Synergistic use of synchrotron radiation techniques for biological samples in solution: a case study on protein-ligand recognition by the peroxisomal import receptor Pex5p

W. A. Stanley,[a] A. Sokolova,[b] A. Brown,[c] D. T. Clarke,[c] M. Wilmanns[a]* and D. I. Svergun[a,b]

[a]EMBL-Hamburg, c/o DESY, Notkestraße 85, 22603 Hamburg, Germany, [b]Institute of Crystallography, Russian Academy of Sciences, Moscow 117333, Russia, and [c]Synchrotron Radiation Department, CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK. E-mail: wilmanns@embl-hamburg.de

Circular dichroism spectropolarimetry and X-ray scattering data, obtained using synchrotron radiation, can yield information about the secondary and tertiary structure of proteins in solution. These techniques have been used to analyse the architecture and shape of a complex of two proteins in solution. The crystal structures of two separate proteins, the C-terminal domain of Pex5p and SCP2, are available but their complex has not previously been structurally characterized. Circular dichroism spectropolarimetry indicated that complex formation requires little secondary structure rearrangement. X-ray scattering data fit an elongated irregular 'shoe'-shaped particle of the complex of the two proteins, with dimensions of the order of 30 Å × 40 Å × 90 Å. Comparison with the known crystal structures suggests that this 'shoe' shape requires a conformational change of the C-terminus of SCP2 to appropriately locate its peroxisomal targeting signal type-1 recognition motif into the binding pocket of the Pex5p receptor. Implications of the combined use of synchrotron-based circular dichroism spectropolarimetry and X-ray scattering in structural biology and proteomics are discussed.

Keywords: SRCD; WAXS; SAXS; Pex5p; SCP2; protein structure; peroxisomal targeting; PTS1.

1. Abbreviations

BSA: bovine serum albumin.
CD: circular dichroism spectropolarimetry.
CSA: (+)-10-camphosulphonic acid.
DA: dummy atom.
d_{max}: maximum particle dimension.
DR: dummy residue.
DTT: 1,4-dithio-DL-threitol.
PDB: Protein Data Bank.
Pex5p: peroxin 5 protein, import receptor for PTS1 peroxisomal matrix proteins.
Pex5p(C): peroxin 5 protein C-terminal fragment, residues 315–639.
$p(r)$: distance distribution function.
PTS1: peroxisomal targeting signal type 1.
R_g: radius of gyration.
s: momentum transfer vector, $\mathbf{s} = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 1.5$ Å is the incident X-ray wavelength.
SAXS: small-angle X-ray scattering.
SCP2: sterol carrier protein 2, PTS1 containing protein.
SRCD: synchrotron radiation circular dichroism spectropolarimetry.
TPR: tetratricopeptide repeat motif.
WAXS: wide-angle X-ray scattering.
UV: ultraviolet light.
VUV: vacuum ultraviolet light.

2. Introduction

Molecular biology seeks to understand living cells by examining the structure and function of individual biomolecules in the cell. Often these biomolecules do not act alone but are specifically incorporated into higher-order assemblies, or 'molecular machines'. Considerable effort is expended on understanding their structure–function relationships, particularly those of protein–protein complexes, at various levels of resolution and fidelity.

Synchrotron radiation circular dichroism spectropolarimetry (SRCD) has emerged as a powerful technique for evaluating the secondary structure of proteins in solution (Wallace, 2000; Wallace & Janes, 2001; Wallace *et al.*, 2003). Possible applications of SRCD exceed those from conventional CD spectropolarimeters, which use low-intensity light sources, thus limiting secondary structure analysis to a narrow spectral range in the far-UV (190–250 nm). Synchrotron sources provide a higher photon flux ($\sim 10^3$–10^5-fold), allowing high-quality data to be extended to the VUV range ($\sim$150–250 nm) with reduced noise from buffer components. The increased spectral range allows additional polypeptide backbone electron transitions to be monitored, hence allowing less ambiguous assignment of secondary structure.

Small-angle X-ray scattering (SAXS) can provide information on the shape of proteins in solution. In this technique, the intensity of scattered X-rays, $I(\mathbf{s})$, is measured as a function of the momentum transfer vector, $\mathbf{s} = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the incident X-ray wavelength, usually $\sim$1.5 Å, depending on the beamline used (Svergun & Koch, 2002; Koch *et al.*, 2003). Diverse protein systems have benefited from SAXS analysis, recent examples including a study of the domain structure of the multidomain Bruton tyrosine kinase (Marquez *et al.*, 2003); viral capsid assembly (Sokolova *et al.*, 2001); protein-RNA distribution in a bacterial 70s ribosome (Svergun & Nierhaus, 2000); and, in combination with geometric docking simulations, analysis of the purine nucleoside phosphorylase trimer (Filgueira de Azevedo *et al.*, 2003). Depending on experimental conditions, the method can be extended to higher angles, covering the **s** range $\sim$0.5–2.5 Å^{-1}, to perform a more detailed analysis of protein structure. While this medium- to wide-angle X-ray scattering (WAXS) can readily be applied to protein samples with a high degree of regularity, *e.g.* spider silk fibres (Riekel & Vollrath, 2001), protein solutions give considerably weaker WAXS signals; indeed, the **s** range 0.5–1.0 Å^{-1} is generally regarded as the wide-angle range for protein solutions. Nonetheless, it has been demonstrated that meaningful information on protein architecture can be obtained from protein solutions (Hirai *et al.*, 2002).

In combination, therefore, these techniques, SRCD and SAXS/WAXS, may be used synergistically to study the conformation of proteins in solution, ranging from the secondary, tertiary and to quaternary structural level. Their combined use could allow for studies of individual proteins or higher-order complexes where other techniques may be either of limited use or not applicable. Additionally, they are well suited to investigate conformational changes that may be triggered by, for example, complex formation. We have chosen a complex of two peroxisomal proteins, Pex5p and SCP2, to test the synergistic potential of these methods. While the binding domains of these two proteins are structurally well characterized, little is known about the structure of their complex, which is essential to understand their function.

  doi:10.1107/S090904950402504X

The first protein, Pex5p, is a cytosolic receptor for the majority of proteins destined for the peroxisome (Dodt *et al.*, 1995). Towards the C-terminus, Pex5p contains an array of tetratricopeptide repeats (TPR) with known three-dimensional structure (Gatto *et al.*, 2000; Kumar *et al.*, 2001). The TPR domain of Pex5p specifically recognizes the peroxisomal targeting signal type 1 (PTS1), which is a C-terminal tripeptide (alanine–lysine–leucine, or a conserved variant), carried by most proteins sorted to the peroxisome (Elgersma *et al.*, 1996; Gatto *et al.*, 2003; Lametschwandtner *et al.*, 1998). Recognition of the PTS1 by the TPR domain of Pex5p is the initial step required to transport proteins to the peroxisome [peroxisomal import mechanisms are reviewed by Holroyd & Erdmann (2001) and van der Klei & Veenhuis (2002)]. In the present study, we used a human Pex5p C-terminal construct covering its TPR domain (residues 315–639), hereafter designated Pex5p(C).

The second protein, sterol carrier protein 2 (SCP2), has been used as a model PTS1 cargo protein. It exists in three forms in the cell, a bifunctional 3-oxoacyl-CoA thiolase: SCP2 protein (SCP-x), a cytosolic precursor form (preSCP2), and a processed mature peroxisomal form (mSCP2). The first two forms originate from different mRNA transcripts of the same gene, and the peroxisomal form (mSCP2) results from proteolytic cleavage of either SCP-x or preSCP2 (Stolowich *et al.*, 2002). While only little is known about structure and function of the bifunctional SCP-x form, extensive data are available for the two latter forms, preSCP2 and mSCP2 (reviewed by Stolowich *et al.*, 2002). This study focuses on these two forms only. Several NMR and crystal structures of mSCP2 (Choinowski *et al.*, 2000; Dyer *et al.*, 2003; Garcia *et al.*, 2000) and preSCP2 (Weber *et al.*, 1998) reveal that it has a mixed α/β structure, with a hydrophobic pocket to accommodate lipid ligands (Choinowski *et al.*, 2000). Common to these structures is that both termini appear to be rather flexible. The C-terminal PTS1 motif is packed against the surface of mSCP2 and poorly exposed for TPR domain recognition. Whether and to what extent the presence of the N-terminal presequence has implications on the overall fold of SCP2 has remained controversial (Stolowich *et al.*, 2002).

We present SRCD data demonstrating that the two forms of SCP2 in a complex with Pex5p(C) contain the same overall secondary structure, with the 20 residues present in preSCP2 in a coil conformation. This presequence of preSCP2 seems to increase the polydispersity of the complex such that scattering experiments are of limited value. However, the complex of mSCP2 with Pex5p(C) can be prepared at high concentration and monodispersed, allowing acquisition of useful X-ray scattering data. A low-resolution envelope structure of the Pex5p(C)/mSCP2 complex is proposed and compared with known crystal structures of its single components (Gatto *et al.*, 2000; Choinowski *et al.*, 2000). Our SRCD, SAXS and WAXS data indicate changes in its tertiary structure while only little alteration could be detected at the level of the secondary structure content of either protein upon complex formation.

3. Materials and methods

3.1. Materials

Chemicals were obtained from Sigma-Aldrich at the highest available purity. Details of the preparation of preSCP2, mSCP2 and Pex5p(C) proteins will be described in detail elsewhere (Stanley *et al.*, unpublished data). Briefly, recombinant proteins were expressed in *E. coli* and purified to >99% homogeneity by affinity and size-exclusion chromatography. Purity was verified by gel electrophoresis and mass spectrometry. Complexes were prepared by mixing Pex5p(C) with excess SCP2 followed by size-exclusion chromatography to separate uncomplexed SCP2.

3.2. SRCD measurements

Protein complexes were exhaustively dialysed against 10 m*M* potassium phosphate (pH 7.4) and subsequently diluted to 1 mg ml^{-1}. Protein concentrations were determined throughout this study by $A_{280\,\mathrm{nm}}$ of protein diluted with 8 *M* urea. An extinction coefficient of 42 530 M^{-1} cm^{-1} was calculated using the method of Gill & von Hippel (1989) for both Pex5p(C)/preSCP2 and Pex5p(C)/mSCP2 assuming a (1:1) stoichiometry.

SRCD spectra were obtained on station CD12 of the CCLRC Daresbury Laboratory's Synchrotron Radiation Source (Clarke & Jones, 2004). Prior to measuring protein spectra, a (+)-10-camphosulphonic acid (CSA) spectrum was measured for ellipticity calibration (Woody, 1995). CSA was used at 10 mg ml^{-1} in a 0.1 mm quartz cuvette (Hellma).

Approximately 30 µl protein at a concentration of 1 mg ml^{-1} was used to fill a 0.1 mm-pathlength quartz cuvette. The sample chamber was purged with dry nitrogen and the cuvette maintained at 298 K. Complete spectra were obtained in rapid single-scan exposures of 3 min each to minimize the effects from possible radiation damage owing to the high photon flux at SRS beamline CD12 (Clarke & Jones, 2004). Spectra were measured over the range 260–168 nm, with 0.5 nm intervals and 1 s integration time per interval. Spectra were corrected for buffer background and scaled against the CSA ellipticity calibration. The high spectral quality did not call for data smoothing; however, standard deviations from sets of five separate measurements for each complex were determined.

3.3. SRCD data analysis

A modified version of the program *SELCON* (Sreerama & Woody, 1993; Clarke & Jones, 1999) was used to analyse SRCD spectra for protein secondary structure content. *SELCON* uses the singular value decomposition algorithm to assign secondary structure by comparison with a basis set of spectra from proteins of known structure, repeated iteratively to self-consistency. The estimated secondary structure content was compared with that of known crystal structures available from the Protein Data Bank (PDB; Berman *et al.*, 2000) of Pex5p(C) (PDB code 1FCH, chain A) and mSCP2 (PDB code 1C44). The programs *PROMOTIF* (Hutchinson & Thornton, 1996) and *XTLSSTR* (King & Johnson, 1999) were used to evaluate the secondary structure content of the crystal structures.

3.4. SAXS and WAXS measurements

Protein complexes were exhaustively dialysed against 10 m*M* potassium phosphate (pH 7.4). 1 m*M* of freshly prepared 1,4-dithio-DL-threitol (DTT) was added immediately prior to measurement. Approximately 120 µl samples were used to fill a 1 mm mica cuvette. SAXS/WAXS curves from protein complexes were obtained at concentrations of 8, 13 and 16 mg ml^{-1} with intermittent buffer background measurements. Forward-scattering calibrations were conducted with 7 mg ml^{-1} bovine serum albumin (BSA) in 50 m*M* sodium-HEPES (pH 7.5) adding 1 m*M* of fresh DTT immediately prior to measurement. All measurements were conducted at 298 K.

Measurements were carried out on beamline X33 at EMBL/DESY, Hamburg, Germany (http://www.embl-hamburg.de/ExternalInfo/Research/Sax/). SAXS patterns for both complexes were recorded with a linear delay line readout proportional gas chamber (Boulin *et al.*, 1988). The sample-to-detector distance was 1.8 m, thus covering the momentum-transfer vector range 0.018 Å^{-1} < **s** < 0.45 Å^{-1}. For

the Pex5p(C)/mSCP2 complex, WAXS patterns were recorded on a second detector at 0.9 m covering the momentum transfer vector range 0.25 $Å^{-1} < \mathbf{s} < 0.9$ $Å^{-1}$. Data were collected over several 1 min frames to monitor for radiation damage.

Data were normalized to incident-beam intensity, corrected for detector response, buffer background subtracted, scaled to protein concentration and extrapolated to zero concentration following standard procedures. The data collected over the two scattering vector ranges for Pex5p(C)/mSCP2 were merged to yield the final composite scattering pattern. All data-processing steps were conducted using the program *PRIMUS* (Konarev *et al.*, 2003).

The maximum particle dimension d_{max} was estimated using the orthogonal expansion program *ORTOGNOM* (Svergun, 1993). Porod analysis (Porod, 1982) was used to estimate the excluded particle volume. The radius of gyration R_g was evaluated using the Guinier approximation (Koch *et al.*, 2003) and with the indirect transform package *GNOM* (Svergun, 1992). *GNOM* was also used to provide the distance distribution function $p(r)$ of the particles. Protein molecular mass was calculated by comparison of the forward-scattering intensity with the reference BSA sample.

3.5. Model building

A low-resolution model of the Pex5p(C)/mSCP2 complex was built from the X-ray scattering data *ab initio* with *GASBOR* (Svergun *et al.*, 2001; Petoukhov & Svergun, 2003), either using reciprocal-space data (version 18) or using a real-space algorithm (version 20). The most probable model was obtained by averaging the reciprocal-space and real-space models using the program *DAMAVER* (Volkov & Svergun, 2003). Chain-compatible dummy residue (DR) models restored by *GASBOR* from SAXS and WAXS data were compared with models calculated using *DAMMIN* (Svergun, 1999), which restores a densely packed (*i.e.* not chain-compatible) dummy atom (DA) model using only SAXS data.

CRYSOL (Svergun *et al.*, 1995) was used to calculate scattering patterns from the known crystal structures of Pex5p(C) (Gatto *et al.*, 2000) and mSCP2 (Choinowski *et al.*, 2000). Subsequently, the heterodimer modelling function of *MASSHA* (Konarev *et al.*, 2001) was used to perform rigid-body modelling of the mSCP2 protein positioned in close proximity to the Pex5p(C) domain. Initially, the C-terminus of mSCP2 was positioned manually to fit the location of a minimal PTS1 peptide in the Pex5p(C) crystal structure, as a starting point for rigid-body refinement. Thus, this modelling procedure was dependent only on prior knowledge of the crystal structures and not on the *ab initio* model. Translations and rotations of the mSCP2 rigid body were tested to find the orientation of the two proteins with the best fit to the scattering data from the Pex5p(C)/mSCP2 complex.

4. Results and discussion

4.1. Determination of secondary structural content by SRCD

SRCD spectra of Pex5p(C)/mSCP2 and Pex5p(C)/preSCP2 were measured to gain insight into their secondary structural content (Fig. 1). The spectra are strongly characteristic of predominantly α-helical protein, with spectral bands attributable to electron transitions in the amide groups of the protein backbone. The carbonyl oxygen lone-pair rotational transition nπ^* gives rise to the minimum at ~222 nm, indicative of a right-handed α-helix. The features at ~208 nm and ~190 nm are indicative of exciton splitting of the $\pi\pi^*$ absorption band; the minimum at ~208 nm results from $\pi\pi^*$ polarized parallel to the helical axis, while the maximum at ~190 nm results from $\pi\pi^*$ polarized perpendicular to the helical axis (Woody, 1996; Wallace, 2000). Further bands can be seen in the VUV part of the spectrum: the shoulder at ~175 nm results from the nσ^* transition of the carbonyl oxygen lone-pair and the negative trend below ~170 nm implies another transition, characteristic of α-helical proteins, with minimum ~165 nm (Wallace, 2000; Wallace *et al.*, 2004). The local minimum at ~222 nm is somewhat deeper than the minimum at ~208 nm. This is indicative of the β-structure present in the protein–protein complexes. The nπ^* transition in β-strands gives a negative band at ~215 nm. A similar feature is observed from type-II β-turns, but red-shifted by 5–10 nm (Woody, 1996).

Two additional features of these spectra are of note. Firstly, the minor fluctuation near 260 nm appears to be an instrumental artefact or a contribution to the CD signal from aromatic amino acids (Woody & Dunker, 1996); therefore, data above 255 nm have been omitted from further analysis. Secondly, the reduced spectral quality below ~180 nm is expected to be due to the high absorbance of light at these wavelengths by the aqueous buffer in the 0.1 mm-pathlength cell used. However, the signal-to-noise ratio of the data was more than adequate for secondary structure analysis to 168 nm.

From our SRCD measurements we estimated an overall amount of regular secondary structure of about 82 ± 6.6% and 78 ± 8.8% for the Pex5p(C)/mSCP2 and Pex5p(C)/preSCP2 complexes, respectively (Fig. 1, Table 1). The reduced α-helical content in the Pex5p(C)/preSCP2 complex suggests that the additional 20 N-terminal residues in preSCP2 do not adopt a regular secondary structure, in accordance with previous NMR data from preSCP2 (Weber *et al.*, 1998). Further, the distribution of secondary structure classes estimated from the SRCD measurements of the Pex5p(C)/mSCP2 complex matches that of the combined values of the independent Pex5p(C) and mSCP2 crystal structures. These results suggest that little alteration in secondary structure appears to be required for the two proteins to form a complex.

4.2. Shape and architecture of the Pex5p(C)/mSCP2 complex by SAXS/WAXS

A composite SAXS/WAXS pattern has been obtained from the Pex5p(C)/mSCP2 complex up to a resolution of about 7 Å (Fig. 2*a*, Table 2). Fig. 2(*b*) shows a comparison of Guinier plots for the two complexes: the Pex5p(C)/mSCP2 complex shows a linear relationship, indicative of a monodispersed population of scattering particles (Koch *et al.*, 2003), while the plot for Pex5p(C)/preSCP2 is non-linear, indicative of polydisperse scatterers. All further SAXS/WAXS analysis, therefore, was conducted using Pex5p(C)/mSCP2. The forward-scattering intensity compared with BSA calibration indicates

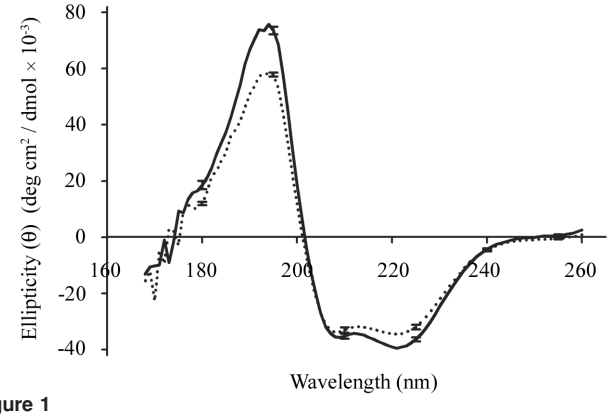

Figure 1
SRCD spectra of Pex5p(C)/mSCP2 (solid line) and Pex5p(C)/preSCP2 complexes (dotted line). Standard deviations were calculated from five independent measurements of each sample.

Table 1
Secondary structural content of the Pex5p(C)/mSCP2 and Pex5p(C)/preSCP2 complexes evaluated from SRCD spectra using *SELCON*, with standard deviations in parentheses (Sreerama & Woody, 1993; Clarke & Jones, 1999).

For comparison, the secondary structural content from the crystal structures of mSCP2 (PDB code 1C44) and Pex5p(C) (PDB code 1FCH, chain A), and the expected secondary structure content for a complex of the two assuming there is no alteration upon complex formation, calculated by *XTLSSTR* (King & Johnson, 1999), is shown. An alternative calculation for the complex is provided by *PROMOTIF* (Hutchinson & Thornton, 1996).

	Secondary structure %						
	α-Helix	β-Sheet	3_{10}-Helix	Turn	Polyproline type-II helix	Other	Total
Pex5p(C)/mSCP2 (*SELCON*)	56.9 (2.0)	3.7 (1.9)	9.8 (1.0)	9.7 (0.9)	1.7 (0.4)	18.2 (4.0)	100
Pex5p(C)/preSCP2 (*SELCON*)	47.4 (1.7)	8.8 (3.4)	8.5 (0.8)	9.6 (1.7)	3.1 (1.2)	22.6 (4.0)	100
Pex5p(C) (*XTLSSTR*)	62.7	0.9	5.8	6.9	2.5	21.2	100
mSCP2 (*XTLSSTR*)	34.6	18.7	1.2	9.0	2.4	34.1	100
Pex5p(C)/mSCP2 (*XTLSSTR*)	57.4	6.1	4.7	7.8	2.6	21.4	100
Pex5p(C)/mSCP2 (*PROMOTIF*)	57.8	6.5	2.0	nd	nd	33.7	100

Table 2
Summary of SAXS and WAXS analysis of the Pex5p(C)/mSCP2 complex.

Theoretical molecular mass (kDa)	49.6
Molecular mass from forward scattering (kDa)	54.0
Radius of gyration, R_g, Guinier analysis (Å)	28.70 ± 0.5
Radius of gyration, R_g, *GNOM* (Å)	28.70 ± 0.2
Maximum particle dimension, d_{max} (Å)	90
Particle volume from Porod analysis (Å^3)	80000 ± 10000
GASBOR (SAXS/WAXS merged)	
Particle excluded volume (Å^3)	77180
Model fit, χ against raw data (real space)	0.7637
Model fit, χ against raw data (reciprocal space)	0.8571
Total number of dummy residues	457
Radius of gyration, R_g, of model (Å)	28.70
DAMMIN (SAXS)	
Particle excluded volume (Å^3)	87730
Model fit, χ against raw data	0.6145
Total number of dummy atoms	473
Radius of gyration, R_g, of model (Å)	28.54
CRYSOL/*MASSHA* rigid-body modelling	
Model fit, χ against raw data	2.09

a particle of about 54 kDa, in close agreement with a heterodimeric Pex5p(C)/mSCP2 complex, which has a calculated mass of 49.6 kDa. The particle has R_g = 28.7 Å and d_{max} = 90 Å, implying an elongated shape.

Porod analysis (Porod, 1982) indicated a particle volume of 80000 ± 10000 Å^3. The excluded volumes for dehydrated Pex5p(C) and mSCP2, as calculated from the available X-ray coordinates using the program *CRYSOL* (Svergun *et al.*, 1995), are 41210 Å^3 for Pex5p(C) and mSCP2 occupies 17650 Å^3, making a total of 58860 Å^3 for the heterodimeric Pex5p(C)/mSCP2 complex. The Porod volume (Porod, 1982) is somewhat larger owing to the hydration shell of the complex in solution. It should also be noted that in the Pex5p(C) crystal structure (Gatto *et al.*, 2000) a total of 28 residues remained invisible, which may further account for the discrepancy in measured and calculated particle volume. In addition, such comparison does not account for possible conformational alterations of the protein components upon complex formation.

Comparison of the scattering patterns shown in Fig. 2 with those of Hirai *et al.* (2002) suggest a predominantly helical protein architecture.

Two complementary approaches were taken to model the solution structure of the Pex5p(C)/mSCP2 complex. Firstly, *ab initio* models were built using *DAMMIN* (Svergun, 1999) and *GASBOR* (Svergun *et al.*, 2001; Petoukhov & Svergun, 2003), which use simulated annealing to build dummy atom (DA) and dummy residues (DR) models, respectively, fitting the X-ray scattering data. The second approach was to model the complex by using rigid-body modelling using the software *MASSHA* (Konarev *et al.*, 2001). In this approach we kept the coordinates of the Pex5p(C) crystal structure fixed while moving the mSCP2 crystal structure to find the best fit to the experimental data.

Parameters describing the *ab initio* DA and DR models are summarized in Table 2. The excluded volume of the DA model is 87700 Å^3 as a total of 473 DAs have been incorporated and R_g = 28.54 Å, values in good agreement with the experimental data. The chain-compatible DR model is generated using a fixed number of 457 DRs, defining an excluded volume of 77200 Å^3. The radius of gyration is derived from the distance distribution function (28.70 Å) and the model is built into the defined search volume. Fig. 2(*a*) shows *GASBOR* and *DAMMIN* fits to the experimental data in reciprocal space and Fig. 2(*b*) shows the distance distribution function for the Pex5p(C)/mSCP2 complex calculated by *GNOM* (Svergun, 1992) with a representative fit to this distribution by *GASBOR* (version 20). Several models restored by *GASBOR* were averaged using *DAMAVER* (Volkov & Svergun, 2003). A model representing the most populated volume from the averaging procedure is shown in Fig. 3. The particle is elongated and 'shoe'-shaped, apparently consisting of two globular domains oriented at ∼140° with respect to each other. The overall dimensions of the particle are about 30 Å × 40 Å × 90 Å. The two domains are of different sizes, the smaller occupying about one-third of the particle volume. It is tempting to speculate that the smaller domain represents mSCP2 and the larger one represents Pex5p(C).

An alternative model has been calculated using the second approach by fitting the crystal structures of the two protein components into the available SAXS/WAXS data (Fig. 3). In this model, the PTS1 motif of mSCP2 is rather distant from the PTS1 binding groove in Pex5p(C), as defined in the X-ray structure of the Pex5p(C)/PTS1 peptide complex (Gatto *et al.*, 2000). Graphical inspection indicated that it is not possible to move the PTS1 motif of mSCP2 into the PTS1 peptide binding site of Pex5p(C) when treating Pex5p(C) and mSCP2 as rigid bodies. Therefore, the SAXS/WAXS model suggests an outward movement of the C-terminus of mSCP2, comprising the PTS1 motif, to penetrate the binding groove. This hypothesis is

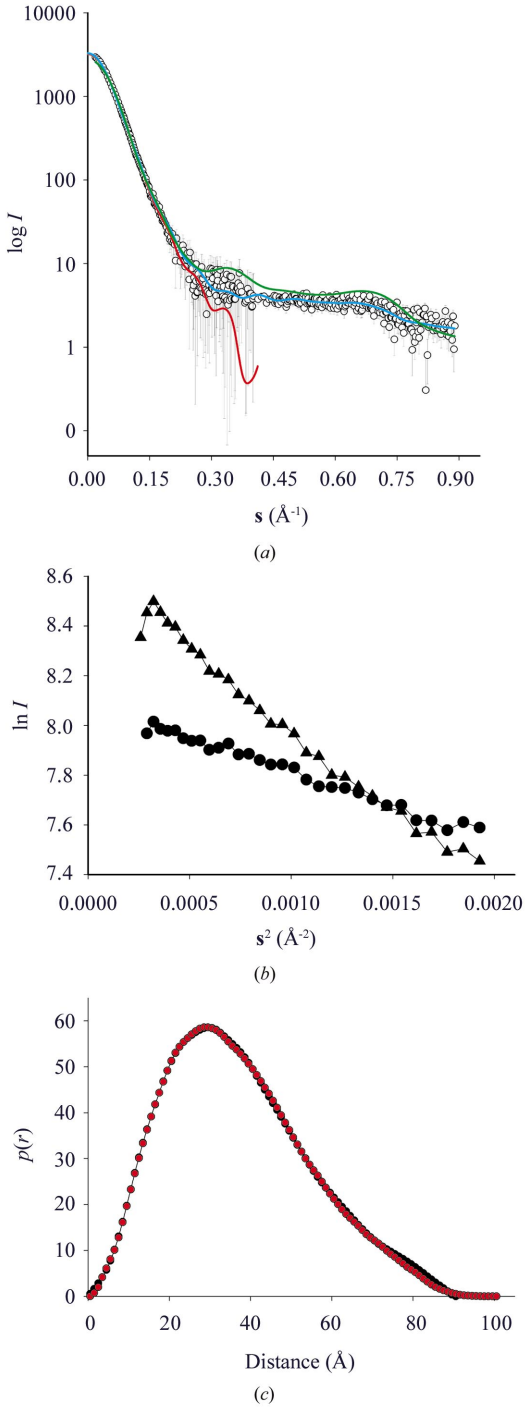

Figure 2
(*a*) Composite SAXS/WAXS pattern of the Pex5p(C)/mSCP2 complex. SAXS data (0.018 Å^{-1}< **s** < 0.45 Å^{-1}) and WAXS data (0.25 Å^{-1} < **s** < 0.90 Å^{-1}) have been merged. The protein sample was at 13 mg ml^{-1} in 10 m*M* potassium phosphate, pH 7.4, and 1 m*M* DTT. Raw scattering data are shown as open circles, with error bars. Data fits for *ab initio* models restored by *DAMMIN* (Svergun, 1999) (red line, final χ value against the raw data = 0.6145) and *GASBOR* (version 18) (Svergun *et al.*, 2001) (blue line, $\chi = 0.8571$) are shown. The rigid-body model fit from *MASSHA* is shown as a green line, $\chi = 2.09$. (*b*) Guinier plots for Pex5p(C)/mSCP2 (circles) and Pex5p(C)/preSCP2 (triangles). Only the Pex5p(C)/mSCP complex shows a linear $\mathbf{s}^{-2}/\ln I$ relationship, indicative of a monodispersed population of scattering particles (Koch *et al.*, 2003). (*c*) Real-space fit to $p(r)$ using *GASBOR* (version 20) (Petoukhov & Svergun, 2003). $p(r)$ output from *GNOM* (Svergun, 1992) is shown in black and the model fit in red. d_{max} is 90 Å. The final χ value against the raw data is 0.7637.

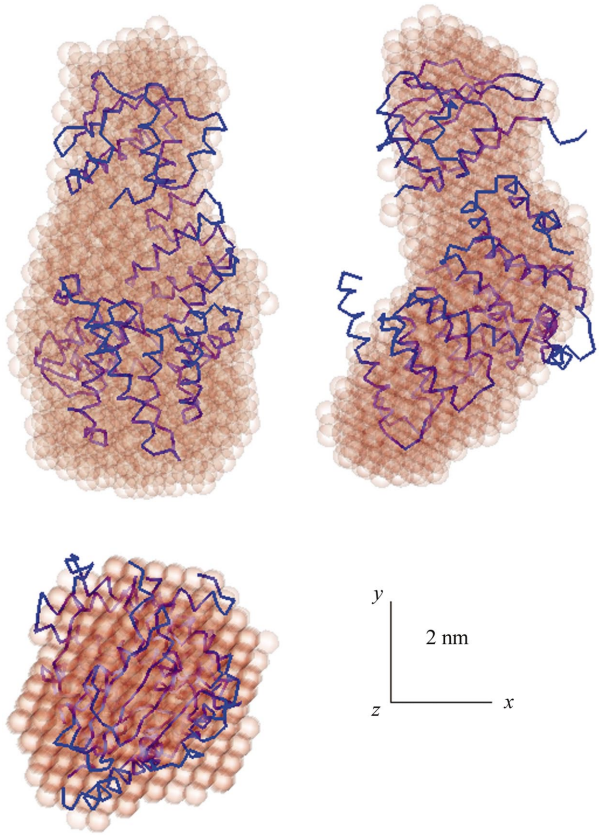

Figure 3
Superimposition of the rigid-body model (shown in blue) generated using *MASSHA* (Konarev *et al.*, 2001) on the 'shoe'-shaped *ab initio* envelope structure (shown in orange) of the Pex5p(C)/mSCP2 complex derived from merging four envelopes restored by *GASBOR* (version 18) and six restored by *GASBOR* (version 20) (Petoukhov & Svergun, 2003; Svergun *et al.*, 2001). Three views of the envelope are shown, rotated by 90° respective to each other. The image was generated using *ASSA* (Kozin *et al.*, 1997).

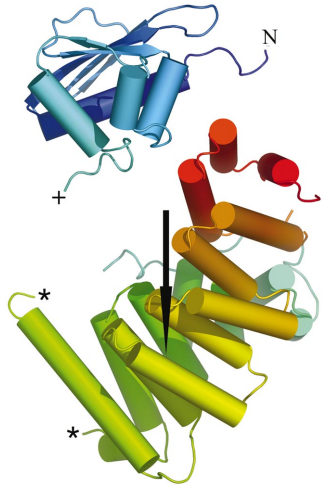

Figure 4
Rigid-body model of the Pex5p(C)/mSCP2 complex. Rigid-body modelling using *MASSHA* (Konarev *et al.*, 2001) was used to determine the orientation of the two proteins that best fits the scattering curve. mSCP2 (Choinowski *et al.*, 2000) is shown in blue, the C-terminal PTS1 indicated with a (+) and the N-terminus with an (N). The three-helix bundle at the C-terminus of Pex5p(C) is shown in red and the linker region from which Pex5p(C) residues are missing (unresolved in the crystal structure; Gatto *et al.*, 2000) is indicated by asterisks. The arrow shows the approximate position where the PTS1 peptide resides within Pex5p(C). See text for details.

supported by the crystal structure of mSCP2, in which the ten C-terminal residues are not in a regular secondary structural conformation. Furthermore, the high-temperature factors associated with these residues indicate that they are more mobile than other parts of the mSCP2 structure (Choinowski *et al.*, 2000). Given that the SRCD data do not indicate a significant alteration of secondary structure distribution upon complex formation, we assume that the C-terminus of mSCP2 remains void of regular secondary structure. In addition, the SAXS/WAXS model suggests that the C-terminal three-helix bundle of Pex5p(C) is in direct contact with the surface helices of mSCP2. Another implication of the model is that the N-terminus of mSCP2, to which the presequence attaches in preSCP2, points away from the TPR domain and does not interfere with binding to Pex5p(C). Indeed, we have not detected a major change in binding affinities when using preSCP2 instead of mSCP2 (Stanley *et al.*, unpublished).

The *ab initio* model and rigid-body model are consistent with each other: both display the same 'shoe' shape, with two distinct globular domains obtusely angled with respect to each other. A key difference, however, emerges from the missing flexible inter-TPR linker in the rigid-body model (Fig. 4). Its proposed location in the rigid-body model suggests that this residue segment could contribute to the elongated appearance of the Pex5p(C)/mSCP2 complex in solution. According to our SAXS/WAXS model, binding of Pex5p(C) and mSCP2 requires additional 'accessory' epitopes on mSCP2. From the rigid-body model, it seems plausible that these mSCP2 accessory epitopes may be topographically remote with respect to the PTS1 motif. Further characterization of such accessory motifs will allow for improved prediction of peroxisomal localization. Physiologically, they may provide an additional sorting mechanism, allowing the Pex5p(C) to select for correctly folded proteins and not simply for the presence of a PTS1. However, in the absence of a crystal structure of a complete Pex5p(C)/mSCP2, their precise topology remains elusive.

4.3. Perspectives for the synergistic use of SRCD and SAXS/WAXS in structural biology and structural proteomics

While the expected complete repertoire of protein folds is already well represented in the PDB (Berman *et al.*, 2000; Thornton *et al.*, 2000), at least for soluble proteins, high-resolution structures of protein–protein complexes are still scarce. Although predictive algorithms for the structural basis of protein–protein interactions are presently under stringent development (Aloy *et al.*, 2004; Eisenberg *et al.*, 2000), there is a requirement for an overwhelming number of experimentally determined high-resolution structures from protein–protein complexes to unravel the precise basis of their molecular interactions. Since protein–protein complexes naturally tend to be larger and more complex than their single components, the experimental demands for their preparation for structural investigation increases and the choice of suitable high-resolution methods is more limited. Hence, we expect that many of them will be initially determined using molecular structural biology methods, such as SAXS/WAXS or electron microscopy, aiming to elucidate their low-resolution shapes (Sali *et al.*, 2003; Svergun & Koch, 2002). Since these methods do not provide information about their secondary structural content and distribution of different types of secondary structure *per se*, SRCD provides an attractive option to provide this type of data. Our own data support previous observations which consider the use of synchrotron radiation to be essential for this type of interpretation (Wallace & Janes, 2001).

Using SRCD and SAXS/WAXS, we have been able to provide insight into some of the structural requirements on Pex5p(C)/mSCP2 complex formation. We have been able to fit the suggested 'shoe' shape of this complex to previous biophysical data. However, our data also reveal the requirement of a considerable conformational change of the C-terminus of mSCP2 to be capable to interact with Pex5p(C) *via* the binding groove that had been previously characterized (Gatto *et al.*, 2000). Further, we have observed that one isoform, mSCP2, can be integrated into a more soluble and ordered complex with Pex5p(C) than preSCP2, without significantly altering the overall nature of the complex. The combined use of SRCD and SAXS/WAXS, as demonstrated in this study of the Pex5p(C)/mSCP2 complex, may serve as a paradigm to provide useful data both at the secondary, tertiary and quaternary structural level of protein–protein complexes, preceding and in some cases complementing available high-resolution data.

5. Conclusions

We have synergistically used SRCD, SAXS and WAXS to study the structure of a complex of two proteins, Pex5p(C) and SCP2, in solution. We have shown that the complex formation does not require major rearrangements in secondary structure but that SCP2 apparently undergoes a change in tertiary structure to facilitate the binding mode. Further, our data suggest a previously uncharacterized extensive interface between the two proteins, which may have important functional implications. The combined use of SRCD and SAXS/WAXS, as demonstrated in this study of the Pex5p(C)/mSCP2 complex, may serve as a paradigm to provide useful data both at the secondary, tertiary and quaternary structural level of protein–protein complexes, preceding and in some cases complementing available high-resolution data.

This work was supported by grants BIO-CT97-2180 and CT-2001-01663 by the European Commission to MW. Access to beamline CD12 at CCLRC Daresbury Laboratory's Synchrotron Radiation Source was supported by the European Commission's Framework 5 IHP Large Scale Facility Programme, 'Access to Research Infrastructures Action for Improving Human Potential Programme' (grant number 39163). We thank Marc Niebuhr and Michel Koch for helpful discussions.

References

Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A. C., Bork, P., Superti-Furga, G., Serrano, L. & Russell, R. B. (2004). *Science*, **303**, 2026–2029.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucl. Acids Res.* **28**, 235–242.

Boulin, C. J., Kempf, R., Gabriel, A. & Koch, M. H. J. (1988). *Nucl. Instrum. Methods A*, **269**, 312–320.

Choinowski, T., Hauser, H. & Piontek, K. (2000). *Biochemistry*, **39**, 1897–1902.

Clarke, D. T. & Jones, G. R. (1999). *Biochemistry*, **38**, 10457–10462.

Clarke, D. T. & Jones, G. R. (2004). *J. Synchrotron Rad.* **11**, 142–149.

Dodt, G., Braverman, N., Wong, C., Moser, A., Moser, H. W., Watkins, P., Valle, D. & Gould, S. J. (1995). *Nature Genet.* **9**, 115–125.

Dyer, D. H., Lovell, S., Thoden, J. B., Holden, H. M., Rayment, I. & Lan, Q. (2003). *J. Biol. Chem.* **278**, 39085–39091.

Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. (2000). *Nature (London)*, **405**, 823–826.

Elgersma, Y., Vos, A., Van Den Berg, M., Van Roermund, C. W., Van Der Sluijs, P., Distel, B. & Tabak, H. F. (1996). *J. Biol. Chem.* **271**, 26375–26382.

Filgueira De Azevedo, W. Jr., Dos Santos, G. C., Dos Santos, D. M., Olivieri, J. R., Canduri, F., Silva, R. G., Basso, L. A., Renard, G., Da Fonseca, I. O., Mendes, M. A., Palma, M. S. & Santos, D. S. (2003). *Biochem. Biophys. Res. Commun.* **309**, 923–928.

Garcia, F. L., Szyperski, T., Dyer, J. H., Choinowski, T., Seedorf, U., Hauser, H. & Wuthrich, K. (2000). *J. Mol. Biol.* **295**, 595–603.
Gatto, G. J. Jr., Geisbrecht, B. V., Gould, S. J. & Berg, J. M. (2000). *Nature Struct. Biol.* **7**, 1091–1095.
Gatto, G. J. Jr., Maynard, E. L., Guerrerio, A. L., Geisbrecht, B. V., Gould, S. J. & Berg, J. M. (2003). *Biochemistry*, **42**, 1660–1666.
Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
Hirai, M., Iwase, H., Hayakawa, T., Miura, K. & Inoue, K. (2002). *J Synchrotron Rad.* **9**, 202–205.
Holroyd, C. & Erdmann, R. (2001). *FEBS Lett.* **501**, 6–10.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
King, S. M. & Johnson, W. C. (1999). *Proteins*, **35**, 313–320.
Koch, M. H. J., Vachette, P. & Svergun, D. I. (2003). *Q. Rev. Biophys.* **36**, 147–227.
Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.
Konarev, P. V., Volkov, V. V., Sokolova, A., Koch, M. H. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Kozin, M. B., Volkov, V. V. & Svergun, D. I. (1997). *J. Appl. Cryst.* **30**, 811–815.
Kumar, A., Roach, C., Hirsh, I. S., Turley, S., Dewalque, S., Michels, P. A. & Hol, W. G. (2001). *J. Mol. Biol.* **307**, 271–282.
Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J. & Hartig, A. (1998). *J. Biol. Chem.* **273**, 33635–33643.
Marquez, J. A., Smith, C. I., Petoukhov, M. V., Lo Surdo, P., Mattsson, P. T., Knekt, M., Westlund, A., Scheffzek, K., Saraste, M. & Svergun, D. I. (2003). *EMBO J.* **22**, 4616–4624.
Petoukhov, M. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 540–544.
Porod, G. (1982). *General Theory.* London: Academic Press.
Riekel, C. & Vollrath, F. (2001). *Int. J. Biol. Macromol.* **29**, 203–210.
Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. (2003). *Nature (London)*, **422**, 216–225.
Sokolova, A., Malfois, M., Caldentey, J., Svergun, D. I., Koch, M. H., Bamford, D. H. & Tuma, R. (2001). *J. Biol. Chem.* **276**, 46187–46195.
Sreerama, N. & Woody, R. W. (1993). *Anal. Biochem.* **209**, 32–44.
Stolowich, N. J., Petrescu, A. D., Huang, H., Martin, G. G., Scott, A. I. & Schroeder, F. (2002). *Cell Mol. Life Sci.* **59**, 193–212.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D. I. (1993). *J. Appl. Cryst.* **26**, 258–267.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I. & Koch, M. H. (2002). *Curr. Opin. Struct. Biol.* **12**, 654–660.
Svergun, D. I. & Nierhaus, K. H. (2000). *J. Biol. Chem.* **275**, 14432–14439.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). *Biophys. J.* **80**, 2946–2953.
Thornton, J. M., Todd, A. E., Milburn, D., Borkakoti, N. & Orengo, C. A. (2000). *Nature Struct. Biol. (Suppl.),* **7**, 991–994.
Van Der Klei, I. & Veenhuis, M. (2002). *Curr. Opin. Cell Biol.* **14**, 500–505.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Wallace, B. A. (2000). *J. Synchrotron Rad.* **7**, 289–295.
Wallace, B. A. & Janes, R. W. (2001). *Curr. Opin. Chem. Biol.* **5**, 567–571.
Wallace, B. A., Lees, J. G., Orry, A. J., Lobley, A. & Janes, R. W. (2003). *Protein Sci.* **12**, 875–884.
Wallace, B. A., Wien, F., Miles, A. J., Lees, J. G., Hoffmann, S. V., Evans, P., Wistow, G. J. & Slingsby, C. (2004). *Faraday Discuss.* **126**, 237–243; discussion 245–254.
Weber, F. E., Dyer, J. H., Garcia, F. L., Werder, M., Szyperski, T. & Wuthrich, K. (1998). *Cell. Mol. Life Sci.* **54**, 751–759.
Woody, R. W. (1995). *Methods Enzymol.* **246**, 34–71.
Woody, R. W. (1996). *Theory of Circular Dichroism of Proteins*, in *Circular Dichroism and the Conformational Analysis of Biomolecules*, edited by G. D. Fasman, pp. 25–67. New York: Plenum.
Woody, R. W. & Dunker, A. K. (1996). *Aromatic and Cystine Side-Chain Circular Dichroism in Proteins*, in *Circular Dichroism and the Conformational Analysis of Biomolecules*, edited by G. D. Fasman, pp. 109–157. New York: Plenum.