997)
29. Wriggers, W., Birmanns, S. J Struct Biol 133, 193-202(2001)
30. Konarev, P.V., Petoukhov, M.V., Volkov, V.V., Svergun, D.I. J App Cryst 39 ,277-286(2006)
31. Konarev, P.V., Petoukhov, M.V., Svergun, D.I. J App Cryst 34, 527-532(2001)
32. Kozin, M.B., Volkov, V.V., Svergun, D.I. J App Cryst 30, 811-815(1997)

33. Petoukhov, M.V., Svergun, D.I. Biophys. J. 89 , 1237-1250(2005)
34. Andreu, J.M., Garcia de Ancos, J.G., Starling, D., Hogkingson, J.L., Bordas, J. Biochemistry 28, 4036-4040(1989)
35. Andreu, J.M., Bordas, J., Diaz, J.F., Garcia de Ancos, J., Gil, R., Medrano, F.J., Nogales, E., Pantos, E., Towns-Andrews, E. J Mol Biol 226, 169-184(1992)
36. Andreu, J.M., Diaz, J.F., Gil, R., de Pereda, J.M., Garcia de Lacoba, M., Peyrot, V., Briand, C., Towns-Andrews, E., Bordas, J. J Biol Chem 269, 31785-31792(1994)
37. Diaz, J.F., Andreu, J.M., Diakun, G., Towns-Andrews, E., Bordas, J. Biophys J 70, 2408-2420(1996)
38. Diaz, J.F., Pantos, E., Bordas, J., Andreu, J.M. J Mol Biol 238, 214-225(1994)
39. Diaz, J.F., Valpuesta, J.M., Chacon, P., Diakun, G., Andreu, J.M. J Biol Chem 273, 33803-33810(1998)
40. Bordas, J., Mandelkow, E.M., Mandelkow, E. J Mol Biol 164, 89-135(1983)
41. Klug, A., Crick, F.H.C., Wyckhoff, H.W. Acta Crystallogr 11, 199-213 (1958)
42. Wang, H.W., Nogales, E. Nature 435, 911-915(2005)
43. Nogales, E., Wang, H.W., Niederstrasser, H. Curr Opin Struct Biol 13, 256-261(2003)
44. Chretien, D., Fuller, S.D., Karsenti, E. J Cell Biol 129, 1311-1328(1995)
45. Lopez, R., Garcia, E. FEMS Microbiol Rev 28 , 553-580(2004)
46. Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J.L., Menendez, M. Structure 11, 1239-1249(2003)
47. Hermoso, J.A., Lagartera, L., Gonzalez, A., Stelter, M., Garcia, P., Martinez-Ripoll, M., Garcia, J.L., Menendez, M. Nat Struct Mol Biol 12, 533-538(2005)
48. Fernandez-Tornero, C., Lopez, R., Garcia, E., Gimenez-Gallego, G., Romero, A. Nat Struct Biol 8, 1020-1024(2001)
49. Fernandez-Tornero, C., Garcia, E., Lopez, R., Gimenez-Gallego, G., Romero, A. J Mol Biol 321, 163-173(2002)
50. Varea, J., Saiz, J.L., Lopez-Zumel, C., Monterroso, B., Medrano, F.J., Arrondo, J.L., Iloro, I., Laynez, J., Garcia, J.L., Menendez, M. J Biol Chem 275, 26842-26855(2000).