BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment

Nicholas Skar-Gislinge,[a]*‡ Søren A. R. Kynde,[a]* Ilia G. Denisov,[b] Xin Ye,[b] Ivan Lenov,[b] Stephen G. Sligar[b] and Lise Arleth[a]*

[a]Structural Biophysics, Niels Bohr Institute, Faculty of Science, University of Copenhagen, Copenhagen, Denmark, and [b]Department of Biochemistry, School of Molecular and Cellular Biology, University of Illinois, 505 South Goodwin Avenue, Urbana, IL 61801, USA. *Correspondence e-mail: skar-gislinge@nbi.ku.dk, kynde@nbi.ku.dk, arleth@nbi.ku.dk

Received 27 March 2015
Accepted 5 October 2015

Edited by P. Langan, Oak Ridge National Laboratory, USA

‡ Current address: Department of Physical Chemistry, Lund University, Lund, Sweden.

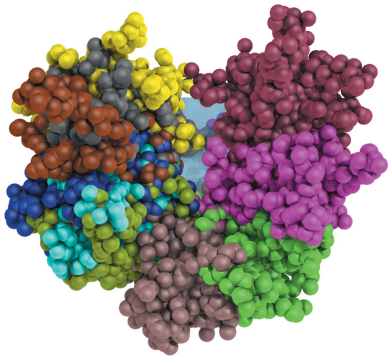
Keywords: small-angle scattering; cytochrome P450 3A4; nanodiscs; membrane proteins.

Membrane proteins reconstituted into phospholipid nanodiscs comprise a soluble entity accessible to solution small-angle X-ray scattering (SAXS) studies. It is demonstrated that using SAXS data it is possible to determine both the shape and localization of the membrane protein cytochrome P450 3A4 (CYP3A4) while it is embedded in the phospholipid bilayer of a nanodisc. In order to accomplish this, a hybrid approach to analysis of small-angle scattering data was developed which combines an analytical approach to describe the multi-contrast nanodisc with a free-form bead-model description of the embedded protein. The protein shape is then reconstructed *ab initio* to optimally fit the data. The result of using this approach is compared with the result obtained using a rigid-body description of the CYP3A4-in-nanodisc system. Here, the CYP3A4 structure relies on detailed information from crystallographic and molecular-dynamics studies of CYP3A4. Both modelling approaches arrive at very similar solutions in which the α-helical anchor of the CYP3A4 systematically stays close to the edge of the nanodisc and with the large catalytic domain leaning over the outer edge of the nanodisc. The obtained distance between the globular domains of CYP3A4 is consistent with previously published theoretical calculations.

1. Introduction

Membrane proteins constitute 26% of the proteins in the human proteome (Fagerberg *et al.*, 2010). They perform many critical functions, such as cell signalling, transport, photosynthesis, biosynthesis and metabolism. However, owing to the difficulty in handling them, high-resolution structural studies of membrane proteins are challenging. Only 1% of the protein structures resolved by X-ray crystallography or NMR and deposited in the Protein Data Bank represent membrane proteins, and almost all of them have been obtained using proteins solubilized in detergent-based micellar systems. Recently, several structures of GPCRs and other types of integral membrane proteins have been resolved based on crystallization in lipid cubic phases (Cherezov, 2011; Rosenbaum *et al.*, 2011). This was a remarkable success, which was honoured by the Nobel Prize in 2012. However, the general problem of structural characterization of a membrane protein inserted into the native membrane, including detailed information on the geometry and mode of protein incorporation into the lipid bilayer and functionally important protein–lipid interactions, is still unresolved.

Cytochromes P450 belong to the broad superfamily of haem enzymes (more than 20 000 isozymes identified in genomes of organisms from all biological kingdoms; Nelson, 2009), which



catalyse multiple reactions that are essentially important for biosynthesis as well as for the catabolism of xenobiotics (Ortiz de Montellano, 2005). Cytochromes P450 exist in all eukaryotic organisms and function as monotopic membrane proteins, with the N-terminal 30–50 residues inserted into the membrane. In humans, 56 different cytochromes P450 are involved in the biosynthesis and regulation of steroid hormones and vitamins, and also play a major role in drug metabolism. Importantly, the catalytic activity of drug-metabolizing cytochromes P450 critically depends on their interactions with the redox partner, which is another monotopic membrane-bound flavoprotein, NADPH-dependent cytochrome P450 reductase (CPR). As a result, the membrane properties, such as lipid composition, physical state and the charge and mobility of bilayer components (Schuler *et al.*, 2013), as well as the mode of incorporation of both proteins into the membrane, are critical for the function of the dynamic complexes of cytochromes P450 with CPR (Grinkova *et al.*, 2013). The position and depth of insertion of the cytochrome P450 in the lipid bilayer define the accessibility and available binding pathways for the lipophilic substrates distributed in the membrane (Yu *et al.*, 2013). In addition, the mobility and mutual orientation of the cytochrome P450 and CPR with respect to the membrane are the key determinants for their interactions and formation of the functional protein–protein complex. Lastly, the membrane electrostatic field may perturb the midpoint potentials of both cytochrome P450 and CPR and significantly change the reduction kinetics and overall catalytic properties of the P450 unit.

Recently, the mode of incorporation of the human cytochromes P450 into the membrane has been the subject of a number of studies. Most of them used various realisations of molecular dynamics (Otyepka *et al.*, 2012; Denisov *et al.*, 2012; Baylon *et al.*, 2013) in order to evaluate the position and orientation of several isozymes in the model bilayer. Several experimental attempts were based on accessibility probes of the surface amino acids of cytochrome P450 (Mast *et al.*, 2009). However, the structural information obtained by such methods is semi-quantitative at best and usually requires specific experimental conditions for the optimal yield of chemical steps. In contrast, small-angle scattering (SAS) methods can be performed under a broad range of conditions (temperature, pH, pressure and solvent composition) with no restriction on the lipid composition of the membrane and using a relatively small amount of sample (50 µl of 3–10 mg ml^{-1} solution).

Nanodiscs are phospholipid bilayer discs of 10 nm in size (Bayburt *et al.*, 2002) in which membrane proteins can be incorporated. They consist of a central lipid bilayer of 120–150 phospholipids, depending on the lipid type, and two amphipathic membrane scaffolding proteins (MSPs) stabilizing the lipids by wrapping round the rim of the bilayer in a belt-like manner (Denisov *et al.*, 2004; Skar-Gislinge *et al.*, 2010). The use of nanodiscs to stabilize membrane proteins in solution in order to perform solution studies has previously been demonstrated in several cases (Bayburt & Sligar, 2003; Duan *et al.*, 2004; Shaw *et al.*, 2006; Kynde *et al.*, 2014). It has also previously been shown that cytochrome P450 3A4 (CYP3A4) can be homogeneously incorporated as monomers into phospholipid nanodiscs (Baas *et al.*, 2004), which is a strict requirement for the analysis of SAS data using the *ab initio* methodology.

SAS enables studies of macromolecules in solution in a native-like environment, but is generally perceived as a low-resolution technique with a relatively low information content. This perception, however, is not completely fair. It is true that the Bragg condition, along with the experimental noise, sets a lower limit on the shortest distance observable, at around 1 nm. However, it is often possible to resolve distances longer than this limit with a resolution as good as 2 Å (Andersen *et al.*, 2011). While the information content remains low compared with methods such as NMR and X-ray crystallography, the use of several data sets and the combination of small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) can greatly increase the information available for a detailed structural analysis of SAS data (Skar-Gislinge *et al.*, 2010; Pedersen *et al.*, 2014).

In order to maximize the amount of information available through SAXS, two data sets, one for loaded and one for unloaded discs, were analysed in the present study. CYP3A4 is only anchored to the membrane *via* a hydrophobic α-helix extending into the hydrophobic part of the phospholipid bilayer and it was reasonable to assume that the perturbation of the bilayer lipids by CYP3A4, on average, was minimal. An unloaded POPC nanodisc prepared under the same conditions as the POPC nanodisc loaded with CYP3A4 was used as a reference sample, from which the bilayer parameters could be extracted. These were subsequently used in the modelling of the scattering data from the CYP3A4-loaded nanodisc.

In this publication, we describe for the first time how the *ab initio* shape-reconstruction approach, originally developed to determine the low-resolution structure of water-soluble proteins from small-angle scattering data (Svergun, 1999), can be generalized to also cover membrane-anchored proteins. This is made possible through combination with a recently established modelling approach for phospholipid nanodiscs with incorporated membrane proteins (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2011; Kynde *et al.*, 2014).

We used this newly developed hybrid approach (Kynde *et al.*, 2014) to investigate the shape and mode of incorporation of the main human xenobiotic metabolizing cytochrome P450, CYP3A4, into the phospholipid membrane. This is to our knowledge the first time that the shape of a membrane protein has been reconstructed from small-angle scattering data while located in a native-like lipid environment.

In addition to this, the obtained SAXS data were also analyzed using a rigid-body approach in which a molecular-dynamics (MD) model of CYP3A4 (Denisov *et al.*, 2012) is fitted to the data in combination with the model for the nanodisc. This is in principle the same approach as was used in our recent study of the 7TM bacteriorhodopsin embedded in nanodiscs (Kynde *et al.*, 2014).

The resulting *ab initio* models of CYP3A4 are in very good agreement with the structure predicted by the corresponding

MD simulation. The localizations of CYP3A4 as determined from the *ab initio* and rigid-body analyses of the SAXS data are in practice identical and the distance between the catalytic domain of CYP3A4 and the phospholipid bilayer surface are in very good agreement with theoretical prediction (Denisov *et al.*, 2012) as well as with the result obtained from a protein crystallographic study of a close structural homologue (Monk *et al.*, 2014).

The *ab initio* approach (Svergun, 1999) has proved to be a very powerful method for studying monodisperse soluble proteins using small-angle scattering (Volkov & Svergun, 2003; Svergun *et al.*, 2001) and has greatly accelerated the study of proteins in solution. It is our hope that the use of the nanodisc system and SAS, in conjunction with complementary structural techniques such as cryo-EM (Frauenfeld *et al.*, 2011) and NMR (Hagn *et al.*, 2013), will be able to accelerate studies of membrane proteins that have previously been inaccessible to structural studies.

2. Materials and methods

2.1. Preparation of CYP3A4 POPC nanodiscs

2.1.1. Expression and purification of CYP3A4. Cytochrome P450 3A4 (CYP3A4) was expressed from the NF-14 construct in the PCWori+ vector provided by Dr F. Peter Guengerich, Vanderbilt University, Nashville, Tennessee, USA. CYP3A4 was expressed and purified as described previously (Denisov *et al.*, 2006) and its concentration was determined using carbonmonoxy difference spectra (Omura & Sato, 1964).

2.1.2. Purification of the membrane-scaffold protein. The expression and purification of MSP1D1(−) has previously been described (Denisov *et al.*, 2004). The MSP1D1(−) was engineered with a *Tobacco etch virus* (TEV) protease site to cleave the N-terminal hexahistidine tag used for purification, as described previously (Denisov *et al.*, 2004). The MSP1D1(−) was cleaved using a 100:1 MSP1D1(−):TEV protease weight ratio in 20 m*M* Tris–HCl pH 7.4, 0.1 *M* NaCl, 0.01% NaN_3 (standard disc buffer) at 28°C overnight. MSP1D1(−) was purified on an Ni–NTA column and concentrated. The concentration was determined by spectrophotometric analysis at 280 nm.

2.1.3. Self-assembly and purification of nanodiscs containing CYP3A4. The self-assembly of CYP3A4 nanodiscs is based on previously described procedures (Baas *et al.*, 2004). The dried POPC phospholipids (Avanti Polar Lipids, Alabaster, Alabama, USA) used in these experiments were solubilized using 100 m*M* cholate solution to give a final lipid concentration of 50 m*M*. CYP3A4 solubilized in 0.1%(*v*/*v*) Emulgen 913 (Karlan Research Products, Santa Rosa, California, USA) was combined with MSP1D1(−), POPC phospholipids and cholate in a molecular ratio of 0.1:1:64:128 [CYP3A4:MSP1D1(−):POPC:cholate]. After incubation, the self-assembly process was initiated by removing cholate and Emulgen 913 by a 4 h incubation on an orbital shaker at 4°C in the presence of Amberlite XAD-2 adsorbent (0.5 g per millilitre of mixture).

The assembled nanodiscs were separated into fractions containing unloaded nanodiscs and nanodiscs containing CYP3A4 on an Ni–NTA column using a pentahistidine affinity tag on the CYP3A4 and no affinity tag on the MSP1D1(−). Nanodiscs with CYP3A4 were eluted with buffer containing 0.3 *M* imidazole. Residual imidazole and any aggregates were removed by size-exclusion chromatography (SEC) on a Superdex 200 10/300 (GE Healthcare) column equilibrated with standard disc buffer running at 0.5 ml min^{-1} using a Waters (Milford, Massachusetts, USA) HPLC system equipped with a photodiode-array detector.

2.1.4. Self-assembly and purification of unloaded POPC nanodiscs. The POPC phospholipids used in these experiments were solubilized using 100 m*M* cholate solution to give a final lipid concentration of 50 m*M*. Cholate-solubilized MSP1D1(−) was added to the solubilized POPC phospholipids at a ratio of 65:1 [POPC:MSP1D1(−)]. The mixture was incubated at 4°C for 15 min. The self-assembly process was initiated by removing cholate by a 4 h incubation on an orbital shaker at 4°C in the presence of Amberlite XAD-2 adsorbent. The mixture was then filtered and fractioned on a Superdex 200 10/300 column equilibrated with 20 m*M* Tris pH 7.4, 0.1 *M* NaCl buffer running at 0.5 ml min^{-1}.

2.2. Small-angle scattering measurements

Samples of POPC nanodiscs with CYP3A4 at approximately 0.015 m*M* and without CYP3A4 at 0.025 m*M* were measured at 25°C on the BM29 BioSAXS beamline at the ESRF, Grenoble, France using the fixed settings of the instrument (Pernot *et al.*, 2010). The recorded detector intensities were automatically radially averaged, background-subtracted and absolute-scaled with an H_2O standard (Orthaber *et al.*, 2000) using the software at the beamline (Pernot *et al.*, 2010). The scattering curves were subsequently logarithmically rebinned to about 100 experimental data points in order to improve the statistics of the single data points. The finally obtained calibrated scattering curves give the differential cross-section per unit volume of the sample as a function of $q = 4\pi\sin(\theta)/\lambda$, where λ is the wavelength of the X-ray source and θ is half the angle between the incident and the scattered beam.

2.2.1. Indirect Fourier transform. The same scattering data can be represented by a real-space function by performing a so-called indirect Fourier transform (Glatter, 1977), yielding the pair distance distribution function (PDDF). All PDDFs presented here were obtained using the Bayesian indirect Fourier transform (BIFT) method (Hansen, 2012; Vestergaard & Hansen, 2006). The BIFT method is able to estimate the information content in the data in terms of the number of good parameters, N_g. The number of good parameters is a more realistic estimate of the number of parameters required to represent the obtained data than the number of Shannon channels, as it takes into account the experimental errors in the data and the match between the size of the investigated structures and the covered q range.

2.3. Modelling of small-angle scattering data

2.3.1. General model assumptions. It is assumed that CYP3A4 is anchored in the bilayer membrane *via* a single α-helix extending into the hydrophobic interior. Previous MD simulations (Denisov *et al.*, 2012; Baylon *et al.*, 2013) suggest that on average about 15 POPC molecules per disc, or about 10%, interact directly with the α-helix. Owing to this relatively low number of lipids in contact with the membrane anchor, it is assumed as a first approximation that the pertubation owing to the presence of CYP3A4 is negligible. This allows the determination of the structural parameters of the disc from the fit to the data for the unloaded disc. These parameters are then kept fixed in the fit to the data for CYP3A4-loaded discs, with the exception that the few POPC molecules replaced by the embedded α-helix were removed from the model.

2.3.2. Continuous modelling. The quantity measured by small-angle scattering is the differential scattering cross-section $\mathrm{d}\Sigma/\mathrm{d}\Omega$ as a function of the magnitude, q, of the scattering vector $\mathbf{q}$. This quantity can also be calculated from any dilute system of monodisperse particles in solution as

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}\Omega}(q) = n\langle|A(\mathbf{q})|^2\rangle_\Omega, \tag{1}$$

where n is the number density of the particles, $A(\mathbf{q})$ is the scattering amplitude of a single particle and $\langle\ldots\rangle_\Omega$ denotes the spherical average with respect to $\mathbf{q}$. The scattering amplitude contains the structural information about the particle and can be calculated as the Fourier transform of the excess scattering-length density, $\Delta\rho$, of the particle,

$$A(\mathbf{q}) = \int \Delta\rho(\mathbf{r}) \exp(i\mathbf{r}\cdot\mathbf{q})\,\mathrm{d}\mathbf{r}. \tag{2}$$

From this equation, the scattering amplitudes of many simple scattering-length density functions $\Delta\rho(\mathbf{r})$ have been calculated analytically or semi-analytically (Skov Pedersen, 1997).

In the present work, we use this approach to represent the nanodisc. The phospholipid bilayer of the nanodisc is divided into five elliptical cylinders of continuous scattering-length density, representing either phospholipid head groups, fatty-acid tails or the methyl end groups of the fatty acids. The two MSPs wrapping around the rim of the phospholipid bilayer are represented by a hollow cylinder with an inner radius corresponding to the radius of the lipid bilayer and a volume based on a composition of two MSPs. The model has been described in more detail in a previous publication (Skar-Gislinge & Arleth, 2011).

2.3.3. Discrete modelling. Solving the integral in (2) becomes tedious for objects with complex shapes, such as proteins. In this case, the complicated shape may more optimally be constructed by a number of discrete points such that the differential scattering cross-section of this assembly can be written as a sum of spherical harmonics (Svergun *et al.*, 1995),

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}\Omega}(q) = n\sum_{l=0}^{L}\sum_{m=-l}^{l}|B_{lm}(q)|^2, \tag{3}$$

$$B_{lm}(q) = i^l(2/\pi)^{1/2}\sum_{j=1}^{M}\Delta b_j J_l(qr_j)Y_{lm}^*(\theta_j,\varphi_j). \tag{4}$$

Here, J_l is the Bessel function of the first kind, Δb_j is the scattering length of the jth point, r_j, θ_j and φ_j are the coordinates of the jth point and Y_{lm}^* is the complex conjugate of the spherical harmonic. This point-model approach is used to describe the CYP3A4 molecule.

2.3.4. Hybrid modelling. By expanding the form-factor amplitude of the geometrical nanodisc model in spherical harmonics, $A_{lm}(q)$, the scattering of a particle represented by a hybrid model consisting of a analytical form factor and an assembly of points can be calculated as (Kynde *et al.*, 2014)

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}\Omega}(q) = n\sum_{l=0}^{L}\sum_{m=-l}^{l}|A_{lm}(q)+B_{lm}(q)|^2. \tag{5}$$

This hybrid approach is utilized in two approaches to describe the combined system of CYP3A4 reconstituted into a phospholipid nanodisc: an *ab initio* model where no prior knowledge of the shape of the CYP3A4 molecule is assumed and a rigid-body model utilizing external knowledge about the structure of CYP3A4.

2.3.5. Hybrid modelling: *ab initio* modelling. Representing the membrane protein through a number of points, while representing the nanodisc by a geometrical model, allows *ab initio* shape reconstructions to be performed along the lines described by Svergun *et al.* (2001). The parameters describing the nanodisc are first determined by fitting the nanodisc model to the data from the POPC nanodisc without the CYP3A4 protein. With these parameters locked, the *ab initio* procedure is used to determine the shape of the protein by fitting (5) to the data from the CYP3A4 nanodisc.

Initially, the points representing each amino-acid residue of the protein are placed randomly in a sphere with a radius of 40 Å. The model is minimized to the measured data by a simulated-annealing procedure in the following way: a random residue is selected and moved giving a new configuration, X. The move is then evaluated *via* a target function $F(X)$,

$$F(X) = \chi^2(X) + \alpha C(X) + \beta H(X) + \gamma T(X). \tag{6}$$

$\chi^2(X)$ is the deviation of X from the experimental scattering data and $C(X)$ measures the degree of connectedness of X. $H(X)$ compares the point–point distance distribution of X to an empirical distance histogram calculated from 20 entries in the PDB. In addition to these constraints, if there are less than 14 amino acids in the lipid bilayer then $T(X)$ is calculated as $T(X) = (n-14)^2$ and if there are more than 14 then $T(X) = 0$. This was performed in order to promote configurations that extend into the bilayer. If a new configuration, X_2, had $F(X_1) > F(X_2)$ it was accepted; on the other hand, if $F(X_1) < F(X_2)$ then the move to configuration X_2 was accepted only with the probability

$$P = \exp\{-[F(X_2) - F(X_1)]/B\}. \tag{7}$$

One pass is the completion of n_{aa} successful moves, where n_{aa} is the number of movable points, which in this case is the

number of amino acids. Initially B was set to $\chi^2(X_0)/10$, but after each pass B was decreased by 5%.

2.3.6. Hybrid modelling: rigid-body modelling. To evaluate the results of the *ab initio* modelling, a rigid-body model of the CYP3A4 molecule was constructed from a previously published molecular-dynamics (MD) simulation of CYP3A4 inserted into a POPC bilayer (Denisov *et al.*, 2012). We expect that this provides a sufficiently realistic model of the solution structure of the protein inserted into a lipid bilayer. In contrast to the *ab initio* approach, the shape of the protein is now defined by the atomic model and only the orientation with respect to the nanodisc is fitted.

As in previous work (Kynde *et al.*, 2014), the atomic structure from the MD model is coarse-grained by placing a point scatterer at the centre of scattering and with a scattering length corresponding to the atomic composition of the corresponding amino-acid residue. The scattering amplitude is then calculated using (4).

In the rigid-body approach, (5) is again fitted to the scattering data from the CYP3A4 nanodisc, but this time with a more traditional steepest-descent algorithm (Skov Pedersen, 1997), minimizing χ^2 with respect to the orientation and the relative position of the CYP3A4 protein and the nanodisc.

When combining the CYP3A4 and nanodisc models, the scattering length of every point placed in the bilayer is recalculated in order to take into account that the protein displaces lipids corresponding to the volume of the amino acids represented by the points. The model was implemented in the open-source SAS fitting framework *Willitfit?* (Pedersen *et al.*, 2013).

3. Results

3.1. Initial analysis of SAXS data from nanodiscs with and without embedded CYP3A4

The SAXS data from the POPC nanodiscs with and without embedded CYP3A4 are plotted in Fig. 1(*a*). Fig. 1(*b*) plots the corresponding pair distance distribution functions (PDDFs) obtained by Bayesian indirect Fourier transformation (BIFT). Both the scattering curves and the PDDFs are in good agreement with previously published scattering data on nanodiscs (Skar-Gislinge *et al.*, 2010; Denisov, Makris *et al.*, 2005). The PDDFs show that the maximum distance present in the samples increases from 110 to 135 Å when CYP3A4 is reconstituted in the nanodiscs.

The data are consistent with those presented in Denisov, Makris *et al.* (2005) and the increase in the D_{max} suggests that the large globular domain of CYP3A4 is located over the rim of the disc rather than over the central part of the disc. Negative contributions to the PDDF are present owing to the alternating signs of the scattering-length density present in the nanodisc (Glatter, 2002). This effect is present in both samples but is most pronounced in the case of the unloaded nanodisc, where the relative amount of negative-contrast hydrophobic carbon chains to positive-contrast protein is highest. Owing

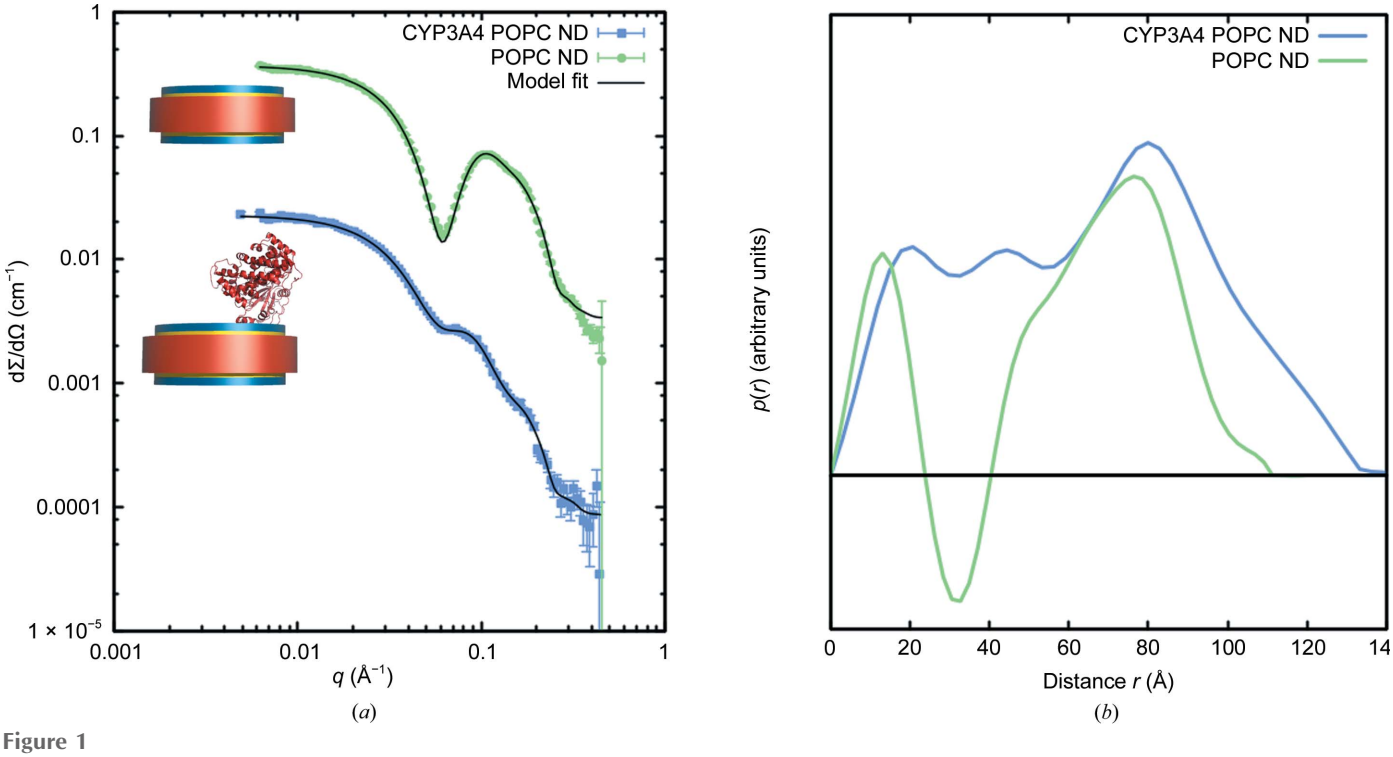


Figure 1
(*a*) SAXS data for POPC nanodiscs (green) and POPC nanodiscs with reconstituted CYP3A4 (blue). The full black curves show the corresponding model fits. The models are sketched in the insets, with protein in red, hydrophobic lipid tail groups in yellow and hydrophilic head groups in blue. The POPC nanodisc data are scaled by a factor of ten for clarity. (*b*) Pair distance distribution functions of the data shown in (*a*) obtained using the Bayesian indirect Fourier transform method. IFT fits to data are not shown.

to these complex interactions between phases of different contrast a detailed interpretation of the PDDFs is not meaningful without the use of an underlying model.

3.1.1. Estimation of information content. In order to avoid overfitting the data when performing model analysis, the information content in terms of the number of good parameters, N_g, was estimated through the BIFT analysis. For the POPC nanodisc the number of good parameters $N_g = 11.3 \pm 0.1$; for the CYP3A4 nanodisc $N_g = 8.0 \pm 0.5$.

3.2. Model-based structural analysis

3.2.1. POPC nanodisc. The scattering data from the POPC nanodiscs were interpreted using a previously published molecular constrained model to calculate the small-angle scattering from phospholipid nanodiscs (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2011), in which the nanodisc is modelled as a stack of elliptical cylinders representing the phospholipid bilayer core and a hollow elliptical cylinder representing the protein belt spanning the rim of the bilayer. The scattering can then be calculated according to (2). By systematically incorporating constraints based on the

Table 1
Fitting parameters for the nanodisc model.

Parameters are reported for POPC-loaded nanodiscs (POPC ND) and CYP3A4-loaded nanodiscs (CYP3A4 POPC ND). In the fit to the CYP3A4-loaded nanodiscs all parameters except the interface roughness terms were assumed to be identical to those obtained in the model fit to the POPC nanodisc data. A_{head} is the area taken up by one phospholipid head group in the disc, N_{lipids} is the number of phospholipids per nanodisc, N_{H_2O} is the number of hydration water molecules associated with each phospholipid head group, roughness gives the average interfacial roughness of all interfaces, η_{Belt} is the partial specific molecular volume of one MSP, η_{POPC} is the partial specific molecular volume of one POPC molecule and η_{CYP3A4} is the partial specific molecular volume of CYP3A4.

Parameter	POPC ND	CYP3A4 POPC ND
ε	1.47 ± 0.01	–
A_{head} (Å^2)	64.9 ± 0.67	–
N_{lipids}	124.20 ± 0.31	–
N_{H_2O}	0.91 ± 0.23	–
Roughness (Å)	4.56 ± 0.02	5.01 ± 0.07
η_{Belt} (Å^3)	25261.5 ± 71.7	–
η_{POPC} (Å^3)	1292.91	–
η_{PO}† (Å^3)	961.9	–
η_{PC}† (Å^3)	331.01	–
η_{CYP3A4} (Å^3)	–	73661 ± 252

† Parameter was not fitted but was calculated from η_{POPC}.

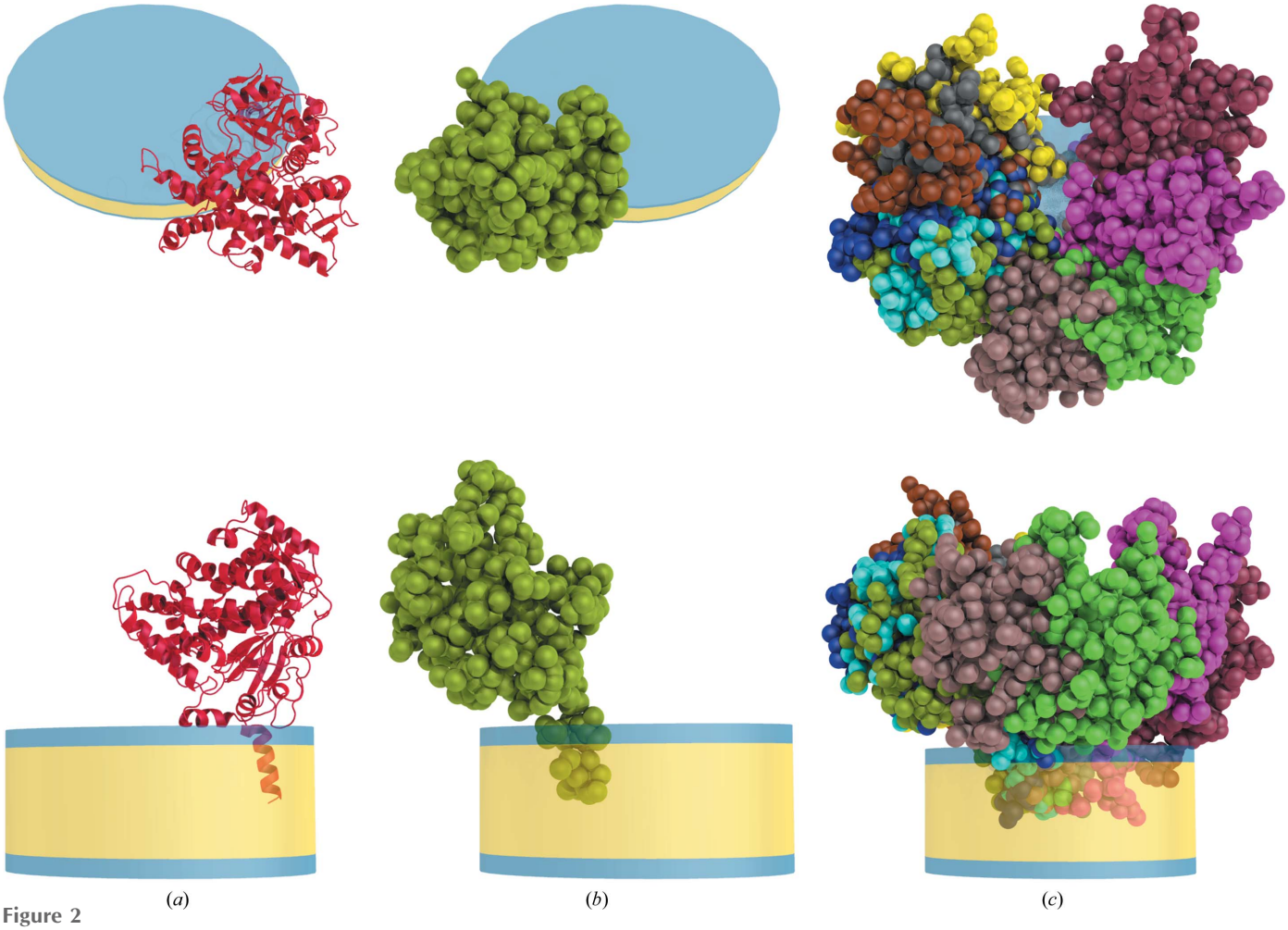


Figure 2
Model results. The lipid bilayer of the nanodisc is shown in blue (hydrophilic) and yellow (hydrophobic). The membrane-scaffolding belt has been omitted for clarity. (*a*) Result of rigid-body analysis of the scattering data. (*b*) Top and side view of one particular result of an *ab initio* analysis of the scattering data. (*c*) Top and side views of ten results of subsequent *ab initio* analyses. The phospholipid head and tail groups of the nanodisc are represented by the blue and yellow discs, respectively. The belt protein is not shown.

molecular composition of the nanodisc, the number of free parameters in this model is reduced to a total of seven, which is significantly less than the N_g for the data set. The fit to the measured POPC nanodisc data using this model is shown in black in Fig. 1(*a*) and the resulting fitting parameters are shown in Table 1. In accordance with previous work (Skar-Gislinge *et al.*, 2010), we find that the POPC nanodiscs have a slightly elliptical cross-section and that the average POPC area per head group is slightly larger than that expected for POPC in an infinitely large bilayer structure.

The small shoulder present in the data at q values between 0.3 and 0.4 Å^{-1} is most likely owing to internal structures of the protein belt. Similar shoulders are commonly observed in scattering from α-helical proteins (Svergun *et al.*, 2001). The feature is not captured by the nanodisc model, which assumes a homogeneous distribution of density inside the elliptical cylinders.

3.2.2. POPC nanodisc with embedded CYP3A4: *ab initio* shape reconstruction. By describing the membrane protein as an assembly of points representing each amino acid, the hybrid model allows an *ab initio* reconstruction to be made of the shape of a membrane protein anchored in a nanodisc.

Fig. 2 shows the results of ten *ab initio* shape reconstructions. It can be seen that the orientations of the ten *ab initio* models are different in relation to the major axis of the ellipse-shaped nanodisc. However, all solutions exhibit an overall similar shape; in particular, they all exhibit a stalk-like domain inserted into the membrane and a globular domain extending above the membrane. All solutions also have approximately the same distance between the globular domain and the bilayer. No solutions are found where the CYP3A4 molecule is placed near the centre of the nanodisc. Instead, all of the solutions have the membrane-bound stalk inserted about 20 Å from the rim of the phospholipid disc, with the globular domain placed over the MSP rim of the nanodisc. The number of POPC molecules replaced by the embedded part of the protein was found to be 3.5 ± 0.8. This gives confidence to the assumption that the perturbation of the disc is negligible.

The models were aligned and combined to produce a consensus model using *DAMAVER* (Volkov & Svergun, 2003). By remapping the models onto a grid, *DAMAVER* aligns the models and counts the frequency with which a given point in the grid occurs in the reconstructed models. The combined model is shown in two representations in Figs. 3(*a*) and 3(*b*). In Fig. 3(*a*) each point of the combined model is represented by a sphere. In Fig. 3(*b*) each point of the combined model is represented by a sphere with a radius of $N^{1/2}$, where N is the number of times that the point occurs in the ten aligned shape reconstructions. As is generally the case for *ab initio* models, the corresponding fits were in perfect agreement with the measured data and will not be shown here. This gives a picture of how conserved each point is in the ten *ab initio* models. Both representations of the combined model are compared with the model of CYP3A4 used in the rigid-body analysis (shown in Fig. 3*c*). Using *DAMAVER*, a filtered model containing the most occurring points cut off at the average volume of the ten shape reconstructions was also calculated. This model is shown as a semi-transparent surface in Fig. 3(*b*).

3.2.3. POPC nanodisc with embedded CYP3A4: rigid-body analysis. As described in §2, a rigid-body model of the CYP3A4 molecule was constructed using the relative atomic positions of a molecular-dynamics simulation of CYP3A4 embedded in a POPC bilayer (Denisov *et al.*, 2012). Using the hybrid approach, this was combined with a model of the nanodisc using the same model and parameters as those determined for the POPC nanodisc. The resulting number of fitting parameters in this model was six: three spatial co-ordinates and the rotation around the bilayer normal of CYP3A4, an adjustment factor for the density of the CYP3A4 molecule and the roughness of the phospholipid bilayer.

The model fit to the scattering data for the CYP3A4 nanodiscs is shown in black in Fig. 1(*a*) and the corresponding direct-space model is shown in Fig. 2(*a*). CYP3A4 is found to be oriented with the hydrophobic membrane anchor inserted into the membrane at the edge of the phospholipid disc, the membrane-facing part in contact with the upper part of the bilayer and part of the soluble domain extending over the edge of the bilayer. The placement at the rim of the phospholipid bilayer is in agreement with previous SAXS measurements of CYP3A4 reconstituted into nanodiscs.

Comparing the *ab initio* results with the rigid-body analysis, it can be seen that the solution found using the rigid-body analysis is in fact one of several possible solutions. This shows that by using the *ab initio* approach, which employs a simulated-annealing minimization, we are able to probe the solution space more effectively, finding possible solutions that would otherwise been overlooked in a rigid-body approach. The rigid-body approach, on the other hand, uses a steepest-descent approach and may become trapped in local minima.

4. Discussion

The *ab initio* molecular envelope of CYP3A4 inserted into a POPC nanodisc shown in Fig. 3 agrees well with the previously published molecular-dynamics simulation of the membrane-anchored protein (Baylon *et al.*, 2013; Denisov *et al.*, 2012) and the crystal structure of the soluble domain of the protein (Yano *et al.*, 2004). Furthermore, the suggested mode of incorporation, *via* a helix extending into the bilayer, is confirmed. This clearly demonstrates the general possibility of using *ab initio*-based analysis to resolve the shape of membrane-anchored proteins incorporated into nanodiscs.

The distance between the globular domain of CYP3A4 and the lipid membrane can be quantified by comparing the centre of mass of the different *ab initio* predictions of the protein as well as of the rigid-body representation and comparing these with the position of the lipid bilayer midplane. For the *ab initio* predictions, an average distance of 49.8 ± 3 Å was found based on the ten *ab initio* runs, whereas the rigid-body analysis yielded a distance of 45.9 Å. The half-bilayer thickness of the nanodisc is found to be around 20 Å (calculated from η_{POPC}/A_{head}). This small difference between the *ab initio* analysis and the rigid-body analysis is very likely to be a reflection of the

higher degrees of freedom of the *ab initio* analysis and was not interpreted further. The result of the rigid-body analysis can be directly compared with the previously obtained MD-based results (Denisov *et al.*, 2012; Baylon *et al.*, 2013) through the position of the centre of mass of the G′-helix (Baylon *et al.*, 2013) in relation to the membrane surface. From the rigid-body analysis a distance of 0.5 Å between the G′-helix and the upper hydrophilic layer of the membrane was found, indicating a close contact between the G′-helix and the membrane surface. This is in very good agreement with the MD-based results, which suggest that the G′-helix is effectively embedded in the upper hydrophilic part of the bilayer. The same observation was made when comparing our results with a proposed model of incorporation proposed from a recent full-length crystal structure of the related cytochrome P450 CYP51 (Monk *et al.*, 2014). This consistency is very interesting, as the model presented here is the first to be directly based on experimental data of CYP3A4 in a membrane-like environment.

Both the *ab initio* and rigid-body analyses of the measured data systematically place the CYP3A4 molecule at the rim of the phospholipid bilayer, in agreement with what was hypothesized from previous SAXS experiments (Baas *et al.*, 2004). The observation that none of the *ab initio* models gave a solution with the catalytic domain near the centre of the nanodisc does not necessarily mean that CYP3A4 is never found there; however, it does show that the centre of the disc is not preferred over the rim of the nanodisc. It has previously been shown that the lipids along the edge of the disc are slightly perturbed by the proximity of the MSPs (Denisov,

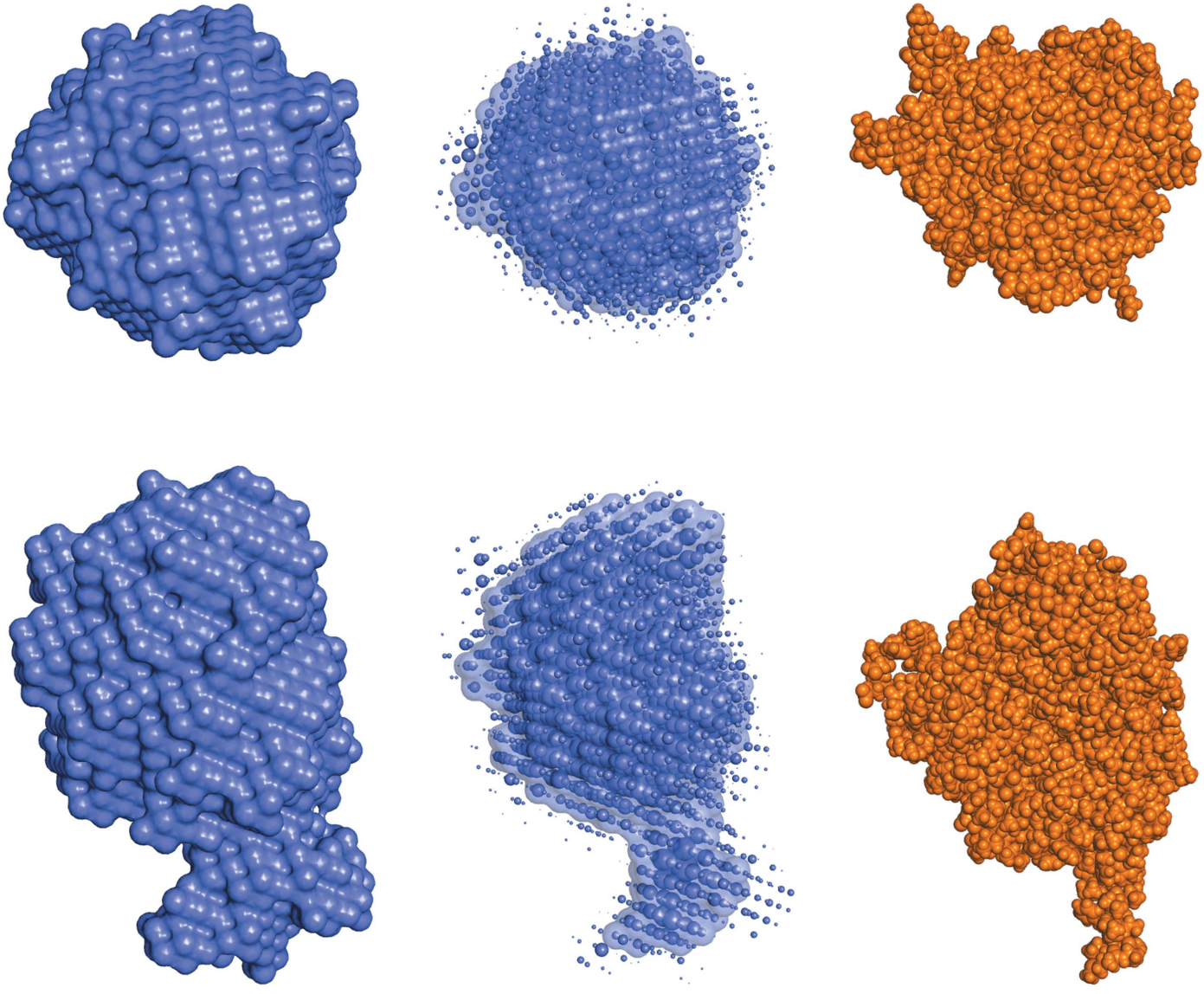


Figure 3
Ab initio shape reconstructions of CYP3A4 in nanodiscs. (*a*) Average of all ten models. (*b*) The average model is shown as spheres, where the radius of each sphere is the square root of the frequency with which it appears in all ten models. The transparent surface represents a filtered model calculated *via* the *DAMFILT* routine. (*c*) Space-filling rendition of the atomic model of CYP3A4 used for the rigid-body analysis. All models are drawn on the same scale.

McLean *et al.*, 2005; Skar-Gislinge *et al.*, 2010); this perturbation may be lessened to some degree when CYP3A4 is located at the rim, replacing some of the perturbed edge phospholipids. CYP3A4 is thought to be able to move along the surface of the membrane in order to locate its redox partner CPR. The driving force for this movement may very well be the relief of tension between the lipids and the protein belts that also drives CYP3A4 towards the edge of the discs.

In the present case, the data agree with the simple preassumption that the bilayer is not significantly perturbed by incorporation of the CYP3A4 membrane anchor. However, the preferred positioning of CYP3A4 near the edge of the discs contains indirect information about the abovementioned lipid–protein interactions.

The apparent lack of a preferred horizontal orientation of CYP3A4 in relation to the horizontal orientation of the disc (see Fig. 2) is an indirect sign that there is no interaction between CYP3A4 and any specific part of the scaffolding proteins which would constrain the entire CYP3A4–nanodisc complex. As a consequence, a level of structural dispersion with respect to the horizontal orientation of CYP3A4 relative to the lipid disc is to be expected in the samples. While it would in principle be trivial to incorporate this flexibility into the analysis, it would also require more fitting parameters in the model and has therefore not been attempted based on the present data. However, the fact that good and consistent reconstructions of the CYP3A4 molecular envelope are obtained and perfect fits can be obtained with the rigid-body approach shows that the effect on the data of this (hypothetical) flexibility is only small.

In a broader perspective, we find that it is possible to obtain a low-resolution shape from SAXS data of a membrane protein when reconstituted into a nanodisc. This becomes possible through the hybrid *ab initio* approach, which is directly compatible with the tools and methodology that are already widely applied for interpreting *ab initio* data from water-soluble proteins. In this context, it is particularly interesting to note that even though the sample is probably not fully structurally homogeneous, with respect to the horizontal orientation of CYP3A4 in the nanodisc the information on the shape of the membrane protein can still be reconstructed through hybrid-approach *ab initio* modelling.

SAS data from membrane proteins incorporated into nanodiscs will, in addition to information on the structure of the membrane protein, also contain direct structural information about the surrounding phospholipid bilayer.

The approach presented in this article provides perspectives for extracting such information under conditions that mimic those of the biological membrane, along with information about protein mobility and protein–protein interactions. This type of information could become accessible through systematic variation of the lipid composition of the nanodiscs. Also, the local packing of the lipids could be indirectly modulated through the size of the discs or through variation of the surrounding buffer.

This information, which is not easily accessible using other experimental methods, is valuable for detailed understanding of membrane-protein functions and understanding of protein–protein interactions in functional complexes of membrane proteins, such as human cytochrome P450s and their redox partner CPR.

Acknowledgements

This work was supported by NIH grants GM33775 and GM110428 to SGS.

References

Andersen, H. D., Wang, C., Arleth, L., Peters, G. H. & Westh, P. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 1874–1878.
Baas, B. J., Denisov, I. G. & Sligar, S. G. (2004). *Arch. Biochem. Biophys.* **430**, 218–228.
Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. (2002). *Nano Lett.* **2**, 853–856.
Bayburt, T. H. & Sligar, S. G. (2003). *Protein Sci.* **12**, 2476–2481.
Baylon, J. L., Lenov, I. L., Sligar, S. G. & Tajkhorshid, E. (2013). *J. Am. Chem. Soc.* **135**, 8542–8551.
Cherezov, V. (2011). *Curr. Opin. Struct. Biol.* **21**, 559–566.
Denisov, I. G., Grinkova, Y. V., Baas, B. J. & Sligar, S. G. (2006). *J. Biol. Chem.* **281**, 23313–23318.
Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. (2004). *J. Am. Chem. Soc.* **126**, 3477–3487.
Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. (2005). *Chem. Rev.* **105**, 2253–2278.
Denisov, I. G., McLean, M. A., Shaw, A. W., Grinkova, Y. V. & Sligar, S. G. (2005). *J. Phys. Chem. B*, **109**, 15580–15588.
Denisov, I. G., Shih, A. Y. & Sligar, S. G. (2012). *J. Inorg. Biochem.* **108**, 150–158.
Duan, H., Civjan, N. R., Sligar, S. G. & Schuler, M. A. (2004). *Arch. Biochem. Biophys.* **424**, 141–153.
Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. (2010). *Proteomics*, **10**, 1141–1149.
Frauenfeld, J., Gumbart, J., van der Sluis, E. O., Funes, S., Gartmann, M., Beatrix, B., Mielke, T., Berninghausen, O., Becker, T., Schulten, K. & Beckmann, R. (2011). *Nature Struct. Mol. Biol.* **18**, 614–621.
Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Glatter, O. (2002). *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, edited by P. Linder & T. Zemb, ch. 3. Amsterdam: Elsevier.
Grinkova, Y. V., Denisov, I. G., McLean, M. A. & Sligar, S. G. (2013). *Biochem. Biophys. Res. Commun.* **430**, 1223–1227.
Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. (2013). *J. Am. Chem. Soc.* **135**, 1919–1925.
Hansen, S. (2012). *J. Appl. Cryst.* **45**, 566–567.
Kynde, S. A. R., Skar-Gislinge, N., Pedersen, M. C., Midtgaard, S. R., Simonsen, J. B., Schweins, R., Mortensen, K. & Arleth, L. (2014). *Acta Cryst.* D**70**, 371–383.
Mast, N., Liao, W.-L., Pikuleva, I. A. & Turko, I. V. (2009). *Arch. Biochem. Biophys.* **483**, 81–89.
Monk, B. C., Tomasiak, T. M., Keniya, M. V., Huschmann, F. U., Tyndall, J. D. A., O'Connell, J. D., Cannon, R. D., McDonald, J. G., Rodriguez, A., Finer-Moore, J. S. & Stroud, R. M. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 3865–3870.
Nelson, D. R. (2009). *Hum. Genomics*, **4**, 59–65.
Omura, T. & Sato, R. (1964). *J. Biol. Chem.* **239**, 2379–2385.
Orthaber, D., Bergmann, A. & Glatter, O. (2000). *J. Appl. Cryst.* **33**, 218–225.
Ortiz de Montellano, P. R. (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed. New York: Kluwer Academic/Plenum Publishers.
Otyepka, M., Berka, K. & Anzenbacher, P. (2012). *Curr. Drug Metab.* **13**, 130–142.

Pedersen, M. C., Arleth, L. & Mortensen, K. (2013). *J. Appl. Cryst.* **46**, 1894–1898.
Pedersen, M. C., Hansen, S. L., Markussen, B., Arleth, L. & Mortensen, K. (2014). *J. Appl. Cryst.* **47**, 2000–2010.
Pernot, P., Theveneau, P., Giraud, T., Fernandes, R. N., Nurizzo, D., Spruce, D., Surr, J., McSweeney, S., Round, A., Felisaz, F., Foedinger, L., Gobbo, A., Huet, J., Villard, C. & Cipriani, F. (2010). *J. Phys. Conf. Ser.* **247**, 012009.
Rosenbaum, D. M. *et al.* (2011). *Nature (London)*, **469**, 236–240.
Schuler, M. A., Denisov, I. G. & Sligar, S. G. (2013). *Methods Mol. Biol.* **974**, 415–433.
Shaw, A. W., Pureza, V. S., Sligar, S. G. & Morrissey, J. H. (2006). *J. Biol. Chem.* **282**, 6556–6563.
Skar-Gislinge, N. & Arleth, L. (2011). *Phys. Chem. Chem. Phys.* **13**, 3161–3170.
Skar-Gislinge, N., Simonsen, J. B., Mortensen, K., Feidenhans'l, R., Sligar, S. G., Lindberg Møller, B., Bjørnholm, T. & Arleth, L. (2010). *J. Am. Chem. Soc.* **132**, 13713–13722.
Skov Pedersen, J. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.
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., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Vestergaard, B. & Hansen, S. (2006). *J. Appl. Cryst.* **39**, 797–804.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D. & Johnson, E. F. (2004). *J. Biol. Chem.* **279**, 38091–38094.
Yu, X., Cojocaru, V. & Wade, R. C. (2013). *Biotechnol. Appl. Biochem.* **60**, 134–145.