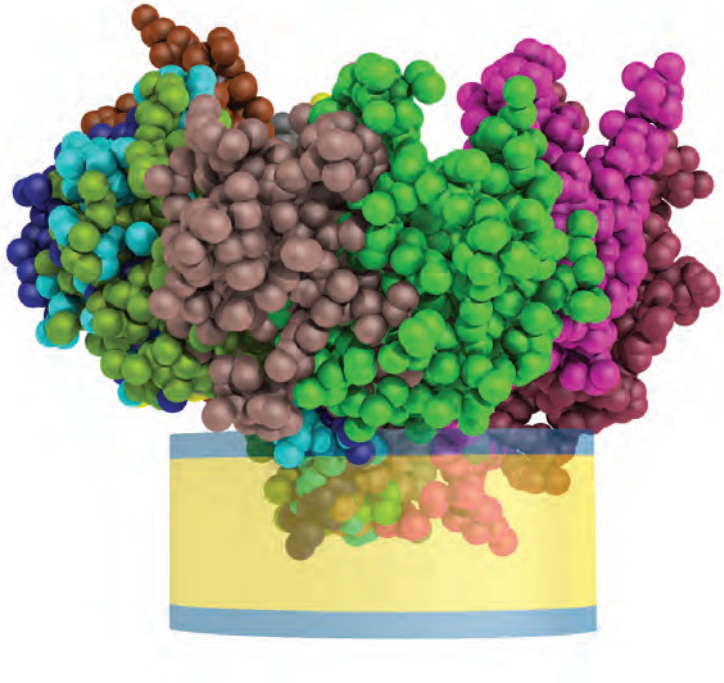
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

PhD thesis

Nicholas Skar-Gislinge

Developing Nanodiscs as a Tool for Low Resolution Studies of Membrane Proteins



Academic advisor: Lise Arleth

Submitted: 28/02/2014

Abstract

Phospholipid nanodiscs are $\sim$ 10 nm sized disc shaped particles consisting of roughly 150 phospholipids arranged in a central bilayer, stabilized by two amphipathic protein "belts" that wrap around the rim of the bilayer. Because they contain a small bilayer leaflet, in which membrane proteins can be incorporated, they can be used as a tool for solution studies of membrane proteins. So far most of the studies of membrane proteins using nanodiscs have been concerning the function of the incorporated membrane protein. However, due to the good control of the size and lipid composition of the nanodisc system, they seem an ideal tool for expanding the use of small angle scattering from studying water soluble proteins, to also include membrane proteins inserted into nanodiscs. This has been the main goal of this thesis.

In order to reach this goal it is necessary to have a detailed understanding of the nanodisc system, without the membrane protein, as well at the self-assembly process in general, and in particular in relation to incorporation of membrane proteins. This was the aim of work done early in this thesis. Here a detailed model for the small angle x-ray and neutron scattering from the empty nanodisc system was derived and used to describe the nanodisc system with great detail. The high level of detail was achieved by combining data from both small-angle x-ray and neutron experiments performed on the same samples as well as the incorporation of so-called molecular constraints. Using this model it was found that nanodiscs have an elliptical cross-section, and that the phospholipids in the nanodisc are slightly perturbed, compared to the phospholipids in vesicles.

Later, the self-assembly process of nanodiscs was investigated in more detail by varying the self-assembly conditions. Here it was found that the detergent used for solubilizing the phospholipids has a pronounced effect on the self-assembly process. This may be particularly relevant in order to obtain nanodisc samples of high quality and yield, needed for structural studies using small angle scattering. And which has been a challenge during the whole project.

In the later part of the project the main work has been on obtaining high quality samples as well as conducting SAXS and SANS measurements on these samples, and finally analyzing the measured data. The data analysis was done using a new hybrid approach combining continuous geometric and discrete modeling schemes. The hybrid approach was used to study the placement and orientation of the membrane protein bacteriorhodopsin incorporated into a nanodisc. Furthermore, the hybrid approach also allows for *ab. Initio* shape reconstruction of membrane proteins of an unknown shape incorporated into a nanodisc. This approach was investigated for the membrane anchored cytochrome P450 3A4 in the final part of the project.

Resume

Phospholipid nanodiscs er $\sim$ 10 nm skiveformede partikler bestående af omkring 150 phospholipider arrangeret som et dobbeltlag, stabiliseret af to amfipatiske protein "bælter" langs kanten af dobbeltlaget. Fordi de indeholder et lille stykke lipid dobbeltlag, hvori et membran protein kan indkorporeres, kan de bruges som et værktøj til studier af membranproteiner. Indtil videre er de fleste studier baseret på nanodiscs dog blevet foretaget med henblik på funktionen af det inkorporerede membranprotein. På grund af den gode kontrol af både størrelsen og lipidsammensætningen af nanodisc systemet, burde nanodiscs dog være et ideelt værktøj, til at udvide brugen af småvinkel spredning fra studier af vandopløselige proteiner, til også at omfatte membranproteiner i nanodiscs. At undersøge mulighederne for dette har været hovedformålet for denne afhandling.

For at kunne opnå dette mål er det nødvendigt at have en detaljeret forståelse af nanodisc systemet uden et inkorporeret membranprotein, samt en forståelse af den selvsanamlings proces hvorved nanodisc fremstilles. Denne forståelse er særligt vigtig i forhold til at optimere inkorporering af membranproteiner i nanodiscs. At opnå denne forståelse var formålet i den tidlige fase af projektet. Her blev en detaljeret model for små-vinkel røntgen og neutron-spredning fra "tomme" nanodiscs udledt og anvendt til en detaljeret beskrivelse af nanodisc systemet. Den høje grad af detalje blev opnået, ved at kombinere data fra både røntgen og neutron eksperimenter udført på de samme prøver, samt ved at bruge af såkaldte molekylære betingelser i modellen. Ved at anvende denne model blev det observeret, at nanodisc har et elliptisk tværsnit, samt at phospholipiderne i nanodisc'en befinder sig i en let pertuberet tilstand sammenlignet med phospholipid vesikler.

Senere i projektet er selvansamlings processen af nanodiscs blev undersøgt nærmere ved at variere de betingelser hvorunder den foregår. Her blev det konstateret, at den detergent, der anvendes til solubilisering af phospholipiderne, har en kraftig effekt på selvansamling processen. Dette kan især være relevant for at fremstille de nanodisc prøver af høj kvalitet, der er nødvendige for strukturelle studier ved hjælp af små-vinkel spredning. Netop at fremstill prøver af en tilstrækkelig høj kvalitet, viste sig at være udfordring igennem hele projektforløbet.

I den senere del af projektet har det vigtigste arbejde både været at opnå prøver af en tilstrækkelig høj kvalitet, samt at foretage SAXS og SANS målinger på disse prøver og analysere de målte data. Dataanalysen blev foretaget ved anvende en nyudviklet hybrid metode, der kombinerer en traditionel geometrisk og diskret tilgang. Denne nye hybrid metode benyttes til at undersøge placeringen og orienteringen af membranproteinet bacteriorhodopsin, indkorporeret i en nanodisc. Desuden giver hybrid metoden også mulighed for *ab. Initio* shape reconstruction af membranproteiner med en ikke kendt form, indkorporeret i en nanodisc. Dette blev demonstreret for membranproteinet cytokrom P450 3A4, i den sidste del af projektet.

Preface

The work presented in this thesis was done from 2009 to 2012, at the Structural Biophysics (former Biophysics located at the Faculty of Life Sciences) group at the Niels Bohr institute, University of Copenhagen.

The main goal has been to develop the phospholipid nanodisc system into a tool for performing structural studies of membrane proteins using small-angle scattering. This has required a very multidisciplinary approach as I have had to learn a number of different techniques and subjects in addition to the main technique, which has been small-angle scattering. I have performed wet-lab work in order to prepare the nanodisc samples, and participated in several measurements at large scale facilities, such as the ESRF and ILL in France or the PSI in Switzerland. I have written computer programs for analyzing the data and during a one month stay at the university of Illinois, I worked with expressing and purifying the membrane protein Tissue Factor. Although this has been quite a challenge, it has been even more fun.

This thesis consists of two parts. The first part introduce the nanodisc system and describe the theoretical background for small angle scattering, setting the stage for the main scientific work presented in the second part. This part consist of 5 papers, of which 3 have been published and 2 are drafts for publications. Included in this part is also a short experimental report of the work done on the membrane protein Tissue Factor.

Acknowledgments

There are several people that deserve a special thanks for making this an enjoyable experience both on a personal level as well as on an academic level. First of all, I would like to thank my supervisor Lise Arleth for giving me the opportunity to continue the work we started in my master thesis in a Ph. D project, and for always being encouraging and interested. I would like to thank the people that I at some point have had to share *my* office with: Søren Kynde, Martin Cramer Pedersen, Pie Huda, Jesper Nyaard and Jens B Simonsen: It has been a pleasure. I would like to thank former and present members of the biophysics group: Kell Mortensen, Lars Øendal, Steen L. Hansen, Marianne Jensen, Jacob JK Kirkensgaard, Søren S. Nielsen, Grethe Vestergaard, Rasmus H. Nielsen and Selma Maric. I would also like to thank the people in the Sligar lab at university of Illinois, in particular, Steven Sligar, Illia Denisov and Mark McLean, for letting me loose in the lab. In particular I would like to thank Søren R. Midtgaard for numerous beamtimes and road trips across Europe from one large scale facility to the other.

Finally I would like to acknowledge the UNIK - Center For Synthetic Biology, the Graduate School for Molecular Biophysics and the Faculty of Life Sciences for funding of the Ph D project.

Contents

List of Abbreviations

$\Delta\rho$ Excess scattering length density or contrast

$\frac{d\Sigma}{d\Omega}$ Macroscopic differential scattering cross section

$\frac{d\sigma}{d\Omega}$ Differential scattering cross section

ρ Scattering length density

σ Scattering cross section

b Scattering length

apo-A1 Apolipo Protein A1

ATSAS 'All That Small Angle Scattering' is a SAXS data analysis software package developed by the Svergun group at EMBL Hamburg Outstation

CMC Critical micelle concentration

CryoEM Cryo-Electron Microscopy

DDM n-Dodecyl β-D-Maltoside

delTF Recombinant Tissue Factor optimized for expression in E. coli. and containing a histidine tag.

DLPC 1,2-DiLauroyl-sn-glycero-3-PhosphatidylCholine

DMPC 1,2-DiMyristoyl-sn-glycero-3-PhosphatidylCholine

DPPC 1,2-DiPalmitoyl-sn-glycero-3-PhosphatidylCholine

HDL High Density Lipoprotein

HDL particle High Density Lipoprotein particle

HIS tag Histidine Tag

LCAT lecithin-cholesterol acyltransferase

MSP Membrane Scaffolding protein

Ni-NTA nickel-nitrilotriacetic acid

NMR Nuclear Magnetic Resonance

OG n-Octyl β-D-glucopyranoside

P(r) Pair Distance Distribution Function

PDDF Pair Distance Distribution Function

POPC 1-Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphatidylCholine

rHDL Reconstituted High Density Lipoprotein

SANS Small-Angle Neutron Scattering

SAXS Small-Angle X-ray Scattering

SDS Sodium Dodecyl Sulfide

TEV Tobacco Etch Virus

TF Tissue Factor

Chapter 1

Introduction

1.1 Motivation

Membrane proteins are, as the name implies, located in the different membranes of the cell. In many cases they serve as gateways for molecules and signals into the cell making them of great relevance, for example as thargets for pharmaceutical drugs[137]. Their natural environment in the cell membranes, require that they contain hydrophobic domains which makes them inherently unstable in aqueous solution. This consequently makes them very challenging to study using traditional solution based techniques. In order to maintain the stability and function of the protein[136], the hydrophobic environment around these domains has to be maintained in one way or another[27].

This challenge becomes particularly evident when comparing the number of crystallized membrane proteins to the total number of crystallized proteins deposited in the Protein Data Bank (PDB). As of Feb. 27, 2014 there are 445[1] unique membrane protein structures solved using diffraction methods compared to a total 44403[2] unique structures in the PDB.

The main aim of this project has been to investigate the potential of nanodiscs as a tool for low resolution structural studies of membrane proteins using small-angle scattering. Nanodiscs[12] are ∼10 nm wide disc shaped lipoprotein particles containing a central phospholipid bilayer, encircled by two amphiphatic membrane scaffolding proteins (MSP) stabilizing the edge of the phospholipids. This is illustrated in figure 1.1. By reconstituting the membrane protein into the central phospholipid bilayer of the nanodisc, the protein can be kept in a stable and functional state. The nanodisc hence provides a close to native lipid environment[16, 23] that both stabilize the membrane protein in solution and makes it accessible to solution techniques.

In addition to being a tool for studying membrane proteins, nanodiscs are interesting examples

[1] According to the data base of membrane protein structures maintained by Stephen White as of Jan. 16'th, 2013. Counting only entries solved by x-ray diffraction. http://blanco.biomol.uci.edu.

[2] According to a search in the Protein Data Bank Feb. 27, 2014. Counting only protein entries solved by diffraction method and removing similar sequences at 100% identity. http://www.rcsb.org/pdb/search/advSearch.

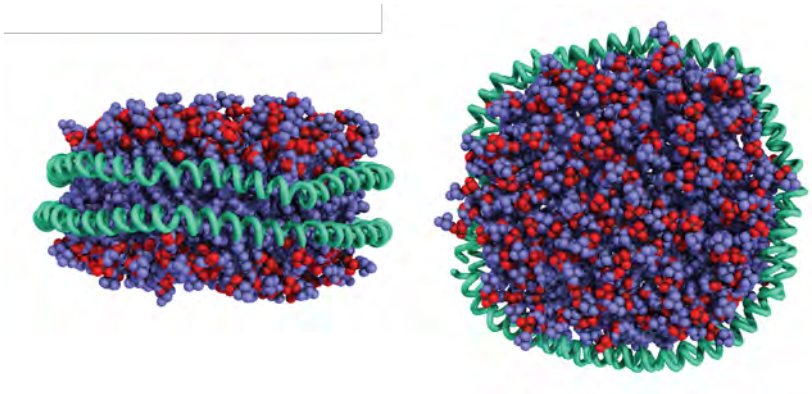


Figure 1.1: Side and top view of an atomic model of a nanodisc. The phospholipid bilayer is shown in blue and red and the two MSPs spanning the rim of the bilayer are shown in teal. The coordinates for the atomic model was kindly provided by Dr. Ilia G. Denisov

of self-assembling amphiphilic particles in their own right. The self-assembly process is governed by the same principles as classical amphiphilic self-assembly but is not yet described in detail. This broader understanding of the self-assembly process is needed for the study of membrane protein using nanodiscs, in order to be able to produce high quality samples in a predictable and reproducible manner.

Using Nanodiscs is not the only way of stabilizing membrane proteins in solution. A stabilizing hydrophobic environment can also be provided by detergent micelles, phospholipid liposomes or amphipols [34, 97], all of which are able to stabilize membrane protein in a functional state. Some of these methods are illustrated in figure 1.2.

In many cases detergents are used for stabilizing membrane proteins during purification process[98], and several crystal structures of membrane proteins have been obtained from detergent stabilized membrane proteins. However, for small-angle scattering the use of detergents is problematic as there also will be a population of micelles consisting purely of detergents present in the sample (in addition to the detergent/membrane protein micelles). The scattering signal from such a sample of mixed micelles will consist of scattering from both membrane protein containing micelles as well as pure detergent micelles, which has to be taken into account. More critically, the particles assembled from detergents often show a considerable structural dispersity, limiting the amount of information that can be extracted form the scattering signal. Finally, detergents only provide a hydrophobic environment, not a native lipid environment, and this may bias the structure or in some cases completely denature the membrane protein. Consequently, several membrane proteins are not functional in detergent micelles.

By stabilizing the membrane proteins in phospholipid liposomes, the phospholipids can provide a true lipid environment, however the mass ratio between the membrane proteins and the lipids is much in favor of the phospholipids. For a scattering experiment this means that the scattering signal originating from the membrane protein is lost in the scattering from the liposomes themselves.

A hybrid between the phospholipid vesicles and the detergent micelles are the so called bilayer

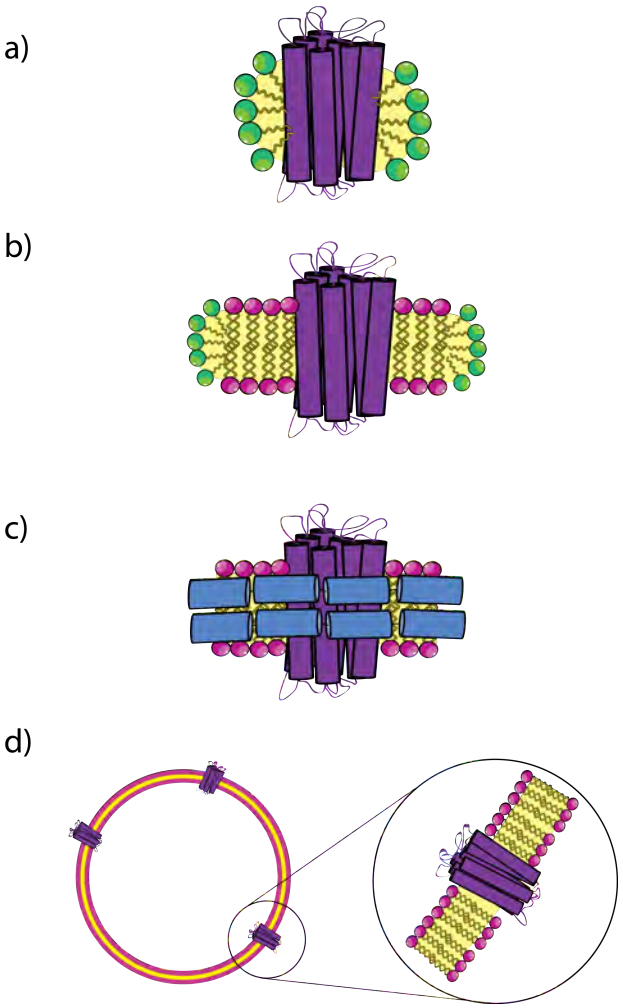


Figure 1.2: Illustration of various strategies for stabilizing membrane proteins in solution. a) detergent micelles, b) bicelles, c) nanodiscs and d) liposomes.

micelles or bicelles. These are produced by mixing short and long chained lipids amphiphiles that separate into regions of different curvatures[106]. This produces disc shaped bilayers with the long chained lipids in the center and the short chained lipids along the rim. Like the vesicles, these particles may contain a lipid environment but they can be made relatively small, giving them a lipid to protein mass ratio more favorable for scattering studies. This has successful been used to grow bicelle stabilized membrane protein crystals, for protein crystallography[132, 42]. However for small-angle scattering, bicelles as detergent micelles can also display a relatively large structural dispersity.

Amphiphatic polymers, known as Amphipols, is another option for stabilizing membrane proteins in solution. However, particles of amphipol stabilized membrane protein have been observed to exhibit a considerable structural dispersity, both with respect to the number of membrane proteins per particles, but also in the formation of larger amphipol aggregates[50].

Due to the two MSPs surrounding the phospholipid bilayer, the nanodisc are formed with a well defined size. The MSP has been engineered with different lengths, giving a precise control over the size of the resulting nanodiscs, by simply choosing a MSP with the desired length. This makes it possible to obtain nanodiscs with a diameter ranging from 7 nm to 20 nm[37][131]. In addition to a precise control of the size, the local phospholipid composition can accurately be controlled within the nanodisc[99]. The presence of phospholipids, also opens for systematic studies of the influence on the phospholipids on the membrane protein. As explained in section 2.4 on reconstitution of membrane proteins in nanodiscs, it is possible to prepare a highly homogenous sample consisting of only one membrane protein per nanodisc, overcoming the problem of polydispersity.

The size of the nanodiscs fall well within the 1 to about 100 nm size accessible to small-angle scattering. Combined with the high degree of control over the size, homogeneity and the phospholipid composition, the nanodisc is an ideal candidate for stabilizing membrane proteins in solution, for scattering studies.

1.2 Structural Studies of Membrane proteins in Solution

Small-angle scattering has been very successful in solution studies of water soluble proteins. A clear sign of this success is the existence of several small-angle scattering beamlines dedicated for the study of biomacro-molecules, in particular proteins. In contrast to the more common technique of crystal diffraction, small angle scattering is able to study the proteins in a solution environment. However, this comes at the cost of the atomic resolution that is often provided by protein crystal diffraction.

Recently nanodiscs have also been used for structural studies of membrane proteins using cryo-electron microscopy (cryoEM) and nuclear magnetic resonance (NMR). In cryoEM the sample is flash-frozen to cryogenic temperatures, in order to freeze the solvent in a glass state. As the sample is flash frozen, the macromolecules are frozen randomly oriented, and by recording micrographs of a large number of the sample molecules in different orientations, a 3D electron density map can be reconstructed.

Compared to small-angle scattering, cryoEM has a more relaxed requirement for structural homogeneity of the sample. Small angle scattering probe the bulk of sample, and the scattering signal contains contributions from all molecules in the probed volume. CryoEM, on the other hand, probe each individual particle one at a time. This mens that the samples used for the 3D reconstruction of the electron density are either picked by the operator or automatic software. Because the sample is frozen very rapidly it is in some cases even possible to study proteins caught in different conformational states[135].

In order to avoid radiation damage the intensity of the electron beam is relatively low, resulting in a low signal to noise ratio in the measurement. This means that molecules with a low electron density are not resolved vey well. This is the case for phospholipids. In a study by Frauenfeld *et al.*, cryoEM was used to reconstruct the structure of a large membrane protein channel reconstituted into a nanodisc[45]. While the protein channel is clearly visible the resolution of the phospholipid nanodisc is very weak.

The second technique, NMR, is a proven technique for obtaining high resolution structures of smaller biological macromolecules. However as the size of the molecules increase, the recorded spectra become harder to interpret. One reason for this is broadening of the spectral peaks due to a low rotation rate for large proteins. With respect to membrane proteins reconstituted into nanodisc, the nanodisc will both contribute to the measured signal as well as slow down the rotation.

Using NMR Hagn *et al.*[131] have recently published a high resolution structure of a small membrane protein reconstituted in a nanodisc. This was done using a shorter version of the MSP, in order to make smaller nanodiscs with a higher tumbling rate, and deuteration of both the MSP and the phospholipids in order to obtain a spectrum exclusively from the protonated membrane protein. However, the trick of increasing the tumbling rate, by making the nanodisc smaller, only works for membrane proteins completely integrated into the lipid bilayer. So in most cases the size limitations of NMR remain.

In addition to the NMR and cyoEM, Berthaud *et al.*[19] have successfully modeled the SAXS pattern of a membrane protein of known structure in detergent micelles. This was done using a size exclusion column connected to a synchrotron SAXS setup. They were successful in obtaining a realistic measurement of the pure detergent micelles, that could in turn be subtracted from the detergent/membrane protein measurement in order to obtain only the scattering from the detergent stabilized membrane proteins.

When comparing small angle scattering to NMR and cryoEM approaches described above, small-angle scattering has the property that the phospholipids are well represented in the scattering data. This opens up for investigating the structural interplay between the phospholipids and membrane proteins. Additionally, in neutron scattering it is possible to highlight different structural features based on their hydrogen content. This can either be done by expressing the protein under deuterated conditions or by tuning the scattering contrast by changing the D_2O content in the buffer. Using the combination of neutron and x-ray scattering it is possible to obtain information on the different parts that make up the nanodisc system, increasing the information available for analysis. However, interpreting the scattering signal from a nanodisc containing a reconstituted membrane protein require a very accurate model of the nanodisc in order to extract the information on the membrane protein. A main part of the work done in this thesis has focused on creating such a model. This work is presented in paper 1 and 2.

1.3 Results

During this project, several results has been achieved. These are presented in the included three published papers and two paper drafts. In addition an experimental study carried out at a one month stay at the University of Illinois is also presented. The papers and paper drafts are listed below along with a very short review of the main findings.

Paper 1 Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: casting the roles of the lipids and the protein
Nicholas Skar-Gislinge, Jens Bæk Simonsen, Kell Mortensen, Robert Feidenhans'l, Stephen G

Sligar, Birger Lindberg Møller, Thomas Bjørnholm, Lise Arleth
JACS 132 **2010** 39

It is found that the empty nanodiscs have an elliptical cross section and that the lipids are found in a perturbed state compared to lipid vesicles. This is ascribed as an effect of minimizing the hydrophobic mismatch between the lipids and the MSP.

Paper 2 Small-angle scattering from phospholipid nanodiscs: derivation and refinement of a molecular constrained analytical model form factor
Nicholas Skar-Gislinge & Lise Arleth
Phys. Chem. Chem. Phys. 13 **2010** 8

Detailed description the geometrical model used in paper 1 and 3, and the use of molecular constraints in order to minimize the number of free parameters.

paper 3 What determines the shape and stoichiometry of self-assembled phospholipid nanodics: Speed of self-assembly process, initial lipid:MSP stoichiometry or detergent type?
Nicholas Skar-Gislinge, Rasmus Høiberg-Nielsen & Lise Arleth
Draft Manuscript, to be published.

The self-assembly of nanodiscs is studied in detail. In particular the choice of detergent and the initial mixing ratio of the lipids and the MSPs is found to influence the number of lipids found incorporated in the assembled discs. It is also found that the discs exhibit a small polidispersity.

Paper 4 Small-angle scattering gives direct structural information about membrane protein inside lipid environment
Søren Kynde, Nicholas Skar-Gislinge, Martin Cramer Pedersen, Søren Roi Midtgaard, Jens Bæk Simonsen, Ralf Schweins, Kell Mortensen & Lise Arleth
Accepted in Acta Crystallographica Section D.

Description of how small-angle scattering from nanodiscs with reconstituted monomeric bacteriorhodopsin can be analyzed using a hybrid model combining both the geometrical model of the nanodisc, as presented in paper 2, and a discrete free form model of the membrane protein. It is also found that the incorporation of the membrane protein has a large effect on the surrounding phospholipids.

Paper 5 Ab Initio shape and membrane orientation of the human cytochrome P450 CYP3A4 in Nanodiscs resolved by SAXS
Nicholas Skar-Gislinge, Søren Kynde, Ilia G. Denisov, Xin Ye, Ivan Lenov, Stephen G. Sligar & Lise Arleth
Draft Manuscript, to be published.
The hybrid approach is used in order to do an *ab initio* shape reconstruction of the membrane protein, Cytochrome p450, while incorporated into a nanodisc. The study illustrate how small

angle scattering can be used in order to obtain structural information on a membrane protein of an unknown shape.

Experimental Work on Tissue Factor Incorporated in to Nanodiscs performed during stay at the Sligar Lab. University of Illinois The enzyme activity of the trans membrane blood coagulation enzyme tissue factor is know depend on the phospholipid composition of the host membrane and the presence of calcium ions. This may be linked to a structural change in the enzyme. The aim of this study is to investigate this possible structural change, and to gain experience in producing and handling membrane proteins. In the current state the study show a change in the scattering from the sample upon addition of calcium ions, however the nature of this change is not determined. A main obstacle is to obtain samples that are structurally homogenous enough.

In addition to the five included papers and drafts the author has contributed to a sixth paper that has recently been accepted in Acta Crystallographica Section D, it is not included as a part of this thesis[79].

Chapter 2

Nanodiscs

The aim of this chapter is to introduce the basic theory required in order to discuss the self-assembly of nanodiscs and to give an introduction to the nanodisc system. It is divided into three parts. The first part gives an overview of the principles of self-assembly of amphiphiles. The second part gives an introduction to the nanodisc system, starting with the very analog apo-A1 HDL particles from which the nanodisc system is derived. The final part discusses some of the practical aspects of reconstitution of membrane proteins into nanodiscs.

2.1 Self-Assembly of Amphiphiles

Amphiphiles are molecules that contain a hydrophobic and a hydrophilic part, often termed the head (hydrophilic) and tail (hydrophobic) of the molecule. Because these parts either interact favorable and unfavorable with water, they may self-assemble into larger structures, in which the hydrophobic parts can be screened from the water. The formation of these structures are governed by a combination of the hydrophobic effect [62] and the repulsion between the head groups, as well as the entropy of mixing[90].

Although many molecules display amphiphilic properties, this short review will mainly focus on detergents and phospholipids. Figure 2.1 shows the molecular structure of a number of amphiphilic molecules. All the molecules shown in the figure, perhaps except cholate, have a clearly defined polar or ionic head group and one or two hydrophobic alkyl tails. As will be discussed further, both the nature of the head group and the tail group(s) play an important role in the self-assembly of amphiphiles.

Phospholipids which make up part of the cell membrane and the center of the nanodisc, are triglyceride derived biomolecules containing two hydrophobic acyl tails, connected to a hydrophilic head group containing a phosphate group. Phospholipids with tails of more than 10 acyl groups form bilayer structures in water. They are named according the acyl chains and the type of head group: for example, the case of POPC depicted in figure 2.1, it contains a Palmitoyl (P) and an Oleoyl (O) tail group, and a PhosphatidylCholine (PC) head group. In addition to

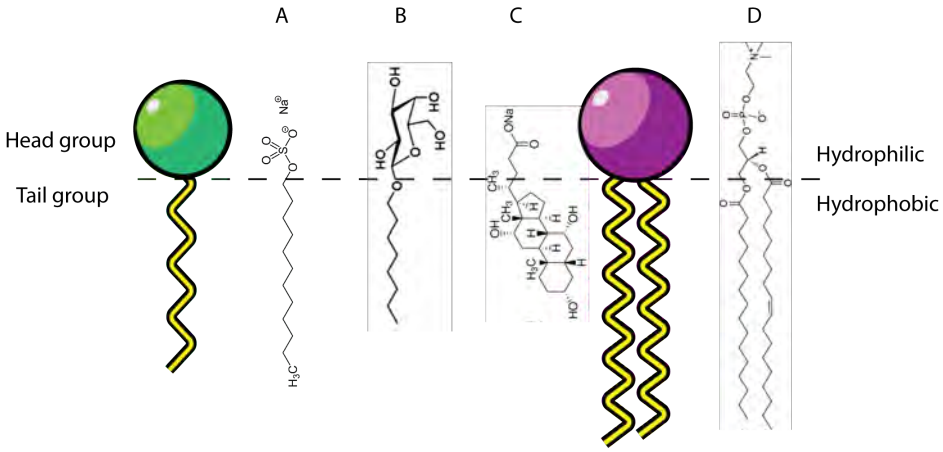


Figure 2.1: Molecular structures and cratoon representation of amphiphiles. The two detergents Sodium Dodecyl Sulfide (SDS) (A) and n-Octyl β-D-glucopyranoside (OG) (B) display a classically amphiphilic structure with a clear hydrophilic head and hydrophobic alkane tail groups. The third detergent, Sodium Cholate (C), does not display as well defined hydrophobic and hydrophilic parts. The phospholipid 1-Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphatidylCholine (POPC) (D) contains two hydrocarbon tails (one monounsaturated) and a zwitter ionic head group.

the zwitter ionic phosphatidylcholine (PC), phospholipids may have other types of head groups. For example, the negatively charged phosphatidylglycerol (PG) or phosphatidylserine (PS). The exact nature of the tail groups and the head group has a large effect on the physical properties of the phospholipid.

The study of the physical properties that govern the self-assembly of amphiphiles is a large and very active field. This is not only due to its close connection to industrial interests, such as cleaning detergents, emulsion stability or the encapsulation and formulation of hydrophobic pharmaceutical drugs, but also because nature uses amphiphiles to assemble very complex structures such as the cell membrane. This also means that this should be considered at short introduction into the key principles of self-assembly of amphiphilic molecules with respect to understanding self-assembly of nanodiscs. A more general and in depth treatment of the subject may be found in the books by Israelachvili [63] and Hyde[61].

2.1.1 The Hydrophobic Effect

The term "the hydrophobic effect" was introduced by Charles Tanford in 1973 and is crucial for understanding self-assembly of amphiphiles. In its essence, it describes the observation that oil and water separate in two phases. In terms of self-assembly of amphiphiles, it acts as a driving force causing the hydrophobic part of the amphiphiles to cluster together in aggregates, in order to minimize the hydrophobic interface to the solution.

It is classically considered a consequence of the apolar tails not being able to contribute to the hydrogen bond network of the surrounding water molecules. This forces the water molecules

Detergent	CMC / mM
Sodium Cholate	9-14
OG	18-26
DDM	0.18
DPPC	$0.46 \cdot 10^{-6}$

Table 2.1: Sources: Cholate: *Reis et al.*[101] OG: Lorber *et al.*[76]. DDM: Priv[98]. DPPC: Avanti Polarlipids Inc. All values are in H_2O at 20° C.

to form a more ordered structure around the hydrophobic part of the amphiphile, in order to maintain the number of hydrogen bonds. This ordering leads to a considerable loss of entropy, which is much larger than the entropy loss due to clustering of the amphiphiles.[63][134].

More recently it is being discussed, that for large (relative to water) molecules the hydrophobic effect should be considered being of a primarily enthalic nature[28]. This view consider that the solvation of a hydrophobic molecule creates an interface that disrupts the hydrogen bond network of the bulk water, giving molecules near the interface, on average, fewer hydrogen bonds. This will favor the formation of larger aggregates with a lower surface to volume ratio.

2.1.2 The Critical Micelle Concentration

Although the hydrophobic effect drives the amphiphiles towards aggregate formation, amphiphiles also exist as monomers in solution (ie. not participating in an aggregate). The formation of aggregates only begin once the monomer concentration reaches a certain critical concentration. This concentration is known as the critical micelle concentration (CMC). Once the CMC has been reached, the monomer concentration will remain at the CMC, whereas additional monomers will form aggregates.

Table 2.1 lists the CMCs of three common detergents and a phospholipid. In general, phospholipids tend to have very low CMCs. This is illustrated by the CMC of the phospholipid DPPC compared to the detergents in table 2.1.

The CMC phenomenon can be understood by considering the system at equilibrium, when the chemical potentials of amphiphiles participating in aggregates and monomers in solution are equal [63]:

$$\mu = \mu_1^\circ + k_b T ln(X_1) = \mu_2^\circ + \frac{k_b T}{2} ln(X_2/2) = \mu_3^\circ + \frac{k_b T}{3} ln(X_3/3)... = \mu_n^\circ + \frac{k_b T}{n} ln(X_n/n), \quad (2.1)$$

where μ_n° is the free energy of an amphiphile molecule in an aggregate of aggregation number n and X_n is the mole fraction of amphiphiles in aggregates of aggregation number n.

From this initial condition it can be shown that when adding amphiphiles to a solution, the concentration of free amphiphiles will rise to the CMC, after wich additional aggregates will start forming aggregates. Equation 2.1 also leads to the following expression, for the CMC for amphiphiles forming large aggregates[63]:

$$X_{1,CMC} = e^{-(\mu_1^\circ - \mu_n^\circ)/k_b T} \quad (2.2)$$

where $X_{1,CMC}$ is the CMC.

For amphiphiles forming small aggregates, the onset of aggregate formation is a more gradual process spanning over a range of concentrations. From 2.2 it can be seen that the CMC is dependent on the temperature. Additionally the free energies, μ_1° and μ_n°, and hence the CMC, also depend on the properties of solution, such as the pH and the ionic strength[101].

2.1.3 The Packing Parameter

In addition to the hydrophobic tails, responsible for driving amphiphiles together in aggregates, amphiphilic molecules also contain a hydrophilic part. The repulsive interactions between these, will contribute positively to the free energy per monomer in an aggregate, thereby working against the self-assembly. This gives rise to an optimal head group area where the hydrophobic attraction between the tails, acting to minimize the aggregate surface to volume ratio, is balanced by the repulsive interactions between the head groups[63][90]. The surface energy of an amphiphile in an aggregate, μ_n°, can as a first order approximation be written as:

$$\mu_n^\circ = a\gamma + \frac{C}{a}, \tag{2.3}$$

where a is the hydrophobic/hydrophilic interface area, γ the interface surface tension between the acyl tails and the solvent and C a constant expressing headgroup-headgroup interaction. Minimizing μ_n° with respect to a gives the optimal area per head group, a_0, expressing the balance between that attractive and repulsive interactions acting at the aggregate interface[63][90].

The shape of the aggregates can elegantly be linked to the molecular geometry of the amphiphilic monomers by the so-called packing parameter, P, introduced by *Israelachvili, Mitchell* and *Ninham* in 1976. This concept takes into account both the minimization of the interfacial area as well as the geometric constraints due to the packing of the hydrophobic tails [90]:

$$P = \frac{v}{l_0 a_0}, \tag{2.4}$$

where a_0 is the optimal head group area, v is the partial specific molecular volume of the hydrophobic tail and l_0 is the length of the fluid tail groups. For amphiphiles containing acyl chains, v and l_0 may be estimated using the Tanford formulas[124]:

$$l_0 \leq l_{max} = 0.154nm + 0.1265nm \cdot n \tag{2.5}$$

$$v \sim 27.4 + 26.9 \cdot n \tag{2.6}$$

Where n is the number of carbon atoms in the tail. For long chained lipids v/l_0 is practically constant with increasing tail length[70][63].

Amphiphiles with p $< 1/3$ are over all wedge or cone shaped and self-assemble into nearly spherical micelles. This is the case for many classical detergents detergents, such as SDS in low salt. Amphiphiles with a packing parameter in the range 1/2 to 1 form flat flexible bilayers. It is experimentally found that phospholipids form bilayers when P > 0.74 which is the case when

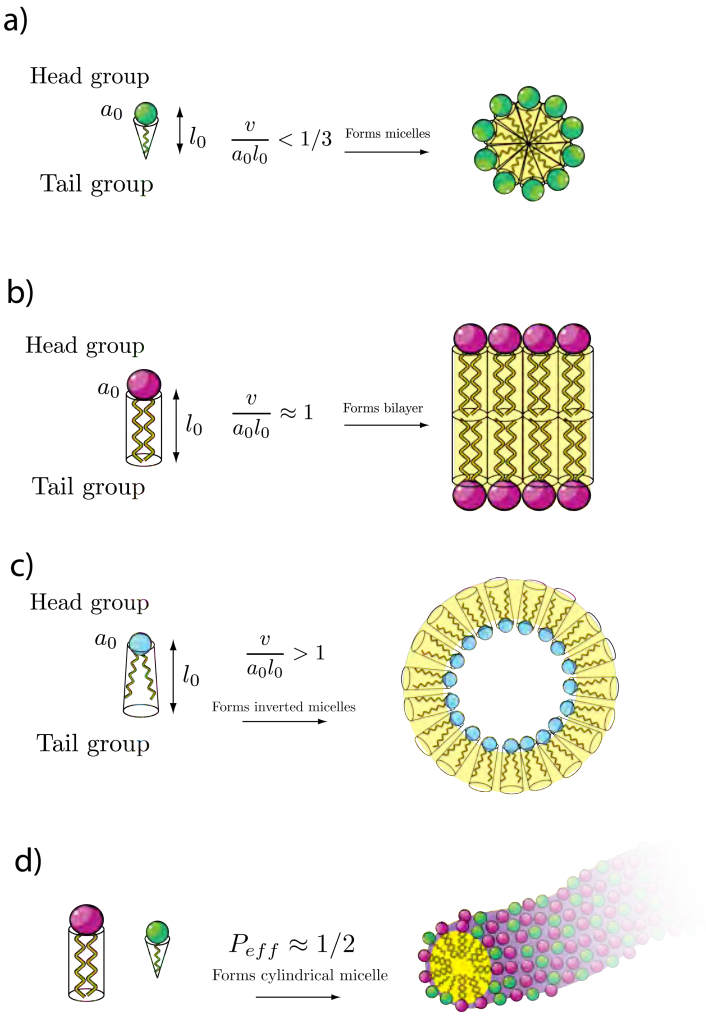


Figure 2.2: Using the packing parameter $P = \frac{v}{l_0 a_0}$ the shape of the self-assembled aggregates can be related to the shape of the monomer. P $< 1/3$ describe amphiphiles as wedges that form micelles (a), P ~ 1 as cylinders forming bilayers (b) and P > 1 as inverted cones forming inverted micelles (c). In many cases the packing parameter is additive and mixtures of amphiphiles form aggregates according to an effective average packing parameter (d)[70].

the tails contain more than 10 carbon atoms[70]. Amphiphiles with a packing parameter larger than 1 will form assemblies such as inverted micelles or cubic phases. The concept of the packing parameter and the associated structures are illustrated in figure 2.2.

The packing parameter can furthermore be related to an intrinsic curvature of the aggregate via[61]:

$$P = 1 - Hl + \frac{Kl^2}{3} \tag{2.7}$$

where K is the mean and H is the gaussian curvature defined as $K = (R_1^{-1} + R_2^{-1})/2$ and $H = R_1^{-1}R_2^{-1}$ and where R_1 and R_2 are the radii of curvature. If the intrinsic curvature cannot be accommodated in the shape of the aggregate, there will be an associated cost in energy as the molecules are forced to arrange themselves in a less optimal configuration. In a bilayer, such a deviation may be due to bending of the bilayer or, as discussed below, due to incorporation of detergents or proteins.

Although the packing parameter provides a convenient way of understanding the shape of an aggregate from the monomers, it only determines the local shape. For large aggregates the global shape is also determined by the flexibility of the aggregate. Finally, the size of the aggregates is furthermore constrained by the entropy of mixing, which favor small aggregates over large aggregates. Further details on this subject may be found in the previously mentioned references by Israelachvili [63] and Hyde *et al.*[61].

2.1.4 Mixtures of Amphiphiles

Mixtures of detergents may lead to the formation of micelles containing a mixture of both types of detergents. The CMC of such a mixed micelle will lie somewhere in between the CMCs of the pure detergents, depending on the ratio between them[57][63]. For ideally mixing detergents the mixed micelle CMC is[32]:

$$C_{cmc,1,2}(X_1) = \left(\frac{X_1}{C_{cmc,1}} + \frac{(X_1 - 1)}{C_{cmc,2}} \right)^{-1}, \tag{2.8}$$

where $C_{cmc,1}$ and $C_{cmc,2}$ are the CMCs of the individual detergents and X_1 is the mole fraction of detergent 1 in the micelle.

The shape of such mixed micelles may be described by an effective packing parameter. For ideally mixing amphiphiles the packing parameter is additive[70] and the effective packing parameter can be found by weighing the individual packing parameters according to the mole faction of each. For two amphiphiles, a and b, the effective packing parameter becomes:

$$P_{eff} = X_a P_a + X_b P_b, \tag{2.9}$$

where P_a and P_b are the packing parameters of amphiphile a and b, and X_a and X_b are their respective mole fractions in the aggregate.

In many systems the mixing is not ideal and there is a segregation of the two amphiphiles into regions with different curvature. This is for example the case in the so-called "bicelle" system. This segregation is often temperature dependent leading to a complex phase behavior[100][91].

2.1.5 Bilayer Detergent Interactions

The phospholipid nanodiscs used in this work are all prepared using double chained phospholipids such as POPC or DLPC that have packing parameters close to 1 and very low CMCs. Consequently they form a lamellar bilayer phase at all the applied concentrations, temperatures and pH. In order to form nanodiscs, the lamellar phase has to be solubilized into a detergent/lipid mixed micelle phase, by the addition a detergent[37].

When detergents are added to the solution, they will distribute between the solution and the bilayer phase. It is found that the detergent concentration in solution often is proportional to the mole ratio of detergent in the bilayer [57]:

$$C_d^{aq} = \frac{n_d^b}{n_l^b}/K_r = \frac{R_b}{K_r}, \tag{2.10}$$

where n_l^b is the number of moles lipids in the bilayer and n_d^b is the number of moles detergent in the bilayer. For an ideally mixing detergent $1/K_r \approx CMC$[58] [57]. As detergent molecules are incorporated into the bilayer (increasing R_b), they change the intrinsic curvature causing a strain in the bilayer. This is associated with a disordering of the bilayer lipid tail groups, a decrease of bilayer thickness and increase in head group area per lipid[57]. The increasing strain due to the incorporation of detergent, will at some point make the bilayer configuration unfavorable and lead to a break up of the structure into smaller micellar particles.

A general model describing the solubilization of lipid bilayers by detergents is the so called "three stage model", where the lipid detergent system transitions between three pseudo phases [58][57]. In the first stage of the model, detergent molecules are incorporated into the bilayer with increasing detergent concentration according to 2.10. The incorporation of detergent continues until $R_b = R_b^{sat}$, where the bilayer is saturated and the strain makes the bilayer geometry unfavourable. The system then transitions into a lamellar-micellar phase, containing a coexisting population of bilayers and cylindrical lipid-detergent micelles. As the detergent concentration increases further, the population of cylindrical micelles increases, whereas the lamellar phase decrease. The final stage begins when $R_b = R_m^{solv}$, and the bilayer phase has been completely dissolved and only micelles are present. In this final stage, the mixed detergent-lipid micelles decrease in size as R_b increases. The whole process is sketched in figure 2.3

There are other pathways for solubilization of bilayers by detergents[57]. However, the three stage model describes the observed solubilization process of lecithin by bile salts [81][39], which in many ways can be considered similar to the solubilization of phospholipid using cholate when preparing nanodiscs.

2.2 Interactions Between Membrane Proteins and Membranes

The interactions between membrane proteins and the membrane is a large and evolving topic. Here it is only viewed in relation to the results presented in the included papers.

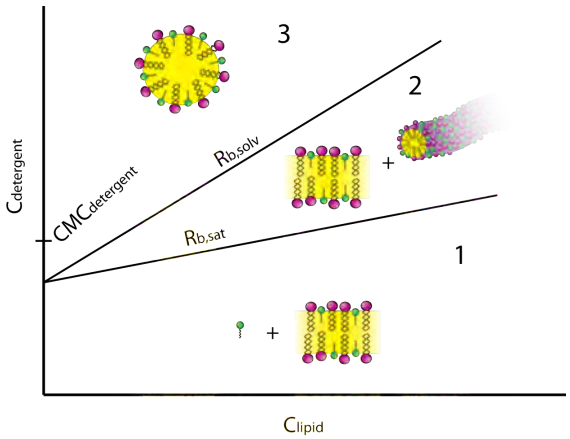


Figure 2.3: Diagram of the three stage model for solubilization of phospholipid bilayers by detergents.

Incorporation of membrane proteins into a lipid membrane may interfere with the balance between the attractive and repulsive interactions along the bilayer normal, perturbing the phospholipids close to the protein. First of all, the membrane protein may physically prevent an optimal packing of the lipid tails, simply due to its shape, and secondly residues on the protein may interact with the phospholipid head groups. Furthermore, there may be a mismatch between the hydrophobic thickness of the bilayer and the membrane protein, causing either hydrophobic residues on the protein or parts of the hydrophobic tails to be exposed to the solvent. Because the lipid and bilayer compressibility are much lower than that of proteins, this mismatch is generally compensated by a perturbation of the lipid packing as compared to the bulk lipids [5].

However it is not only the membrane proteins that affect the bilayer, the interaction goes both ways. The activity of several membrane spanning proteins has been shown to depend on the hydrophobic thickness of the membrane[5][65].

2.3 Nanodiscs and Apo-A1 HDL particles

The main goal of developing the nanodisc system is to provide an alternative to detergents for solubilizing membrane proteins in a functional state[1]. This is achieved by the phospholipid core of the nanodisc particle, which provide a lipid environment that is able to stabilize membrane proteins in solution. The nanodisc system is derived from the HDL particles found in many animals, which in its nascent state, share many features with the nanodisc system.

In the HDL particles the phospholipid core is stabilized by the protein apo-A1, whereas the nanodiscs use an engineered version of the apo-A1, called MSP. Because of the similarities of MSP and apo-A1, the apo-A1 particles will be reviewed first before discussing the nanodisc system.

2.3.1 Apo-A1 HDL Particles

High Density Lipoprotein (HDL) particles are found in human blood plasma, where they are part of the reverse cholesterol transport system transporting excess cholesterol in the body to the liver, using a hydrophobic core lipid core[80]. Because of their connection to the cholesterol transport of the body, HDLs have been studied extensively in relation to coronary artery disease, and a very large volume of literature exists on this subject (see eg. review by Wang and Briggs[133]). However, this falls outside the scope of this work. Here HDLs will be considered only in relation to the nanodisc system.

Initially, the nascent HDL particles are disc shaped particles consisting of phospholipids and the protein Apolipo Protein A1 (apo-A1)[108]. These are matured to mature HDL particle in the presence of the enzyme lecithin-cholesterol acyltransferase (LCAT), that esterifies cholesterol and transfers it into the hydrophobic lipid core of the apo-A1 HDL particle[40]. During this maturation process, the apo-A1 HDL particles change shape from disc shaped nascent HDLs, to spherical mature HDL particles[133]. The disc shaped nascent apo-A1 particles can be reconstituted *in vitro* by mixing phospholipids and apo-A1 in the appropriate ratios (often termed reconstituted HDL or rHDL). The resulting rHDL particles consists of two apo-A1 proteins wrapped around a central core of phospholipids [108, 46, 31]. Just as in the nanodisc, the small lipid sheet in the center of the reconstituted HDL particles can used as a tool for stabilizing membrane proteins in solution, enabling studies of the membrane proteins using solution techniques. This has successfully been used for activity studies of a G Protein Coupled Receptor[67] and zebra fish rhodopsin[9]. It has also been used to develop a commercial kit for *in vitro* cell free expression and purification of membrane proteins[30].

The 243 amino acid human apo-A1 consists of two distinct domains, a lipid associated domain and a so-called globular domain. See figure 2.4. The globular domain consists of four sub domains denoted G0 to G3. Deletion studies indicate that the globular domain is involved in the uptake of cholesterol and activation of LCAT, although not exclusively, as deletion of several of the lipid associated helices also has an effect on the cholesterol uptake[78]. Furthermore, the globular domain may also be involved in the stabilization of the lipid free form of apo-A1 in solution[2][83]. However, in relation to phospholipid nanodiscs the most interesting feature of the apo-A1 protein, is the ability of the lipid associated domain to stabilize phospholipids in solution.

The lipid associated domain consists of 10 helices: Two 11 amino acids helixes (Helix 3 and 9), and eight of 22 amino acids helixes (see figure 2.4). All helices are amphipatic with the hydrophobic amino acids faceing one side of the helix and the hydrophilic facing the other[55]. Several of the helices are initiated by a proline residue that is thought to introduce a kink between the helices giving the protein a curving geometry, compatible with the idea that the protein wraps around the rim of the phospholipids[24][83]. These proline residues are a highly conserved recidues of the apo-A1 protein across many species, indicating that they may play an important role in the functioning of the apo-A1[10]. By deleting the last 58 amino acids (aa 185 to 243) Mei *et al.* has successfully crystalized Δ185-243apo-A1 in a form that may be relevant for the lipid bound structure. This crystal structure nicely illustrates how the prolines are located at the kinks in the helices as well as the amphipatic nature of the helices (see figure 2.5).

Apo-A1 is water soluble but upon binding the protein undergoes a significant structural transition

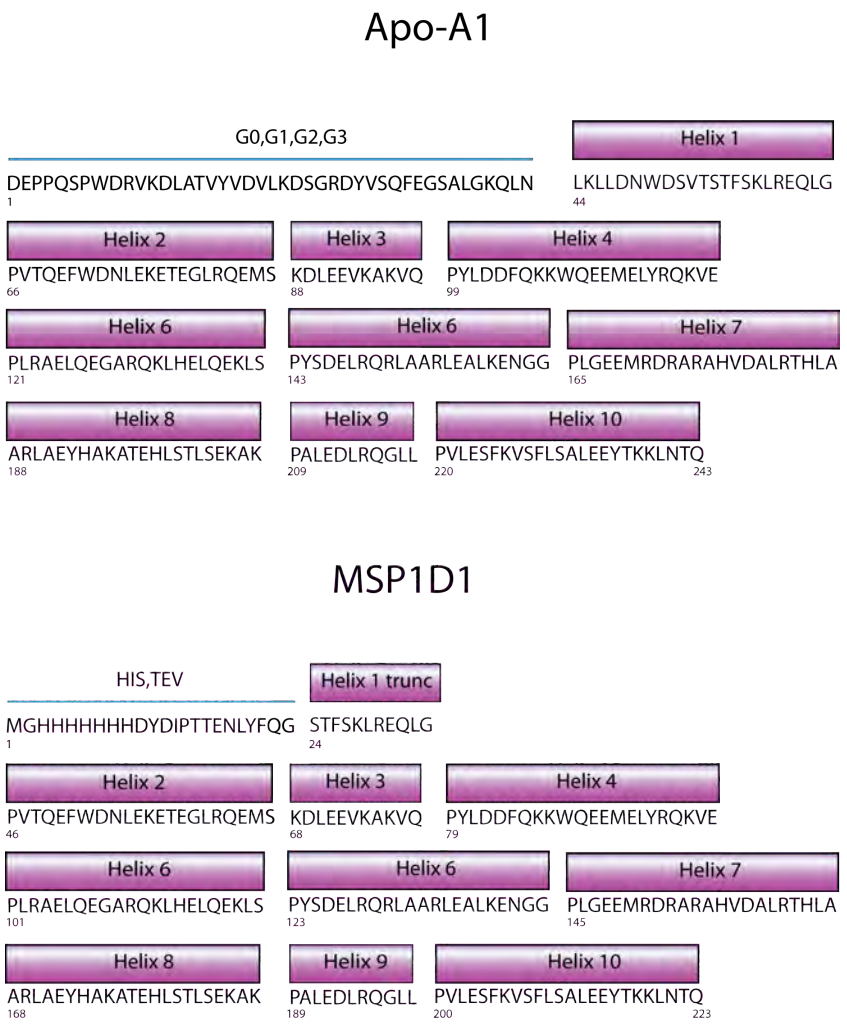


Figure 2.4: Illustration of the apo-A1 and MSP1D1 sequence and helix nomenclature.

changing from ∼50% alpha helix, in the lipid free state, to ∼75% upon binding[105].

Several models have been proposed for the structure of discoidal Apo-A1 HDLs regarding the arrangement of the two apo-A1s towards one another. In the "picket fence" model (Figure 2.6 Top)the apo-A1s are located on opposite sides of the lipid disc covering one half of the disc rim in a zig zag manner[73]. By the beginning of this project an alternative model of the apo-A1 HDL particles containing cholesterol esters was proposed by Wu *et al.*[87], based on contrast variation small-angle neutron scattering. In this model the lipids formed an ellipsoidal core with two apo-A1 proteins wrapping around the lipid core in a helix like manner. However, most work point in the direction of a "belt" model where both apo-A1s run along the whole bilayer rim in a parallel manner (Figure 2.6 Bottom)[55][72][109][125][83].

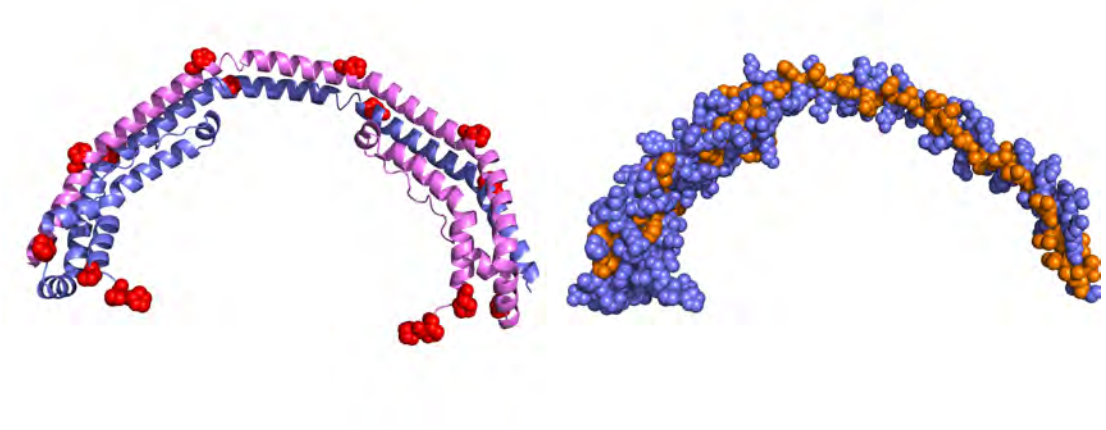


Figure 2.5: Crystal structure of Δ185-243apo-A1 (PDB: 3R2B). Left: Dimer of Δ185-243apo-A1 with prolines highlighted in red. Right: all atom model of the Δ185-243apo-A1 monomer where hydrophobic residues (ala, gly, val, ile, leu, phe and met) are highlighted in orange.

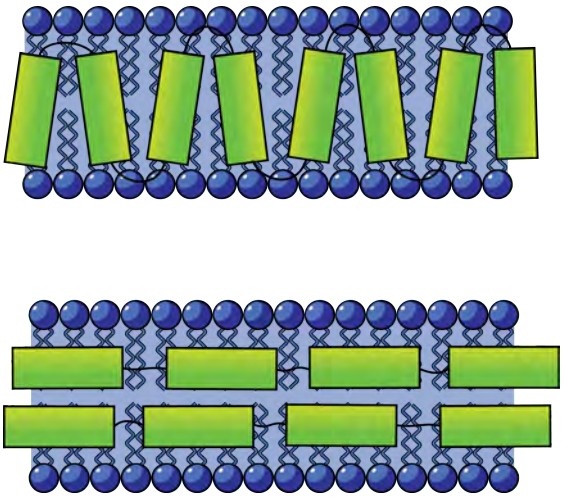


Figure 2.6: Illustration of the "picket fence" (top) and the "belt" model (bottom)for the arrangement of the apo-A1 in discoidal apo-A1 HDLs.

2.3.2 Nanodiscs

The nanodisc system is a reengineered version of the apo-A1 HDL particle developed in the Sligar lab about a decade ago[12] and optimized for forming discoidal homogenous lipoprotein particles. The optimized protein is called Membrane Scaffolding Protein (MSP). As the globular domain of the apo-A1 is not needed for forming disc shaped lippoprotein particles, this has been removed in the MSP together with the first 11 amino acids of helix 1[12]. Instead a histidine tag (HIS tag) has been added to facilitate the purification of the MSP. The HIS tag also includes a Tobacco Etch Virus (TEV) protease cleavage sequence that can be used to remove the HIS tag after the purification[37]. This construct is known as MSP1D1 and unless otherwise stated, MSP will refer to this construct. As seen from the sequence in figure 2.4 the lipid binding domains of MSP and apo-A1 are very similar and the mechanism by which the MSP stabilizes the bound phospholipids is assumed to be the same as apo-A1. However, it has not inherited the biological function with LCAT from apo-A1 and hydrogen deuterium exchange mass spectroscopy studies indicate that the solvent structure of the lipid free MSP is different from the lipid free apo-A1[86].

When reconstituted with phospholipids MSP forms nanodiscs consisting of two MSPs and 120 to 150 lipids, depending on the type of lipid[12][37]. By repeating several of the helices extended MSP variants has been developed, resulting in nanodiscs with a larger diameter, giving a precise control over the size of the self-assembled nanodisc. The nanodisc system has been used to successfully reconstitute a number of membrane proteins for solution studies[23][130][102][14][99][66], and because the size of the discs can be controlled by using MSPs of different lengths, monomers and oligomers can be reconstituted selectively[13][15].

2.3.3 Self-Assembly of Nanodiscs

The nanodisc self-assembly is a so-called directed self-assembly process[37]. It is a two step process where the phospholipids are first dissolved by a detergent into mixed lipid-detergent micelles and then the mixed micelles are transformed into phospholipid bilayer discs by incubating with MSP and subsequently removing the detergent using biobeads. In this context the amphiphatic MSP may also be considered a kind of detergent that will distribute itself between the solution and lipid-detergent micelles. During the incubation the mixed lipid-detergent micelles will bind some MSPs, which will be incorporated into the lipid-detergent micelle. The added Biobeads are small porous polymer beads that will trap detergents[103], effectively removing them from solution and in turn removeing them from the mixed micelles. The removal of the detergents changes the effective packing parameter of the detergent-lipid micelles from spherical to a planar bilayer geometry, which in turn is stabilized by the MSPs[111]. When the mixing stochiometries are right, the majority of the particles formed will be nanodiscs[12] containing 130-150 phospholipids. The formed nanodiscs are finally separated form larger aggregates and excess MSPs using size exclusion chromatography.

By constructing a partial phase diagram, Bayburt *et al.* has shown that in order to for nanodiscs to form, the detergent concentration has to be high enough, that the system is in a mixed micellar state[12] (region 3 in figure 2.3). The self-assembly process is schematically illustrated in figure 2.7. Once formed the nanodiscs are stable and do not fuse[89], in good agreement with the observation that the formed nanodiscs are relatively equal in size.

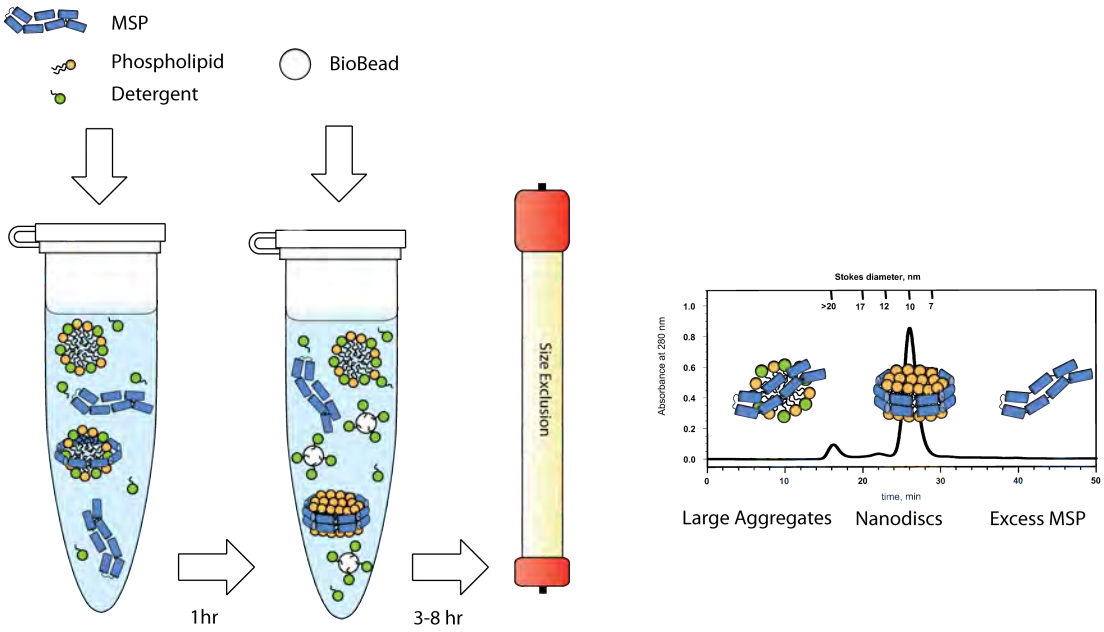


Figure 2.7: Illustration of the self-assembly process of nanodiscs. Initially lipids, detergents and MSP are mixed, the detergents are then removed using biobeads, changing the mixed micelles of lipids and detergents in to planar structures of phospholipids stabilized by the MSPs. The formed nanodiscs are then purified via size exclusion.

An interesting study of the self-assembly of the nanodisc system has been preformed by Shih *et al.* using SAXS and Molecular dynamic. Here the detergent cholate is titrated to a solution of nanodiscs in order to reproduce the cholate removal during self-assembly, although in reverse. The results confirm that the nanodiscs gradually change shape from a spherical mixed micelle, to a planar disc as the detergents are removed [111].

In paper 3 the self-assembly of the nanodisc is investigated in more detail using SAXS and the model for small-angle scattering derived in papers 1 and 2. It is found that the aggregation number of the lipids in the formed nanodiscs is determined by the mixed micelles from which the nanodisc is formed. This can be varied either by the stochiometry of lipids or by the choice of detergent. And It is also found that detergents with smaller packing parameters form discs with a higher lipid content.

2.3.4 Structural Studies of Nanodiscs

At this point there exist no published crystal structure of either nanodiscs or discoidal apo-A1 HDL particles. However a number of studies using other structural methods have been performed. AFM[12] and cryo-EM [43] confirms that nanodiscs and apo-A1 HDLs indeed form small disc shaped objects roughly 10 nm in diameter and four to five nm tall, in good agreement with the thickness of a phospholipid bilayer.

The size of about 10 nm makes the nanodiscs accessible for studies using small angle scattering studies. The first such study, was a SAXS study of apo-A1 HDL published in 1976 by Atkinson *et al.*[116]. The methodology for analysis of the data is in fact quite similar to the one used in paper 1 although more simple, and the actual model is an ellipsoid of revolution core/shell model and not a cylindrical model as the one used in this work. The analysis find the two axis of the ellipsoid to 11.0 nm and 5.5 nm in good agreement with the cryo-EM and later AFM studies. In 1979 Wlodaver *et al.*, using both cryo-EM and contrast SANS, propose a model for the structure of discoidal apo-A1 HDLs very similar to the model presented in paper 2. However this model is not evaluated against the whole scattering curve, only the Gunier region[108]. A similar model is used in a neutron scattering study by Nakano *et al.*[89], and shows good agreement with the measured scattering curve. Moluclar dynamics simulations by Catte[11] and Shih[112][111][110] also show that the nanodiscs and discoidal HDLs may form disc shaped lipid particles stabilized by the MSP or the apo-A1.

2.4 Reconstitution of Membrane Proteins into Nanodiscs

In this work the general protocol for reconstitution of membrane proteins into nanodiscs has been adapted from the work by Sligar *et al.*[8][15][38][99][22]. In this approach the detergent stabilized membrane protein is added along with the MSP and is then incorporated in to the nanodiscs during the incubation and the removal of the detergents. By varying the length of the MSP and the nanodisc to membrane protein ratio it is possible to prepare nanodiscs containing one or more membrane proteins, making it possible to study membrane proteins in different oligomeric states[21][13].

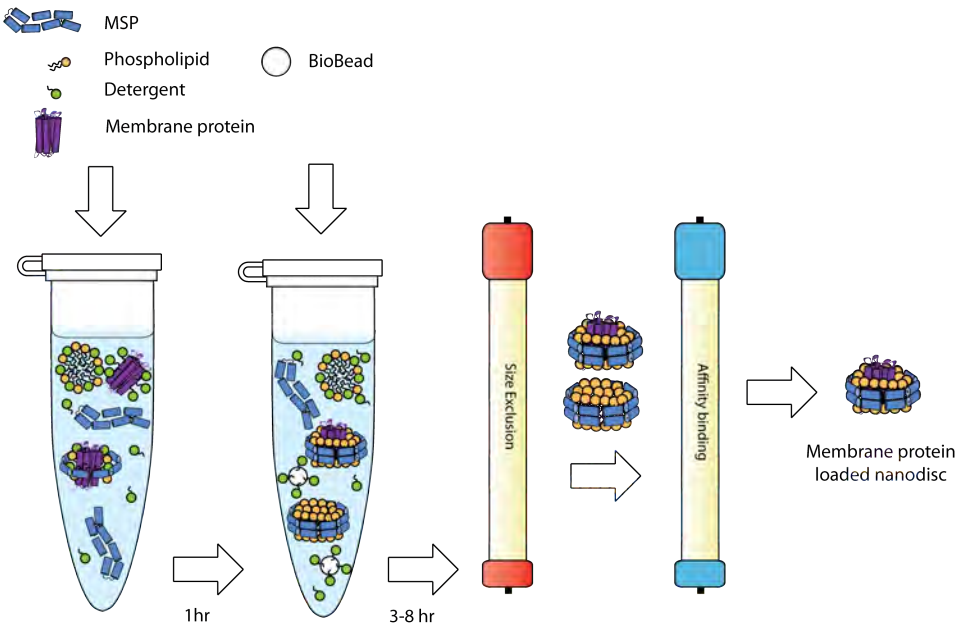


Figure 2.8: Self-assembly and purification scheme of membrane protein "loaded" nanodiscs. The detergent stabilized membrane protein is added along with the MSP in a ratio resulting in a large number of empty discs and a few containing a single membrane protein. "Empty" and "loaded" nanodiscs are purified using size exclusion, and the "loaded" nanodiscs are further separated from the empty using an affinity tag on the membrane protein.

For small-angle scattering a detailed model analysis require that the sample is highly homogenous. Therefore the protocol used in this work, aimed at obtaining a sample consisting of nanodiscs containing one incorporated membrane protein each. This is achieved by keeping the membrane protein to nanodisc ratio low during the reconstitution, thereby making it very unlikely that nanodiscs containing more than one membrane protein were formed. However this approach also produces a large excess of nanodiscs not containing any membrane protein at all, but by purifying using an affinity tag located on the membrane protein the "loaded" nanodiscs could be separated from the "empty". For example, in the case of bacteriorhodopsin in paper 4, the loaded nanodiscs were purified using streptavidin covalently linked to the bacteriorhodopsin, in order to produce a highly homogenous sample consisting of only nanodiscs containing one membrane protein each.

For other types of solution studies, the criteria for very homogenous samples can often be relaxed. For example; in activity studies it is possible to measure very low activities, and only the actual active proteins are measured. Empty nanodiscs and inactive or denatured membrane proteins are not detected.

From personal experience empty nanodiscs are stable for weeks at 5°C as judged from SAXS data, while the incorporation of a membrane protein makes the system much more prone to aggregation. This is presumably due to a small fraction of denatured, or otherwise unstable, membrane proteins, that when incorporated into the nanodiscs act as seeds for further aggregation.

Yield and Concentration

A main challenge during this project has been to obtain enough sample, of a sufficiently good quality, to conduct small angle scattering studies. Membrane proteins are in themselves often challenging to produce and purify. From personal experience, it is not uncommon with a loss of 30% to 40% of the membrane protein during the reconstitution process. However, this depends on the membrane protein.

A typical concentration of nanodiscs containing a membrane protein after purification, is about 1 mg/ml. In order to get a good scattering signal the samples has to be concentrated to 2 to 4 mg/ml depending on the instrument. These sample requirements seriously constraints the amount of sample that can practically be produced for a scattering experiment. For the three membrane proteins investigated in this thesis (Bacteriorhodopsin, Cytochrome P450 3A4 and Tissue Factor) enough sample could typically be prepared in 2 to 3 reconstitutions, yielding $\sim$50 - 100 μL. This is sufficient for a modern SAXS instrument but for a SANS experiment (on the instruments used in this work) 2 to 4 times this volume is required. This effectively limits the membrane proteins that can be studied by both SAXS and SANS to membrane proteins have a high yield of purified protein and can be incorporated effectively.

During this project, the up-concentration of the samples has only been done using spin filters. However this method build up high local concentrations near the filter membrane, increasing the probability of aggregation. An alternative to this may be to use small volume dialysis against glycerol. This is presently being further investigated in the structural biophysics group.

Transport and Storage

Freezing and thawing may also increase aggregation. Adding 15% w/v Glycerol had a positive effect on the stability of Cytochrome P450 3A4 studied in paper 5 and the tissue factor, but not on the bactheriorhodopsin studied in paper 4. Adding a size exclusion step immediately before the measurement was very successful in removing aggregation from the sample studied in paper 5. If there is enough sample, this step is recommended.

Matching of Lipids and Membrane Protein

In order to stabilize the membrane protein in the nanodisc in a native like environment, it is important to consider the hydrophobic height of the membrane protein, compared to the lipids used for the reconstitution. Either a lipid or a mixture of lipids that match both the hydrophobic height of the membrane protein and the MSPs should be used. For bacteriorhodopsin studied in paper 4 and the study of tissue factor, a mixture of PC and PG or PS lipids was found to give better sample preparations. Both of these are trans-membrane proteins with a considerable volume within the hydrophobic part of the bilayer and may therefore be particular sensitive to the surrounding lipid environment. For the p450 studied in paper 5, which only has a single alpha helical membrane anchor, good results were obtained using only one type of lipid.

The fact that the presence of the MSP perturbs the lipids may also have to be taken into account

when choosing the lipid (or lipid mixture) for reconstitution of the membrane protein.

It should be emphasized that within the UNIK project, several lipid mixtures have been investigate without any clear conclusions regarding which lipids optimal are optimal, for the reconstitution of membrane proteins.

2.4.1 Choice of Detergent for reconstitution of Membrane Proteins

In paper 3 it is shown that the choice of detergent used in the self-assembly process of the nanodisc is influencing the structure of the disc. This pose a potential problem as the membrane protein may often only be purified in one kind of detergent, and this detergent may in turn influence the self-assembly of the membrane protein containing nanodisc. The study presented in paper 3 find that n-Dodecyl-β-D-Maltoside (DDM), a popular detergent for stabilizing membrane proteins, form lipid-poor nanodiscs. However, the membrane protein to nanodisc ratio is normally only 1:10 during the reconstitution, leaving the nanodisc reconstitution detergent in excess, which may help reduce this effect.

2.5 Conclusions

The nanodisc system has since its introduction about a decade ago proven a valuable tool for studying the solution properties of membrane proteins. These studies have mostly been of a functional nature and only a few has been concerning structure. During this work it has become clear that it is not trivial to prepare a nanodisc sample containing a membrane protein, that meet the relative more strict requirements in monodispersity and purity required for structural studies using small-angle scattering. This may stem from an incomplete understanding of self-assembly process by which these particles are produced.

The basis of nanodisc self-assembly can be understood in terms of the fundamental principles of self-assembly of amphiphiles. however, the self-assembly process is "directed" by consciously pulling the particles out of equilibrium during assembly. This leads to much less predictable behavior of the particles, which again means that a combined empirical and rational approach is necessary in order to better understand the system.

Chapter 3

Small Angle Scattering

This chapter provides an overview of the key concepts and theory of small angle scattering, as well as a description of the methods and strategies for analyzing small angle scattering data from macromolecules in solution.

The chapter is divided into three parts: The first discusses the experimental setup and initial data treatment of a small angle scattering experiment. The second explains the origin and theory of scattering while the third describes the various methods and strategies for analyzing small angle scattering data.

Several text books and reviews exists on small angle scattering that have served as general references for the whole chapter. These are: *"Small-Angle Scattering of X-Rays"* by Guinier [51], *"Small Angle X-ray Scattering"* by Glatter and Kratky[49], *"Structure Analysis by Small-Angle X-Ray and Neutron Scattering"* by Svergun and Feigin[120], *"Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter"* edited by Lindner and Zemb[74] and *"Small-angle scattering studies of biological macromolecules in solution"* by Svergun and Koch[121].

In this work, both neutron and x-ray small angle scattering has been used to study nanodiscs both with and without embedded membrane proteins. Even though the cause of neutron and x-ray scattering is of a rather different nature, the theory and the interpretation of the data is essentially similar. Therefore the theory described in the following sections applies both to SAXS and SANS.

The reason for using both neutrons and x-rays in this work, is due to the fact that the different elements scatter neutrons and x-rays differently. The measurements are said to be preformed under different "contrast" situations, highlighting different parts of the sample, resulting in a more complete description of the sample.

When a neutron or an x-ray photon passes through the sample, there is a chance that it is scattered by the atoms that make up the sample. The scattered photons or neutrons will interfere with one another and the resulting scattering pattern can be recorded using a detector. For coherent scattering the resulting interference pattern will contain information about the atomic

arrangement within the sample. The main objective of the data analysis of small angle scattering, is to interpret the structural information conveyed in this interference pattern.

3.1 Experimental Setup and Initial Data Treatment

Figure 3.2 shows a typical small angle scattering set up used in this work. Initially the beam emitted from the source is collimated in order to make a highly parallel beam. After the collimation section, the beam passes through the sample where it is scattered, and the scattering is measured on a 2D position sensitive detector. For non-oriented samples the scattering pattern is centrosymmetric, and it can be radially averaged yielding scattering intensity as a function of scattering angle 2θ. The radially averaged detector signal is shown in the insert in figure 3.2. The scattering angle is dependent on the wavelength of the scattered radiation and is therefore converted to the wavelength independent scattering vector $q = 4\pi sin(\theta)/\lambda$.

In order to get a good angular resolution on the detector, it is placed several meters away from the sample. How far, depends on the detector in use and the size of the particles studied in the experiment. In order to minimize absorption and background scattering, the whole beam path, all the way to the detector is placed in vacuum. All SAXS experiments presented in this work were preformed on instruments with a fixed sample detector distance whereas the SANS experiments were conducted at several sample-detector distances as discussed in the next section. However many SAXS instruments also have the option to change the sample-detector distance depending on the q range desired.

A beamstop is placed in front of the detector, in order to protect the detector from the direct beam. The size of the beam and the beamstop limits the minimum scattering angle measurable in the experiment.

In general, the SAXS and SANS experimental setups are very similar. However the neutron flux is generally lower at a SANS than the x-ray flux at a synchrotron SAXS instrument. As a consequence the beam cross section is larger, which require a larger sample. In order to gain additional flux, the wavelength distribution of the neutron beam is often widened. Furthermore, the resolution of a neutron detector is considerably lower than a modern x-ray detector. Therefore measurements are typically done at several sample detector distances, by moving the detector inside the detector tank, in order to obtain a good q resolution in the whole desired q-range. This is illustrated in figure 3.1.

In addition to coherent scattering, many elements can also scatter neutrons incoherently, giving rise to flat incoherent background that does not carry structural information. For biological samples the dominating source of incoherent scattering is hydrogen, with an incoherent scattering cross section of $\sigma_{H,inc} = 80.26 \cdot 10^{-24}$ cm^2 compared to $< 1 \cdot 10^{-24}$ cm^2 for the other biological relevant elements[64]. Deuterium on the other hand only has $\sigma_{D,inc} = 2.05 \cdot 10^{-24}$ cm^2, by changing the solvent from H_2O to D_2O a large source of the incoherent scattering can be removed, improving the signal to noise ratio. Therefor all SANS experiments performed in this work was done in D_2O based buffers.

At so called bioSAXS setups used in this work, the sample environment was optimized for

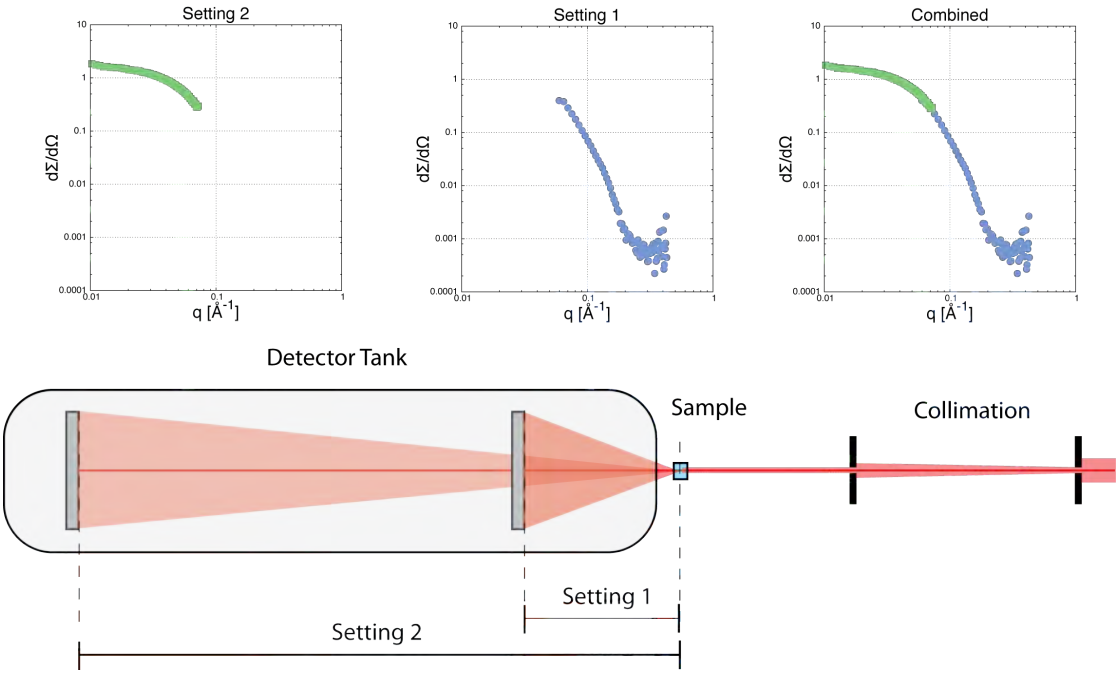


Figure 3.1: Sketch of SANS setup: In order to cover the desired length scale several sample detector distances are often used, this is achieved by moving the detector inside the detector tank. The measurements, measured at different detector distances, are combined to yield the full scattering curve.

measuring macro molecules in solution, consisting of a temperature controlled flow cell with a capillary going through the vacuum. This make it possible to load liquid samples without breaking the vacuum and reestablishing it each time the sample is changed. In addition, several of the SAXS beamlines had automatic sample changer robot, for automatic loading of the samples and subsequently cleaning of the capillary. Due to the high photon flux a synchrotron, the measurement times at synchrotron bioSAXS beamlines are in the order of seconds, and the sample volume is 20μl - 30 μl at a minimum $1-2$ mg/ml sample concentration.

For the SANS experiments performed in this work, the samples were put in cuvettes that were exposed to the neutron beam using a motorized rack. The neutron beam diameter was typically between 8 mm and 16 mm, and the sample thickness varied between 1 mm and 5 mm depending on the different cuvettes. This required sample volumes from 100 μl up to 2-3 ml. Depending on the neutron source and the contrast situation, the measurement times ranged from several minutes to several hours.

3.1.1 Initial Data Treatment

From a sample of particles in solution the intensity scattered neutrons or x-rays on the detector in the direction q can be written as:

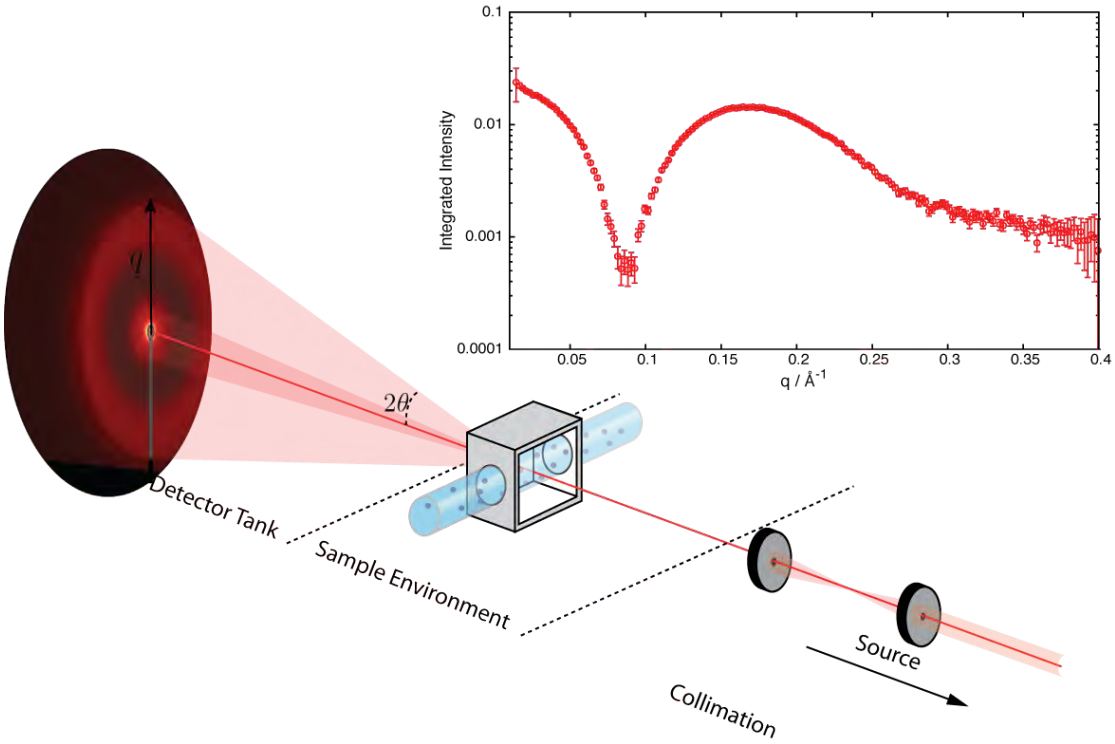


Figure 3.2: Sketch of a typical solution small angle scattering experiment. Insert shows the radially averaged scattering pattern recorded by the detector.

$$I(\mathbf{q})_{sample} = I_0 t A \epsilon(\mathbf{q}) T_{cuv} T_{sample} d_{sample} \frac{d\Sigma}{d\Omega}(\mathbf{q})_{sample} \Delta\Omega + I_{solvent} + I_{cuv} + I_{instrument} \quad (3.1)$$

Where I_0 is the incident flux of neutrons or photons on the sample, t the measurement time, A the cross section of the beam, $\epsilon(\mathbf{q})$ the efficiency of the detector pixel in the direction $\mathbf{q}$, T_{cuv} and T_{sample} are the transmissions of the sample cuvette and the sample respectively, d_{sample} the thickness of the sample and $\Delta\Omega$ is the solid angle spanned by the detector pixel in the direction $\mathbf{q}$. $\frac{d\Sigma}{d\Omega}(\mathbf{q})_{sample}$ is the so-called macroscopic differential scattering cross section and describes how the particles scatter in the direction $\mathbf{q}$ per unit solid angle and unit sample volume. For x-rays, the macroscopic differential scattering cross section is determined by the distribution of electrons in the sample whereas for neutrons it is determined by the isotopic distribution within the samples.

In addition to the scattering from the particles themselves there is a scattering contribution from the solvent, the cuvette and an experimental background. By measuring the scattering from a sample consisting of the only the solvent, the $I_{solvent} + I_{cuv}$ contribution can be determined and by preforming a measurement with a 100% absorbing sample $I_{instrument}$ may be determined. Normalizing for the transmission measurement time and incident flux, these background signals may be subtracted, and the normalized sample scattering can be found:

$$I(\mathbf{q})_{sample-bg} = \left(\frac{I_{sample} - I_{instrument}}{AI_0 t T_{sample}} - \frac{I_{bg} - I_{instrument}}{A_{bg} I_{0,bg} t_{bg} T_{bg}} \right) \tag{3.2}$$

Where I_{bg} denote the scattering from the solvent and the cuvette. A_{bg}, $I_{0,bg}$, t_{bg} and T_{bg} are the beam cross section, flux, measurement time and transmission of the solvent measurement.

The macroscopic differential scattering cross section of sample is then related to the background subtracted intensity by:

$$I(\mathbf{q})_{sample-bg} = \epsilon(\mathbf{q}) T_{cuv} d_{sample} \frac{d\Sigma}{d\Omega}(\mathbf{q})_{sample} \Delta\Omega \tag{3.3}$$

3.1.2 Absolute Calibration

The macroscopic differential scattering cross section has the units of reciprocal length and it is traditionally reported in inverse centimeters. By measuring the scattering on an absolute scale, it is possible to quantitatively measure the macroscopic differential scattering cross section of the studied particles. Scattering data measured on an absolute scale will therefore, in addition to the structural information conveyed by the shape of the scattering curve, also contain information on the amount of scattering. This extra information can be used for molecular mass determination of proteins[88], and as constraints in a more detailed model based analysis. As the macroscopic differential scattering cross section of the particles is independent of the instrument, absolute scaling can be used when comparing measurements performed on different instruments.

For the work presented here the calibration of $I(\mathbf{q})_{sample-bg}$ to instrument independent absolute units, was done using a secondary standard with a known differential scattering length density. In the case of SAXS either a protein standard or H_2O was used, and in the case of SANS H_2O was used.

SANS Water is not necessarily a good choice of standard for absolute calibration of SANS experiments as gives a large multiple incoherent and inelastic contribution to the scattering signal[47]. Nonetheless, water was offered as the standard for calibration at the beamlines used in this work.

For a thin water sample the background subtracted intensity (in this case the background consists of the scattering from the cuvette and the instrumental noise) from water can be approximated by:

$$I_{H_2O-bg}(\mathbf{q}) = \epsilon(\mathbf{q}) T_{cuv} \frac{(1 - T_{H_2O})}{4\pi} G \Delta\Omega \tag{3.4}$$

where T_{H_2O} is the transmission of the water sample. The factor G, is an empirical correction factor for multiple incoherent inelastic scattering by water and is dependent on the wavelength[75].

By dividing 3.3 by 3.4 $\frac{d\Sigma}{d\Omega}_{sample}$ can be found:

$$\frac{d\Sigma}{d\Omega}_{sample} = \frac{I_s \, G \cdot (1 - T_{H_2O})}{I_{H_2O} \; d_{sample} 4\pi} \tag{3.5}$$

For the work presented here, the correct G factor was incorporated in the data reduction software at the beam line. The division is done pixel by pixel as this corrects for any difference in the detection efficiency between the pixels on the detector.

SAXS The macroscopic differential scattering cross section of a globular protein sample of known molecular weight and concentration can to a reasonable accuracy be calculated by[88]:

$$\frac{d\Sigma}{d\Omega}(q=0)_{prot} = \frac{M_w \Delta\rho_M^2 c}{N_A} \tag{3.6}$$

where M_w is the molecular weight of the protein, c is the concentration in mg/ml, $\Delta\rho_m$ the excess scattering length per gram, and N_A Avogadros number. For globular proteins in water $\Delta\rho_m$ has an average value of $2 \cdot 10^{10}$ cm/g.

A measurement of a known protein sample can then be used to calibrate the measurements to absolute scale.

$$K\frac{I(q=0)_{prot-bg}}{\epsilon(\mathbf{q})T_{cuv}d_{sample}\Delta\Omega} = \frac{d\Sigma}{d\Omega}(q=0)_{prot} = \frac{M_w \Delta\rho_M^2 c}{N_A} \tag{3.7}$$

where K is the proportionality constant converting $I(0)_{prot-bg}$ to absolute units.

A useful list of protein standards is reported by *Mylonas et al.*[88]. The absolute scale calibration using this method is generally accurate within 10 %. An important requirement for absolute calibration using protein standards is that the measurement is preformed in the dilute regime where the proteins can be considered ideal particles and that proteins are not aggregating, as this will affect the forward scattering.

An alternative to using a protein standard is to use water. As described by Orthaber *et al.* water produces a relative flat scattering spectrum in the small angle regime that can be used for absolute scale calibration[17]. The macroscopic differential scattering cross section of a water sample can be related to the isothermal compressibility by:

$$\frac{d\Sigma}{d\Omega}(0,T)_{water} = \rho_{water}^2 k_b T \chi(T)_t, \tag{3.8}$$

where $\chi(T)_t$ is the isothermal compressibility as a function of the temperature T. A table of the isothermal compressibility of water can be found in *CRC Handbook of Chemistry and Physics, 93re Edition*[56].

3.1.3 Contrast Variation

Because for the difference between deuterium and hydrogen in neutron scattering, varying the amount of H_2O in the solvent will change the contrast of the sample. For a sample with several contrasts, this can be used to increase the amount of information available for the analysis, as the relative contrasts between the regions change and thus change the scattering pattern. In particular, in a 42% D_2O/58% H_2O based solvent proteins have close to zero contrast[64]. For example for protein-DNA or protein-lipid complexes, this gives an opportunity to study the

whole assembly without seeing the protein. Using deuterated phospholipids it is also possible to study membrane proteins incorporated into phospholipid liposomes using SANS, as the signal form the lipids is greatly decreased[41].

Alternatively proteins can be expressed under high deuterium conditions giving them a different contrast than proteins grown under normal conditions as deuterium is incorporated instead of hydrogen. This can be used to highlight proteins that are part of a larger protein complex[20, 33].

A recent effort within the UNIK project has been to produce contrast matched nanodiscs, where both the MSPs and the phospholipids are deuterated such that they have a minimal contrast in D_2O based solvents. In these disc an incorporated membrane protein will dominate the scattering signal limiting or completely removing the need of a detailed model of the nanodisc, in order to obtain structural information on the membrane protein[79].

Contrast variation can improve the information on a sample available for analysis, but by adding hydrogen to the solvent, the incoherent scattering increases and for weakly scattering samples, the resulting lower signal to noise ratio may wash out the structural signal.

3.1.4 Instrumental Smearing

As mentioned above the monochromatization of the incoming neutrons is in practice relaxed in order to increase the neutron flux on the sample. For all SANS experiment done in this work a wavelength distribution of $\Delta\lambda/\lambda = 10\%$ at FWHM has been used.

As the scattering angle is related to the wavelength, a beam of neutrons with a wavelength distribution, will produce a smeared scattering pattern[48]. In addition, the finite collimation will also introduce a spread in scattering angle, in particular near the beam stop. These two effects can be taken into account by calculating a resolution function, using the wavelength distribution, the size of the beam, the size of the pinholes and the collimation length[115]. The calculated resolution function can then be applied to any model calculations in order to simulate the instrumental smearing present in the measured data. As the instrument geometry is changed for each sample detector distance, data sets acquired in different setting will have to be smeared with the corresponding resolution function.

For the nanodisc data, the effect of smearing was minimal due to the scattering curve having relatively few features, and this was therefore omitted in order to gain computational speed.

Instrumental smearing is in principle also an issue for SAXS. However, for synchotron sources the x-ray beam is in practice monochromatic with a very little divergence, effectively making smearing of the measured data minimal.

3.2 Combination of SAXS and SANS

SAXS and SANS data both provide structural information, but in different contrast situations. For samples where different contrasts are present, such as the nanodisc system, this can be

exploited to enable a more detailed model-based analysis.

One potential obstacle is that, the volume and concentration requirements for SANS are bigger than for SAXS. This may pose a problem for samples that are difficult to obtain in large volumes. In this work obtaining enough sample of nanodiscs with an incorporated membrane protein in order to perform both SAXS and SANS measurements was an issue, that was resolved by using a membrane protein with a highly optimized expression protocol.

Because of the long measurement times for SANS many more samples can be measured at 24 hrs of SAXS beam time then 24 hrs of SANS beamtime. In paper 1 it was only possible to perform neutron scattering at 20 C° whereas a whole temperature scan was done during the SAXS experiment. However, the SANS data were key to determining the aggregation number of lipids in the nanodiscs due to its sensitivity to the hydrogen rich lipids. The number determined from the analysis of the combined SANS and SAXS data, could then be used in the analysis of the SAXS data from the temperature scan.

3.3 General Scattering Theory

3.3.1 Scattering from One Atom

Scattering Length and Scattering Cross Section

Scattering of x-ray photons from an atom is caused by the interaction between the electric field of the photon and the electric field of the electrons of that atom. Scattering of a neutron is due to an interaction between the neutron and the atomic nucleus[3]. This implies that isotopes of the same atom may scatter neutrons differently, while x-rays are only sensitive to the number of electrons in the atom, and hence not isotope sensitive. In other words, x-rays probe the electron distribution in the sample, whereas the neutrons probe the isotopic distribution in the sample.

The amplitude of a scattered photon or neutron is proportional to the so called scattering length, b, which describes the scattering potency of the particular atom:

$$E_{scat}(\theta) = \frac{E_{in} \cdot b}{R} \sqrt{\frac{1 - cos^2(2\theta)}{2}}, \tag{3.9}$$

where E_{scat} is the amplitude of the scattered photons at a distance R and E_{in} is the amplitude of the incident photons or neutrons. As the units of the amplitudes cancel out, b has the unit of length, hence the name scattering length. For historically reasons it is usually given in cm. The scattering lengths of common biological elements for neutrons and x-rays are listed in table 3.3.1. For small angles the square root term approaches unity, and the scattered amplitude can be considered independent of the scattering angle. For x-ray scattering at large angles the scattered amplitude will also be dependent on the angle due to the electronic structure of the atom, and the so called atomic form factor will have to be included in eq 3.9.

The x-ray scattering length of an atom is proportional to the number of electrons in the atom,

Element	b (x-rays) / cm	b (neutrons) / cm
^{1}H	$2.82 \cdot 10^{-13}$	$-3.739 \cdot 10^{-13}$
^{2}H	$2.82 \cdot 10^{-13}$	$6.671 \cdot 10^{-13}$
C	$16.92 \cdot 10^{-13}$	$6.646 \cdot 10^{-13}$
N	$19.74 \cdot 10^{-13}$	$9.36 \cdot 10^{-13}$
O	$22.56 \cdot 10^{-13}$	$5.803 \cdot 10^{-13}$
P	$42.30 \cdot 10^{-13}$	$5.13 \cdot 10^{-13}$
S	$45.12 \cdot 10^{-13}$	$2.847 \cdot 10^{-13}$

Table 3.1: Scattering lengths of elements common in biological samples. The scattering length for x-rays is proportional to the number of electrons whereas the neutron scattering length is isotope specific. Neutron scattering lengths are reproduced from *Sears 1993* [127].

whereas there is no such simple dependence for the neutron scattering length. In practice the neutron scattering lengths have been obtained empirically. One of the most important consequences of differences between x-ray and neutron scattering is the possibility of observing hydrogen by neutron scattering, as ^{1}H has a very different scattering length. This is normally not possible in x-ray scattering, due to the weak signal from hydrogen compared to the heavier atoms.

In a scattering experiment, it is not the amplitudes but the intensity of scattered x-rays or neutrons, I_{scat}, that is measured. The intensity on a detector is the absolute square of the scattered field times the area of the detector: $I_{scat} = |E_{scat}|^2 \cdot A_{detector}$, and the incoming flux is proportional to the square of the incoming amplitude $I_0 = |E_{in}|^2$. The ratio of scattered to incoming flux of neutrons or photons is:

$$\frac{I_{scat}}{I_0} = \frac{|E_{scat}|^2 \cdot A_{detector}}{|E_{in}|^2} = \frac{|b|^2 A_{detector}}{R^2} = |b|^2 \Delta\Omega \tag{3.10}$$

$$\frac{I_{scat}}{I_0 \Delta\Omega} = |b|^2 \tag{3.11}$$

where $\Delta\Omega$ is the solid angle spanned by the detector at a distance R. This gives the relative amount of scattered x-rays or neutrons by the sample in the direction of the detector. In order to obtain the total scattering into all directions, we will have to measure all solid angles:

$$\int_\Omega |b|^2 d\Omega = 4\pi |b|^2 = \sigma \tag{3.12}$$

$$\frac{d\sigma}{d\Omega} = |b|^2. \tag{3.13}$$

σ is known as the scattering cross section, and $\frac{d\sigma}{d\Omega}$ as the differential scattering cross section. It is often useful to express the scattering from a sample as scattering cross section per unit volume, yielding the macroscopic scattering cross section $\frac{d\Sigma}{d\Omega}$. Traditionally reported in cm^{-1}.

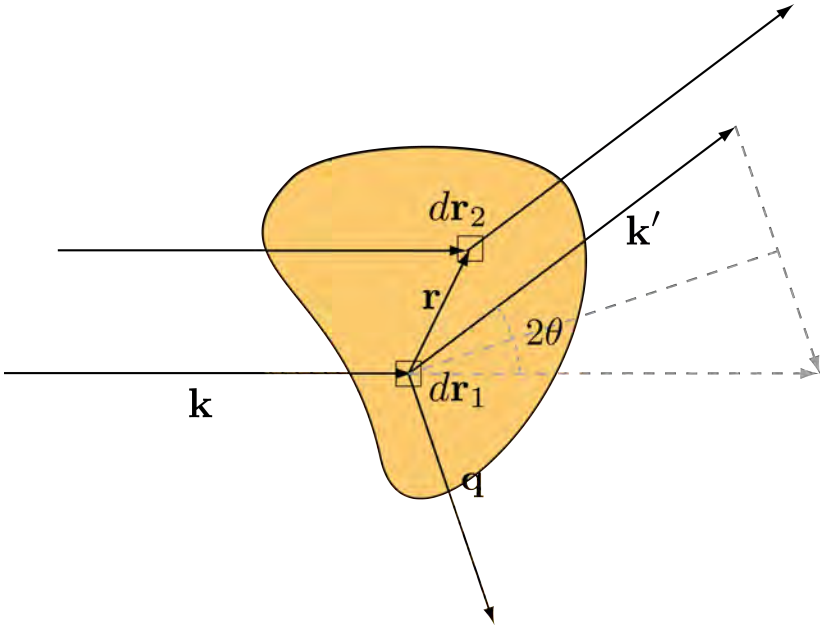


Figure 3.3: Sketch of a small angle scattering experiment. The wave vector of the incoming beam, $\mathbf{k}$, is scattered by the two volume elements $d\mathbf{r_1}$ and $d\mathbf{r_2}$, and the scattered beam is described by the wave vector $\mathbf{k}'$. The scattering vector $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ is connected to the scattering angle 2θ by $|\mathbf{q}| = q = 4\pi sin(\theta)/\lambda$

3.3.2 Scattering from Many Atoms

In small angle scattering, it is not possible to resolve the individual atoms of a sample. Instead of describing the sample as built up by atoms with an associated scattering length, it is described as made up by small volume elements with an associated scattering length density.

Figure 3.3 shows two volume elements, $d\mathbf{r_1}$ and $d\mathbf{r_2}$, in a particle separated by the distance $\mathbf{r}$. The incoming wave is denoted by its wave vector, $\mathbf{k}$, and the scattered wave by $\mathbf{k}'$. Due to the distance between the two volume elements, the scattered wave from the second volume element will have a phase shift of $\mathbf{k} \cdot \mathbf{r} - \mathbf{k}' \cdot \mathbf{r} = \mathbf{r} \cdot \mathbf{q}$ relative to the first. For elastic scattering we can assume $|\mathbf{k}| = |\mathbf{k}'|$. $\mathbf{q}$ is the momentum transfer to the photon or neutron and is related to the scattering angle, 2θ, by $q = 4\pi sin(\theta)/\lambda$.

The combined scattered amplitude from the two volume elements is found by summing the scattering amplitudes: $A(\mathbf{q}) = b_1 + b_2 e^{-i\mathbf{r}\cdot\mathbf{q}} = d\mathbf{v_1}\rho(\mathbf{r_1}) + d\mathbf{v_2}\rho(\mathbf{r_2})e^{-i\mathbf{r}\cdot\mathbf{q}}$, where $\rho(\mathbf{r_1})$ and $\rho(\mathbf{r_2})$ are the scattering length densities of the two volume elements.

To get the total scattered amplitude from the sample, the contributions from all volume elements have to be taken into account by integrating over the total volume:

$$A(\mathbf{q}) = \int_{\mathbf{v}} \rho(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.14}$$

The equation above identifies the scattered amplitude as the Fourier transform of the scattering

length density of the scattering volume. This means that the structural information of the sample is transformed into the scattering pattern as the Fourier transform of the scattering length density of the sample.

The fact that the measured scattering pattern is the Fourier transform of the scattering length density has some quite profound consequences for the behavior of the scattering pattern. First of all, there is a reciprocity between the distance vector, $\mathbf{r}$, and the scattering vector, $\mathbf{q}$, so that large distances in the sample will be described by short $\mathbf{q}$-vectors in the scattering pattern. Secondly, periodic order in the sample will result in peaks in the scattering pattern.

The measured intensity is the absolute square of the amplitudes:

$$\begin{aligned} I(\mathbf{q}) = A(\mathbf{q})A(\mathbf{q})^* &= \int_{\mathbf{v}} \rho(\mathbf{r_1}) e^{-i\mathbf{q}\cdot\mathbf{r_1}} d\mathbf{r_1} \int_{\mathbf{v}} \rho(\mathbf{r_2}) e^{i\mathbf{q}\cdot\mathbf{r_2}} d\mathbf{r_2} \\ &= \int_{\mathbf{v}}\int_{\mathbf{v}} \rho(\mathbf{r_1})\rho(\mathbf{r_2}) e^{-i\mathbf{q}(\mathbf{r_1}-\mathbf{r_2})} d\mathbf{r_2} d\mathbf{r_1} \\ &= \int_{\mathbf{v}}\int_{\mathbf{v}} \rho(\mathbf{r_1})\rho(\mathbf{r_1}-\mathbf{r}) d\mathbf{r_1} e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \\ &= \int_{\mathbf{v}} \gamma(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \end{aligned} \tag{3.15}$$

where the correlation function $\gamma(\mathbf{r}) = \int_{\mathbf{v}} \rho(\mathbf{r_1})\rho(\mathbf{r_1}-\mathbf{r}) d\mathbf{r_1}$. $\gamma(\mathbf{r})$ is proportional to the frequency of a given distance vector, $\mathbf{r}$, in the sample, weighted by the scattering length density of the volume elements at each end.

3.3.3 Contrast

The measured scattering intensity from a sample consisting of particles in solution will consist of scattering from the particles as well as the solvent. By subtracting a background measurement of the solvent, the scattering originating from the particles can be obtained. Assuming that the solvent has a homogenous scattering length density, the scattering from the particles will be seen relative to the solvent. The scatted amplitude in 3.14 now becomes:

$$A(\mathbf{q})_{particle} = \int_{\mathbf{v}} \Delta\rho(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}, \tag{3.16}$$

where $\Delta\rho$ is the excess scattering length density of the particles compared to the solvent,

$$\Delta\rho = \rho_{particle} - \rho_{solvent}. \tag{3.17}$$

The shape of the scattering curve is determined only by the scattering length density deviations from the solvent. If the solvent and the sample have the exact same scattering length density, $\Delta\rho$ vanishes and there will be no scattering from the particles (except at $q = 0$). Therefore, the quantity $\Delta\rho$ is often referred to as the contrast of the sample.

	$\Delta\rho$ x-rays, in H_2O / cm/Å^3	$\Delta\rho$ neutrons, in D_2O / cm/Å^3
Protein	$2.83{\cdot}10^{-14}$	$-3.38{\cdot}10^{-14}$
POPC	$1.06{\cdot}10^{-15}$	$-6.03{\cdot}10^{-14}$
PC	$5.10{\cdot}10^{-14}$	$-4.17{\cdot}10^{-14}$
PO	$-1.61{\cdot}10^{-14}$	$-6.67{\cdot}10^{-14}$

Table 3.2: Scatterings lengths of protein and the phospholipid POPC for neutrons and x-rays. For neutrons the head (PC) and tail (PO) part of the phospholipid have more or less the same contrast whereas there is a considerable difference for x-rays.

As mentioned previously, the scattering length density of the sample buffer in SANS may be tuned by the amount of D_2O and H_2O in the buffer[64]. The use of contrast variation can be used to emphasize regions of different scattering length density in multi contrast samples. Table 3.2 lists the x-ray and neutron contrasts for proteins and the phospholipid POPC.

In most globular proteins, the amino acids packs in more or less the same way, giving them the same excess scattering length density [88]. Due to the limited resolution of small angle scattering, the internal structure of proteins can often not be resolved. Consequently, in SAXS and SANS proteins appear as having a homogenous contrast.

For phospholipids, the picture is a bit more complicated. The electron-rich head-groups have a positive contrast for x-rays, whereas the hydrocarbon tail groups are relatively electron-poor, compared to water, and therefore have a negative contrast for x-rays. For neutrons on the other hand, phospholipids have a relatively homogenous contrast. This implies that different parts of the molecular structure are highlighted by x-rays and neutrons. X-ray scattering will give information on the arrangement of the tail and head groups, whereas neutron scattering will give information on the whole molecule[69].

3.3.4 Solution Small Angle Scattering

Macromolecules in solution are free to move and rotate in the buffer. Assuming that the measurement represent the time and space average of the sample, the scattering from all particles can be written as the rotational average of equation 3.15:

$$I(q) = \frac{\int_0^{2\pi}\int_0^{\pi} I(\mathbf{q})sin(\theta)d\theta d\phi}{\int_0^{2\pi}\int_0^{\pi} d\theta d\phi} = \langle I(\mathbf{q})\rangle_{\mathbf{\Omega}} = \left\langle \int_{\mathbf{v}} \gamma(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r} \right\rangle_{\Omega}$$
$$= 4\pi \int \gamma(r)r^2 \frac{sin(qr)}{qr}dr = 4\pi \int P(r)\frac{sin(qr)}{qr}dr \tag{3.18}$$

Where $\langle\rangle_\Omega$ denotes the orientational average and $P(r) = \gamma(r)r^2$.
Assuming the particles are not interacting, the scattering from the sample becomes the sum of the scattering from each particle:

$$I(q) = n \cdot I(q)_{particle} = n\int_0^{D_{max}} P(r)\frac{sin(qr)}{qr}dr, \tag{3.19}$$

where n is the number of particles, and D_{max} is the maximum length present within the particle. $P(r)$ is the so called pair distance distribution function (PDDF) and contains the structural information about the sample. Using the Indirect Fourier Transform method[48], $P(r)$ is easily calculated from the scattering data and it is often an important part of the initial data analysis. For samples with a homogenous contrast, such as proteins, the $P(r)$ is related to the shape of the sample. However, for samples with an in-homogenous contrast, $P(r)$ may take negative values due to $\gamma(r)$ being weighted by the excess scattering length density. An example of this is the POPC nanodiscs, where the negative contrast in the tail groups give rise to negative values in $P(r)$. Ultimately this means that it is non trivial to relate $P(r)$ to a certain shape, in the case of a particle with an in-homogenous contrast.

$P(r)$ is one dimensional because of the orientational average. This also means that the measured scattering pattern, is only a one dimensional projection of the actual 3d structure of the studied particle. The orientational average, therefore, represents a substantial loss of structural information. It also makes it impossible uniquely reconstruct the 3d structure by simply performing a reverse Fourier transform of the scattering data. This problem is some times known as the "inverse scattering problem". As will be discussed in the next sections, not all structural information is lost, and in addition it is often possible to make up for some of the lost information by assuming some kind of model, based on information on the sample from other sources.

3.3.5 Radius of Gyration

Using $P(r)$, a radius of gyration of the scattering particle can be defined as:

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr} \tag{3.20}$$

For proteins and other particles with homogenous contrast, the radius of gyration will give a measure of the size of the particle. For example the radius of gyration of a sphere is: $R_g = \sqrt{\frac{3}{5}} r$, where r is the radius of the sphere. However, for samples with several contrasts, in particular positive and negative contrasts, the radius of gyration is not easily related to an actual size.

3.3.6 Form Factor and Structure Factor

It is often useful to separate the scattering signal into a part originating from the intra-particle structure and a part due to inter-particle structure. The inter-particle structure arise from particle-particle interactions and is concentration dependent. At low concentrations the inter-particle contribution to the scattering becomes negligible.

At zero scattering angle, there is no phase difference between the scattering originating from each of the volume elements making up the particle and therefore no interference occurs. Using equation 3.15, the scattering intensity from a background subtracted non-interacting monodisperse sample becomes:

$$I(0) = n \cdot \int_{\mathbf{v}} \int_{\mathbf{v}} \Delta\rho(\mathbf{r}) \Delta\rho(\mathbf{r_1} - \mathbf{r}) d\mathbf{r_1} d\mathbf{r} = n \cdot \Delta b^2$$

$$= n \cdot (\langle \Delta\rho \rangle \, v)^2 \tag{3.21}$$

where n is the number of particles and Δb is the excess scattering length of the scattering particle compared to the solvent. $\langle \Delta\rho \rangle$ is the average excess scattering length density of the particle and v is the volume of the particle.

As the scattering angle changes, the scattered waves will start to interfere and the scattering intensity start to change in a q-dependent manner:

$$I(q) = n \left\langle \Delta\rho \right\rangle^2 v^2 F(q). \tag{3.22}$$

$F(q)$ is known as the form factor and describes the interference due to the shape of the sample.

For monodisperse particles of spherical symmetry interacting with a spherical symetric potential 3.22 becomes:

$$I(q) = n \left\langle \Delta\rho \right\rangle^2 v^2 F(q) S(q), \tag{3.23}$$

where $S(q)$ is the structure factor. As mentioned above, the structure factor may be considered equal to unity at low sample concentrations.

Anisotropic Interacting Particles

For anisotropic interacting particles, the calculation of the scattered intensity becomes very complicated as it depends on the shape and orientation of every single particle as well at the interaction between all particles. In the general case the scattering can be written as[93]:

$$I(q) = \left\langle \Delta\rho \right\rangle^2 v^2 \left(\sum_i |A_i(\mathbf{q}, \mathbf{e_i})|^2 + \frac{1}{n} \sum_{i,j} A_i(q, \mathbf{e_i}) A_j(q, \mathbf{e_j}) \left(S_{i,j}(q, \mathbf{e_i}, \mathbf{e_j}) - 1 \right) \right), \tag{3.24}$$

where $A_i(q, \mathbf{e_i})$ is the form factor amplitude of the i'th particle with the orientation given by the unit vector $\mathbf{e_i}$, N is the number of particles, and $S_{i,j}(q, \mathbf{e_i}, \mathbf{e_j})$ is the structure factor between the i'th and j'th particles orientations defined by $\mathbf{e_i}$ and $\mathbf{e_j}$. The amplitude form factor is defined as $F(q) = |A(q)|^2$.

For weakly anisotropic interacting particles equation 3.24 can be simplifyed using the decoupling approximation[68]:

$$I(q) = n \left\langle \Delta\rho \right\rangle^2 v^2 F(q) \left(1 + \beta(q)(S(q) - 1) \right) \tag{3.25}$$

where $\beta(q) = \left\langle F(q, \mathbf{e}) \right\rangle_\Omega^2 / \left\langle F(q, \mathbf{e})^2 \right\rangle_\Omega$. $F(q, \mathbf{e})$ is the form factor of the particle oriented with the unit vector $\mathbf{e}$ and $\langle \rangle_\Omega$ denotes the orientational average.

Polydisperse Particles

In the discussion above, the particles studied have all been assumed to be monodisperse (except for equation 3.24). This may not always be the case. For non-interacting polydisperse particles the scattering is the a sum of the scattering from each kind of sample:

$$I(q) = n_1 \left\langle \Delta\rho_1 \right\rangle^2 v_1^2 F_1(q) + n_2 \left\langle \Delta\rho_2 \right\rangle^2 v_2^2 F_2(q) ... \tag{3.26}$$

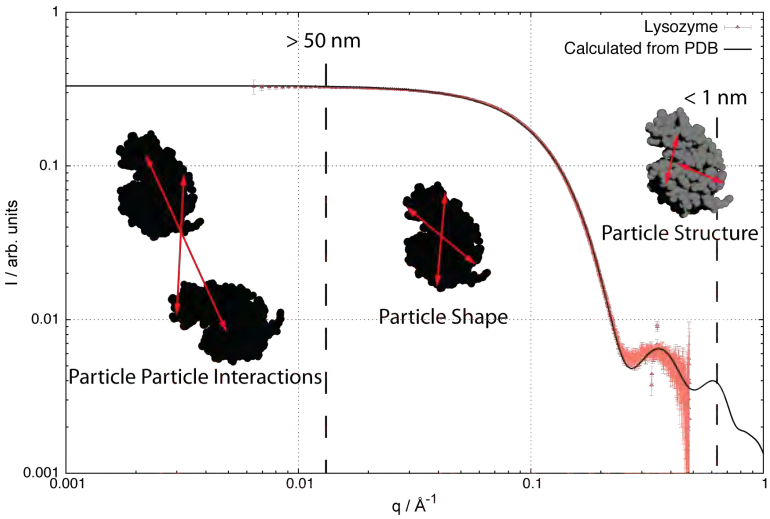


Figure 3.4: Illustration of the information content in a scattering curve. At small q-values the scattering signal contain information on the long length scales. In most cases this will be particle-particle interactions. The intermediate q-values describe the particle shape and the low q-vaules contain information on the internal structure of the particle. Scattering data of lysozyme recorded at the SWING SAXS beamline at SOLEIL is shown in red and scattering of the crystal structure of lysozyme (6LYZ.pdb), calculated using the program CRYSOL, is shown in black.

3.4 Resolution and Information Content

3.4.1 Resolution

The structural resolution can be defined using the Bragg law, $d = \frac{2\pi}{q}$, giving a SAXS experiment with a $q_{max} = 0.63$ Å^{-1} a resolution of about 10 Å. However, this is not a completely fair definition of the resolution of small angle scattering, as much smaller changes in distances are observable, as long as the distance is larger than 10 Å. For example, changes in lipid bilayer thicknesses on the order of 2 Å are clearly detectable[4].

Because the scattering curve is a Fourier transform of the sample scattering length density, short real-space distances are represented at long scattering vectors and long real-space distances by short scattering vectors. Figure 3.4 shows a SAXS curve from the protein lysozyme. In general the scattering curve may be divided in to three parts, containing different kinds of structural information. The central part of the curve contains information on the shape of the protein, whereas low q end describe protein-protein interactions and the high q part describe the molecular arrangement. The limiting factors, at the low q-values are the beam stop and possible parasitic scattering from the direct beam. While at the high q-values it is the fact that the scattering signal becomes weaker than the background noise of the experiment.

3.4.2 Information Content

For small angle scattering data the number of parameters needed to represent a data set from q_{min} to q_{max} can be estimated as the number of Shannon channels[121]:

$$N_s = D_{max}\frac{(q_{max} - q_{min})}{\pi}. \tag{3.27}$$

Because the experimental noise level at large q-values, q_{max} is often not uniquely defined, this leads to an over estimate of the information content present in the data. By analyzing the data using the Bayesian Indirect Fourier Transform (BIFT) method, a better estimate of the information content, called "number of good parameters", N_g, can be determined[128]. This estimate improves on the Shannon channels by also taking into account the experimental noise of the measured data.

The difference may be illustrated by the data shown in figure 3.4 where the number of Shannon channels are found to 5.56, whereas the number of good parameters are 4.57.

Multiple Contrasts

For a detailed analysis, it is preferable to have as much information as possible. For multi contrast particles, such as nanodiscs consisting of both proteins and phospholipids, the information about the scattering particles may be increased by measuring the scattering in different contrast situations. This strategy has been used in papers 1 and 4 by measuring both SAXS and SANS. Furthermore, by changing the deuterium content in the solvent several additional contrast situations can be measured in SANS.

It is not straight forward to quantify the amount of information gained by thes approach, via the number of good parameters. However, it must be less than the sum of good parameters for each data set, as several of the parameters must be correlated, but more than in only one measurement.

By performing a detailed model analysis, an ad-hoc estimate of the increase in information by measuring several contrasts can be performed by examining the change in the certainty of fitted model parameters, when using several contrasts compared to a single contrast.

Multiple Measurements

For a sample made up by several subunits or components an alternative strategy for increasing the information available for data analysis, is to measure the subunits alone. In this case, the result from the analysis should agree with the measured data both the individual subunits and the combined sample. This relies on the assumption that the individual subunits do not change when combined.

This strategy was employed for the data analysis of the Cytochrome P450 3A4 in paper 5. Here, two measurements were done, one of the empty nanodisc, and one of the nanodisc with the incorporated Cytochrome P450 3A4.

3.5 Model Free Data Analysis

The model free analysis methods are often used for an initial analysis and/or quality check of the data as they are relatively simple and quick and can be fully automated.

3.5.1 Guinier Analysis

The first type of model free analysis is the Guinier approximation proposed by Guinier in 1939. This approximation is made by expanding the $sin(qr)$ term in equation 3.19 into a Taylor series:

$$I(q) = \int_0^{D_{max}} \frac{P(r)}{qr}(qr - \frac{q^3r^3}{6} + ...)dr = \int_0^{D_{max}} P(r)dr - \frac{\int_0^{D_{max}} P(r)q^2r^2}{6}dr + ... \quad (3.28)$$

$$= I(0) - \int_0^{D_{max}} P(r)dr\frac{q^2}{3}\frac{\int_0^{D_{max}} P(r)r^2dr}{2\int_0^{D_{max}} P(r)dr} + ... \approx I(0)(1 - \frac{q^2R_g^2}{3}) \quad (3.29)$$

where the forward scattering $I(0) = \int_0^{D_{max}} P(r)dr$. The final approximation is valid only for small $q^2R_g^2$ (typically less than 1.3), where the higher order terms of the expansion are negligible. Furthermore the result above can be considered the first two terms in a powers series expansion of an exponential function yielding:

$$I(0)(1 - \frac{q^2R_g^2}{3}) \approx I(0)e^{-q^2R_g^2/3} \quad (3.30)$$

This provides a very quick way of evaluating the radius of gyration of a measured object as well as the forward scattering at zero angle. The Guinier analysis is mostly of a historic importance, as the more modern Indirect Fourier Transform method is able to provide both of these parameters as well as the pair distance distribution function.

3.5.2 Molecular Mass Determination

At zero scattering angle the molecular mass of a sample can be calculated from data measured on an absolute scale[17][88]

$$\frac{d\Sigma}{d\Omega}(q = 0) = n_v \cdot (\langle \Delta\rho \rangle v)^2 = \frac{M_w \Delta\rho_m^2 c}{N_A}, \quad (3.31)$$

$$M_w = \frac{d\Sigma}{d\Omega}(q = 0)\frac{N_A}{\Delta\rho_m^2 c} \cdot 10^3 \quad (3.32)$$

where n_v is the number density of particles, c is the concentration in mg/ml, $\Delta\rho_m$ the excess scattering length per gram, and N_A Avogadros number. $\Delta\rho_m = (\rho_m - (\rho_{solvent}\nu))$ where ρ_m is the scattering length per gram of the particle, $\rho_{solvent}$ is the scattering length density per unit volume of the solvent and ν is the specific molecular volume of the particle. In SAXS contrast, $\Delta\rho_m$ has an average value of $2 \cdot 10^{10}$ cm/g for globular proteins in water, accurate to within 10 % [88].

3.5.3 Indirect Fourier Transform Method

When looking at equation 3.19, one might be tempted to do a Fourier transform of the measured data, in order to obtain the pair distance distribution function. While this would work in theory, it would require a scattering curve measured from $q = 0$ to $q = \infty$, which is not possible experimentally. At low scattering angles, the beam stop sets a clear limit, whereas at the higher angles the scattering signal is lost in the instrumental background.

The idea proposed by Glatter[48] in 1977, was to propose a $P(r)$ function and evaluate it to the data. However, this is an ill posed problem as there are many $P(r)$ functions that will match the measured data equally well. But by imposing two constraints on the proposed $P(r)$, this problem is overcome. The first constraint is that there is a maximum distance present in the sample, D_{max}. The other is that the proposed $P(r)$ function should be smooth.

This is not a strictly model free method, as we are in fact proposing a model with a smooth $P(r)$ and a certain D_{max}, however, the final D_{max} must be in accordance with the data, and the smoothness constraint can be justified by the limited resolution of small angle scattering.

Using equation 3.19, the proposed function is evaluated to the data by the Least Squares method. The function used by Glatter is a sum of *cubic-b* splines, ϕ_i, spaced equally from $r = 0$ to $r = D_{max}$ weighted by the coefficients c_i:

$$P(r) = \sum_i^k c_i \phi_i(r) \tag{3.33}$$

The scattering from this proposed $P(r)$ is calculated from equation 3.19:

$$I_m(q) = \int_0^{D_{max}} \sum_i^k c_i \phi_i(r) \frac{sin(qr)}{qr} dr = \sum_i^k c_i \int_0^{D_{max}} \phi_i(r) \frac{sin(qr)}{qr} dr = \sum_i^k c_i \chi_i(q) \tag{3.34}$$

and is evaluated to the measured data by minimizing:

$$\chi^2 + \alpha N_c \tag{3.35}$$

where

$$\chi^2 = \sum_n^M \frac{(I(q_n) - \sum_i^k c_i \chi_i(q_n))^2}{\sigma(q_n)^2} \tag{3.36}$$

$$N_c = \sum_i^{k-1} (c_{n+1} - c_n) + c_1^2 + c_n^2. \tag{3.37}$$

χ^2 measures the agreement between the measured data, $I(q_n)$, and the calculated scattering from the proposed $P(r)$ at the same data points, $\sum_i^k c_i \chi_i$. N_c describes the smoothness of the model $P(r)$, and α is a Lagrange multiplier determining the weight of the smoothness constraint relative to χ^2.

In order to perform the IFT three parameters are needed: the maximum distance, D_{max}, the number of spline functions, k, and the magnitude of the Lagrange multiplier, α.

When doing the analysis, one often has a good idea of the D_{max} of the sample in question, and this can be used as an initial guess, and the number of splines should be sufficient to generate a smooth $P(r)$ from $r = 0$ to $r = D_{max}$. α is chosen by finding the point of inflection[48]. This is done by calculating a series of solutions with increasing α and evaluating the χ^2, at the point of inflection, the χ^2 will increase dramatically with only a small increase in α. At the same point, the second derivative of the χ^2 changes sign.

The two constraints imposed on the $P(r)$ are essentially the same as requiring the smoothest best fitting function.

Implementations

Since the IFT method was developed by Glatter and published in 1977, several implementations and improvements have been made, including methods using statistical methods to estimate the maximum distance and the Lagrange multiplier. In this work, the following implementations have been used:

glatSAXS and glatSANS These are FORTRAN implementations originally made by Prof. Jan Skov Pedersen following the original scheme by *Glatter* with a slightly modified smoothness constraint. It estimates the error on the resulting $P(r)$ by Monte Carlo simulations of the experimental data based on the measured errors[113]. The sister program glatSANS works in the same way but is made for handling SANS data which often consist of several data sets measured at different settings and takes the instrumental smearing of the data into account[115].

GNOM The GNOM implementation is part of the very common ATSAS data analysis package package by the *Svergun* group at the EMBL Hamburg Outstation. The output of GNOM is used for some of the more advanced data analysis methods available in the ATSAS package. It is particularly well suited for performing IFTs of globular proteins, and uses 6 parameters to estimate the Lagrange multiplier, and based on these parameters, it provides a quality factor between 0 and 1 of the solution[117]. A related implementation is AUTOGNOM, which estimates D_{max} automatically by evaluating the solutions from $D_{max} = 2R_g$ to $D_{max} = 4R_g$[123].

BIFT This is an implementation of the indirect Fourier transform method based on Bayesian statistics that is able to estimate both the maximum distance as well as the Lagrange multiplier without any input from the user[53, 128]. Because the method uses a statistical approach to estimating the $P(r)$, it provides a realistic error estimate of the result. Furthermore, it also provides an estimate of the information content in the data by estimating the number of good parameters, N_g[128]. It is implemented as a web application available at *www.bayesapp.org*[54].

In this project BIFT or glatSAXS has been used as an initial analysis step for SAXS data. For SANS data glatSANS has been used. BIFT has the great advantage over the other implementations, that it works without input from the user by using a statistical methodology to estimate the smoothness and maximum distance parameters. This makes it free from any personal bias

and provides a realistic estimate of the errors on the resulting $P(r)$. By the end of this project, this method had become very reliable and was the preferred method for IFT.

3.6 Model Based Data Analysis

Due to the rotational average in 3.18, the 3D information on the scattering structure is converted to a 1D projection in form of the pair distance distribution function. This loss of structural information means that it is not possible to uniquely reconstruct the 3D structure of the scattering particles, from the data alone. In the model based data analysis, a model of the scattering particle is constructed based on information from other sources, in order to make up for the lost information. The resulting model is then evaluated against the data. There are traditionally two modeling strategies

The first, here termed *continuous*, is to analytically calculate the form factor using the fact that the scattering is the Fourier transform of the excess scattering length distribution. This approach has been very successful in describing that scattering from liposomes[69, 4], micelles[25, 6] and micro emulsions[29], and it can also be expanded to take polydispersity into account[7]. The calculation of the form factor is generally possible for simple geometrical objects but becomes tedious for complex shapes.

The second method, here termed *discrete*, is to represent a complex shape as an assembly of discreet objects with a simple shape, such as points or spheres, and calculate the scattering as the sum of these[96][114]. This method is very successful in describing the scattering from proteins, that often have non trivial shapes[118, 107], but it is also applicable for other objects of a complex shape, for example: aggregates of silica nanoparticles[92].

During this project, a combination of the two approaches has been developed. This approach has been termed *hybrid* approach. Here, parts of the sample is described as an assembly of discrete objects and other parts as a continuous object.

3.6.1 Continuous Modelling

Using the form factor, the scattering from a particle can be written as:

$$I(q) = \left\langle \left| \int \Delta\rho(\mathbf{r}) e^{-i\mathbf{qr}} d\mathbf{r} \right|^2 \right\rangle_\Omega = (\langle \Delta\rho \rangle \, v)^2 F(q) \tag{3.38}$$

where $\langle \Delta\rho \rangle$ and v is the average excess scattering length density and volume of the particle, respectively. The form factor, $F(q)$, can be calculated as the Fourier transform of the excess scattering length distribution, $\Delta\rho(\mathbf{r})$,:

$$F(q) = \frac{\left\langle \left| \int_v \Delta\rho(\mathbf{r}) e^{-i\mathbf{qr}} d\mathbf{r} \right|^2 \right\rangle_\Omega}{\langle \Delta\rho^2 \rangle \, v^2} = \frac{\left\langle |A(\mathbf{q})|^2 \right\rangle_\Omega}{\langle \Delta\rho^2 \rangle \, v^2} \tag{3.39}$$

The form factor for many geometrical shapes can be calculated. Table 3.6.1 show a list of form

factors used in this work and a more exhaustive list of form factors can be found in *Pedersen 2002*[93] and *Pedersen 1997*[113].

The geometrical form factors can be combined in order to yield new and more complex shapes. For example: The form factor of a hollow cylinder can be calculated by subtracting the form factors of two cylinders with the same scattering length density but different radii. This is illustrated in figure 3.5, where a cylinder with radius R_i is subtracted from a cylinder with radius R_o yielding a hollow cylinder with an inner radius of R_i and outer radius of R_o. Using the form factors from table 3.6.1 the hollow cylinder form factor amplitude is:

$$A_{hc}(q, R_o, R_i, h) = A_2(q, R_o, h) - A_2(q, R_i, h) \tag{3.40}$$

where $F_2(q, R_o, h) = |A_2(q, R_o, h)|^2$. Paper 2 presents a model for the empty nanodisc system based on this principle, where the MSPs spanning the rim of the phospholipid bilayer are modeled by a hollow cylinder with an elliptical cross section.

Building models of this type quickly becomes very complex, containing many parameters. In order to reduce the number of free parameters in the model, constraints based on the molecular composition of the sample can be introduced.

Molecular Constraints

The use of molecular constraints is an attempt to introduce molecular information known from other sources into the model, in order to reduce the number of free parameters. First of all, when the data is measured on an absolute scale, a strong constraint is the sample concentration as,

$$\frac{d\Sigma}{d\Omega}(q)_{measured} = n_{measured} \langle \Delta\rho \rangle^2_{mod} v^2_{mod} F_{mod}(q)$$

will have to be true. Where $\frac{d\Sigma}{d\Omega}(q)_{measured}$ is the experimentally measured scattering intensity on an absolute scale and $n_{measured}$ the measured concentration.

Secondly, the molecular composition of the sample can be used to impose constraints on the geometrical parameters of the model. In the case of the nanodisc model, the MSPs are modeled as a hollow cylinder spanning the rim of the phospholipid bilayer. This is used to connect the radius of the phospholipid bilayer to the inner radius of the MSP, as the model does not allow a gap between the MSP and the lipid rim. Furthermore, the volume of the hollow cylinder is constrained to the volume of two MSP proteins, connecting the height and the width of the hollow cylinder. The volume of the central phospholipid bilayer is constrained by the number of lipids and the partial molecular volume of a single lipid molecule. In this way, the number of free parameters in the model presented in paper 2 are reduced from 20 to 7.

3.6.2 Discrete Modeling

In contrast to the continuous approach, the discrete modeling approach represents the studied particles as a sum of discrete "dummy" scatterers (often points or small spheres) with a scattering length. Depending on the desired resolution, these "dummy" scatterers can represent atoms[119][107], amino acids[122] or simply the scattering length density of the particle[114][52][118].

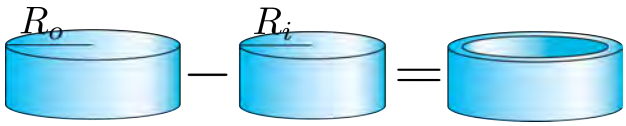


Figure 3.5: Illustration of how form factor amplitudes can be combined to obtain more form factors of more complex shapes.

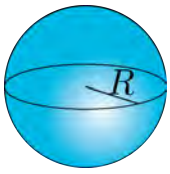
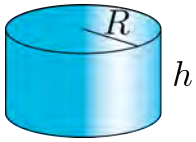
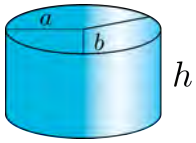
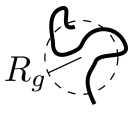
Geometrical Object	Form Factor
Sphere 	$F_1(q,R) = \left\| 3\left(\frac{sin(qR) - cos(qR)\cdot Rq}{(qR)^3}\right)\right\|^2$
Cylinder[44] 	$F_2(q,R,h) = \left\| \int_0^{\pi/2} \frac{2J_1(sin(\alpha)\ qR) sin(cos(\alpha)\ q\frac{h}{2})}{qR\ sin(\alpha)\ cos(\alpha)\ \frac{h}{2}q} sin(\alpha) d\alpha \right\|^2$
Cylinder w Eliptical Cross Section[84] 	$F_3(q,a,b,h) = \left\| \int_0^{\pi/2}\int_0^{\pi/2} \frac{2J_1(sin(\alpha)\ qR(\phi)) sin(cos(\alpha)\ q\frac{h}{2})}{qR(\phi)\ sin(\alpha)\ cos(\alpha)\ \frac{h}{2}q} d\phi sin(\alpha) d\alpha \right\|^2$ $R(\phi) = \sqrt{a^2 sin^2(\phi) + b^2 cos^2(\phi)}$
Gaussian Random Coil[36] 	$F_4(q,R_g) = \frac{2(e^{-R_g^2 q^2} + R_g^2 q^2 - 1)}{R_g^2 q^2}$

Table 3.3: List of form factors from different geometrical objects. Here the same coordinate notation as in *Pedersen 2002*[93] and *Pedersen 1997*[113] has been used: α is defined as the angle to the z axis and ϕ as the angle to the x axis

The Debye Sum

The scattering intensity of discrete scatterers of spherical symmetry can be calculated using the so called Debye sum[35]:

$$I(q) = \left\langle \sum_i^N \sum_j^N A_i(q) A_j^*(q) \frac{sin(qr_{ij})}{qr_{ij}} \right\rangle_\Omega \tag{3.41}$$

where N is the number of scatteres, $A_i(q)$ is the form factor amplitude of the i'th scatterer, r_{ij} is the distance between the i'th and j'th scatterer, * denotes the complex conjugate and $\langle\rangle_\Omega$ denotes the spherical average. Using this formula, the scattering from complex shapes can be calculated by representing the shape by a number of points or small spheres.

Calculation of The Scattering from Atomic Coordinates

For many proteins there is a crystal structure available, that can be used to calculate the small angle scattering pattern in solution. In order to calculate the small angle scattering from the atomic structure, three terms have to be taken into account. These are the scattering of the protein in vacuum, the scattering from the excluded solvent and she scattering from a hydration layer of solvent molecules at the surface of the protein[119]:

$$I(q) = \left\langle \left| A_{protein}(\mathbf{q}) - \rho_{solvent} A_{excluded}(\mathbf{q}) + \Delta\rho A_{border}(\mathbf{q}) \right|^2 \right\rangle_\Omega . \tag{3.42}$$

Where $A_{protein}(\mathbf{q})$ is the scattering amplitude from the protein, $\rho_{solvent} A_{excluded}(\mathbf{q})$ is the scattering amplitude form the excluded volume and $\Delta\rho A_{border}(\mathbf{q})$ is the scattering from the hydration layer with an excess scattering length of $\Delta\rho$, compared to the rest of the solvent.

There are several implementations of this idea[107][126], where the most widely used is the program CRYSOL[119].

CRYSOL This implementation is widely used and was the first to include the scattering contribution from a hydration layer around the protein. The calculation of the scattering amplitude of the protein is done by adding the scattering amplitudes of each individual atom in vacuum, $A(\mathbf{q})$, multiplied by the correct phase factor according to the position of the atom:

$$A_{protein}(\mathbf{q}) = \sum_i A_i(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}} \tag{3.43}$$

The contribution to the scattering from the excluded solvent is calculated by replacing each atom of the protein structure with a dummy atom with the scattering length density of the solvent. The form factor of the dummy atom is a Gaussian sphere with radius corresponding to the replaced atom. The scattering from the hydration layer is found by calculating an envelope function, describing the surface of the protein, and then describing the hydration layer as a shell with the shape of the protein surface and a thickness[119]. By evaluating the calculated scattering profile against experimental data, the thickness and the density of the hydration layer may be fitted. The scattering from the crystal structure of the protein Lysozyme, using CRYSOL, is compared to experimental SAXS data in figure 3.4.

CRYSOL can only calculate the x-ray scattering profile, however the related program CRYSON can be used to calculate the neutron scattering profile from an atomic structure. Both programs are included in the ATSAS software package.

Ab Initio Modeling

Using *Ab initio* modeling it is often possible to reconstruct the 3D shape of a scattering particle, from the measured data in the form of an assembly of spheres or points. The approach is to start out with an initial guess on the shape (often just a sphere), represented by a collection of dummy scatterers. The initial guess is then refined iteratively against the measured data, by moving the dummy scatterers around. After each move the scattering from this new configuration is compared to the data, and the move is either accepted or rejected according to a Mont Carlo procedure, often simulated annealing.[94, 118, 122]

DAMMIN One implementation of the *ab initio* approach is the program DAMMIN, available from the ATSAS software package[118].

In DAMMIN, the sample is considered to be made up by K phases each with N_k beads of contrast of $\Delta\rho_K$. A sample consisting of a protein in solution is in this context regarded as a two phase system: a solution phase $K = 0$ and a protein phase $K = 1$. The scattering is calculated as:

$$I(q) = \left\langle \left(\sum_{k}^{K} \Delta\rho_K A(q)_K \right)^2 \right\rangle_\Omega \tag{3.44}$$

In order to gain computational speed, the scattering amplitude of each phase is expanded as a sum of spherical harmonics

$$A_K(q) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}^K(q) Y_{lm}(\theta, \phi) \tag{3.45}$$

where $\mathbf{q} = (q, \theta, \phi)$ and $Y_{lm}(\theta, \phi)$ is the spherical harmonic of the lm'th order and

$$A_{lm}^K(q) = i^l \sqrt{2/\pi}\, A_{bead}(q) \sum_{j}^{N_K} j_1(qr_j) Y_{lm}^*(\alpha_j, \beta_j) \tag{3.46}$$

j_1 is the first order Bessel function, N_K is the number of dummy atoms in the phase K, and A_{bead} is the form factor amplitude of the beads. Each bead is placed at $\mathbf{r}_j = (r_j, \alpha_j, \beta_j)$.

The scattering intensity obtained by combining equation 3.44 and 3.45 is[118]:

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\sum_{k=1}^{K} (\Delta\rho_k A_{lm}^k(q))^2 + 2 \sum_{n>k} \Delta\rho_k A_{lm}^k \Delta\rho_n\, A_{lm}^n(q)^* \right) \tag{3.47}$$

The initial configuration for DAMMIN is an assembly of beads fixed on a grid within a confined volume, typically a sphere with radius $R = D_{max}/2$. Instead of moving the beads, they are kept

fixed but are randomly shifted from one phase to the other by changing their excess scattering length density. The resulting configuration is evaluated against the data by the χ^2 and a so-called looseness criterion to emphasize compact models relevant for proteins, $P(X)$, by using a Monte Carlo search algorithm to minimize:

$$f(X) = \chi^2 + \alpha P(X) \tag{3.48}$$

$$P(X) = 1 - \left\langle 1 - (e^{-0.5N} - e^{-0.5N_c}) \right\rangle \tag{3.49}$$

For each bead, N is the number of nearest neighbors of the same phase, N_c is the number of nearest neighbors for an ideal compact phase ($N_c = 12$) and $\langle\rangle$ denotes the average over all beads in the model. P(X) is weighted with respect to the χ^2 via the Lagrange multiplier α. The choice of alpha is dependent on the weight of the looseness criterion relative to the χ^2, and should reflect that the criterion should be important at the end of the Monte Carlo simulation when the dimensions of the particle has been established.

Figure 3.6 show 10 *ab initio* models of lysozyme computed from SAXS data. As will be discussed later the solutions produced by the *ab initio* approach are often not unique, due to the large number of free parameters in such a model. However, they are often quite similar.

3.6.3 Hybrid Modeling

The new concept of hybrid modeling has been developed as part of this project, primarily by *Søren Kynde*. It has risen out of a desire to combine the strengths of both the continuous and the discrete modeling approaches. In particular, a desire to describe the scattering from a nanodisc with an incorporated membrane protein, using a continuous model for the nanodisc and a discrete model for the membrane protein.

The key issue has been to combine the calculation of the scattering amplitude from a continuous model with the calculation of the scattering amplitude from a discrete model. This is solved by expanding the continuous form factor amplitude in a series of spherical harmonics and adding it to the amplitudes of a discrete model also calculated as a series of spherical harmonics.

A continuous form factor amplitude can be expanded as a series of spherical harmonics as[85]:

$$I(q) = n(\langle\Delta\rho\rangle\, v)^2 F(q) = n \sum_{l=0}^{L} \sum_{m=-l}^{l} |B_{lm}(q)|^2 \tag{3.50}$$

$$B_{lm} = \left\langle \langle\Delta\rho\rangle\, v A(q,\theta,\phi) \frac{e^{-im\phi}}{\sqrt{2\pi}} \tilde{P}_{lm}\left(cos\left(\theta\right)\right) \right\rangle_{\Omega}, \tag{3.51}$$

where $\langle\Delta\rho\rangle$, v and A is the average excess scattering length density, the volume and the form factor amplitude of the continuous model, respectively. $\tilde{P}_{lm}$ are the Legendre polynomials, and q, ϕ, θ are the coordinates of the scattering vector $\mathbf{q}$ in spherical coordinates. $\langle\rangle_{\Omega}$ denotes the spherical average over the angles α and β.

This can readily be combined with equation 3.46 to produce the scattering intensity from an object described by both a continuous and a discrete part:

$$I(q) = n \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm} + B_{lm}|^2 \tag{3.52}$$

Coarse Graining In paper 4 and 5, the hybrid model was used in order to describe the scattering from nanodiscs with incorporated membrane proteins. In both cases, the protein were represented by discrete coarse grained models derived from all atom atomic model of the crystal structure. The coarse graining was done by representing each amino acid in the crystal structures by a point with the appropriate scattering length, centered at the center of scattering of each amino acid.

Hybrid Ab Initio Modelling

Representing part of the sample by a discrete model opens for performing *ab initio* shape reconstructions. This is used in paper 5 to reconstruct the shape of the membrane protein Cytochrome P450 3A4 anchored to the lipid bilayer in a nanodisc.

This *ab initio* analysis is initiated with a model, where a point representing each amino acid of the membrane anchored protein, is placed randomly within a spherical volume with a radius of 40 Å. Form this initial configuration, a random point is selected and moved, resulting in a new configuration, X. Each move is evaluated using a target function:

$$F(X) = \chi^2(X) + \alpha C(X) + \beta H(X) + \gamma T(X) \tag{3.53}$$

The target function evaluates the new configuration based on four criteria:

$\chi^2(X)$ measures the agreement between the calculated scattering of the configuration, X, and the experimental data and $C(X)$ provides a measure of the disconnectedness of X. $H(X)$ compares the point-point distance distribution of X, to an empirical amino acid distance histogram calculated from 20 representative protein crystal structures, in order to give a more protein like local structure[122]. The final term promotes the placement of points in the phospholipid bilayer of the nanodisc, if there are less than 14 amino acids in the lipid bilayer, $T(X) = (n - 14)^2$ and if there are more than 14, $T(X) = 0$.

The target function is used to determine whether or not a new configuration is accepted. If a new configuration, X_2, has $F(X_1) > F(X_2)$ it is accepted and on the other hand if $F(X_1) < F(X_2)$ it is accepted only with the probability

$$P(X_2, X_1) = e^{-(F(X_2) - F(X_1))/B}. \tag{3.54}$$

The process of selecting a random point, moving it and evaluating the target function for the new configuration, is then repeated. One pass is completed after n_{aa} successful moves, where n_{aa} was the number of movable points (in this case the number of amino acids). Initially, B is

set to $\chi^2(X_0)/10$ but after each pass B is decreased by 5%, thereby decreasing the probability of accepting configurations that does result in a smaller target function. In the case presented in paper 5 the target function converged after about 70 passes.

3.6.4 Uniqueness of Over Parameterized Models

The number of parameters in a model made up of one point per amino acid in case of the hybrid *ab initio* model or $N = (D_{max}/2r)$ (where r =0.8Å using the standard settings) in the case of DAMMIN, is far greater than the number of good parameters present in the measured data. This means that the experimental data are vastly overfitted by such a model, and that the solutions found by such an analysis are not unique.

However, It turns out that the resulting molecular shapes often are very similar. This is illustrated in figure 3.6, where 10 results from a DAMMIN analysis of scattering from Lysozyme are shown. Although the placement of the beads is different in each model, they are strikingly similar in both size and shape, pointing towards a general solution.

DAMAVER

The program DAMAVER provides a method for evaluating several *ab initio* models in terms of consensus between the models and calculates an average model as well as a filtered model[129]. All models are first aligned and superimposed onto one another using the helper program SUPCOMB. Each model is remapped onto the same grid of spheres, making it possible to determine in how many of the models, a given sphere occur.

This can be used to determine what structural elements are most frequently occurring in the reconstructed models and by constructing an average and filtered model. The average model is the sum of all the aligned solutions. The filtered model is constructed by taking only the most occurring elements and cutting of by the average volume of all the compared models.

It is important to keep in mind that, even though all the *ab initio* models fit the experimental data the average model or the filtered model will not. They only represent the average or shared structure of the solutions. Additionally, only regions that are relatively static, will be kept in the filtered model as the flexible parts will be cut off. The filtered model of ten DAMMIN *ab initio* models of Lysozyme is shown in figure 3.6.

3.7 Perspectives

Solution SAXS and SANS are powerful methods for studying particles of nanometer sizes in solution. Compared to other techniques that can access similar size ranges, solution scattering has the advantage that they are able to probe the bulk of the solution and a large number of particles, whereas cryo-EM or AFM probe individually selected particles. However because a large number of particles are probed at the same time only information on the particles on

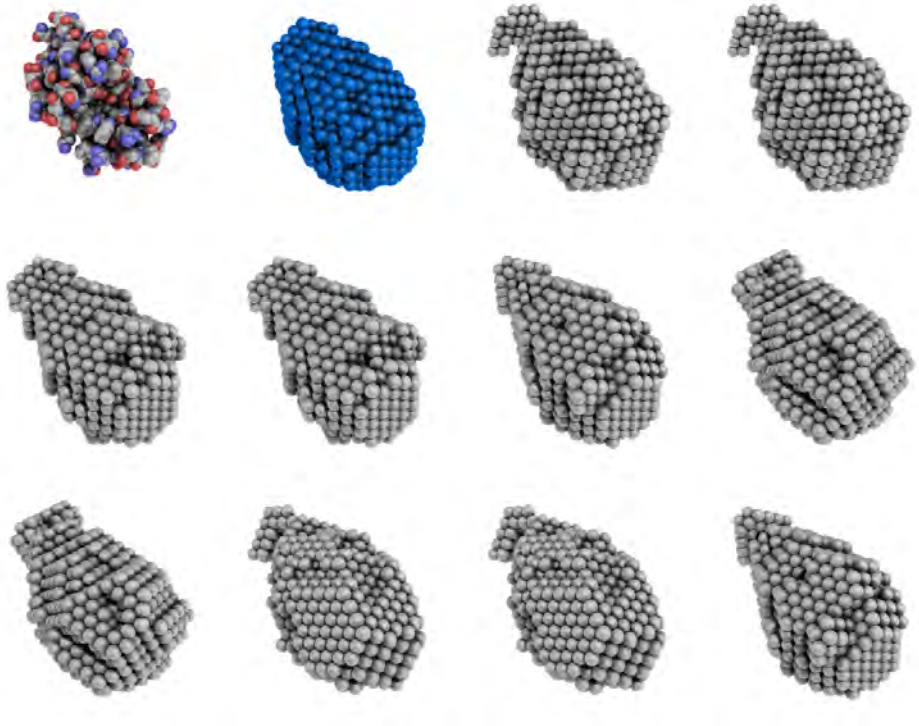


Figure 3.6: DAMMIN shape reconstruction of scattering data from Lysozyme. Grey: models resulting from 10 analysis attempts. Blue: Average structure with outliers filtered out. The crystal structure of Lysozyme (6LYZ.pdb) is shown for reference in the top left. All models are drawn on the same scale.

average can be extracted. Thefore homogenous samples are a requirement when preforming *ab initio* reconstructions of proteins.

The interpretation of small angle scattering data is complicated by the fact that the observed scattering represents the rotational average of the particle. In order to obtain detailed structural information a model based analysis is often required, making the data analysis of scattering data a rather time consuming affair. However, when using a consistent model for analyzing the data it is possible to obtain detailed structural information using small angle scattering.

The work presented here has benefitted hugely from the simultaneous SAXS and SANS beamtimes provided by the ILL and the ESRF. This access has been crucial in order to confidently preform a combined analysis of the measured data, knowing that the measured data are from the exact same samples. Hopefully this possibility may be introduced at other facilities in the future.

The author has been involved in proposing a new BioSANS instrument for the ESS neutron source to be build in Lund, Sweden. The proposed instrument is an attempt to exploit the high neutron flux of the ESS in order to build an instrument, comparable to present BioSAXS

instruments, optimized for measuring small volumes of macromolecular biomolecules. If built this will hopefully help to apply the advantages of SANS to samples that have previously been possible due to sample volume requirements.

The next generation of x-ray sources, free electron lasers, have shown promising results for single molecules scattering[60] and nanocrystal diffraction[59]. Future sources of this type may be able to do single/few molecule scattering in a timescale shorter then the rotational motion of the molecules, overcoming the loss of the 3D information due to the orientational average.

Chapter 4

Results and Conclusion

4.1 Structure and Shape of Nanodiscs

Most previous structural studies of nanodiscs and discoidal HDLs have concerned the size and geometry of the particles. These studies have been done using a broad range of structural techniques, including AFM, cryo-EM and small angle scattering, as well as molecular dynamics simulations. The work presented in paper 1 show that using a molecular constrained model of the small-angle scattering data, it is possible to obtain information not only on the shape and geometry of the nanodisc particles but also information on the molecular environment of the lipid disc. In particular, the effect of the hydrophobic mismatch, between the lipids and the MSPs, is directly observed on the lipids.

It is this high level of detail that provides a basis for advancing to the more complicated problem of studying membrane proteins incorporated into nanodisc using SAXS and SANS.

Shape and Polydispersity

For the nanodisc model presented in papers 1 and 2, an elliptical cross section of the disc was necessary in order to accurately describe the measured data. In small-angle scattering the observation of ellipsoids is often ascribed to the presence of polydispersity, as these two phenomena can be very hard to distinguish based on scattering data alone. This line of argument can be extended to also include cylinders with circular and elliptical cross sections. One might therefore argue that the need to introduce an elliptical cross section, is in fact due to a polydispersity in the size and/or shape of the nanodisc particles.

Caponetti *et. al* have in 1993 thoroughly discussed the case of polydisperse spheres and ellipsoids[26]. The authors argue that because the size distribution that model the ellipsoids is very atypical (i.e. very far from Gaussian), it is in principle possible to distinguish between a system of polydisperse spheres with a Gaussian like distribution of radii and ellipsoids. However, this may in

practice not always be the case, due to the inherent measurement uncertainty on experimental scattering data[26]. The authors conclude that ideally both types of models should be evaluated against the data.

In the discussion by Caponetti *et. al*, a homogenous scattering contrast is assumed. However, in the case of nanodiscs, the presence of both the MSP and the phospholipids give rise to a much more complex contrast situation. This complex contrast situation, gives a more information rich scattering pattern, that may make it more easy to distinguish polydispersity and anisotropy[18]. Additionally, the models used for the nanodiscs, rely not only on the information available from scattering pattern, but also include molecular information via the incorporation of molecular constraints.

Following the recommendation of Caponetti *et. al*, a model with circular nanodiscs with a truncated normal distribution of radii, was evaluated against the data. Compared to the model with an elliptical cross section, it produced slightly less good fits. Additionally the polydisperse discs had very different circumferences, in poor agreement with the fact that these are stabilized by MSPs of the same length. For these reasons, the elliptical model was chosen over the polydisperse model.

One hypothesis to explain the observed elliptical cross section is that the MSPs have a fixed length. If the nanodiscs are prepared at a lower than fully loaded lipid to MSP ratio a consequence of this hypothesis will be that the discs have an elliptic cross section. In paper 3 it is shown that even discs prepared in a high excess of phospholipids still have an elliptic cross section, ruling out that our observations of elliptical discs are artifacts of "under loading" the nanodiscs. By assuming that the MSP has a fixed length, there is also an entropic argument against the formation of perfect circular discs, as such a disc can be formed in a much more limited number of ways compared to a disc with an elliptic cross section. This argument is discussed in detail in Paper 1.

Citing the finding of an elliptical cross section in paper 1 as inspiration, Maleki *et al.* have recently published a theoretical study of the stability of circular nanodiscs[77] under different perturbations. Using a continuum model of the nanodisc they find that under certain conditions the nanodisc may indeed adopt a flat elliptic or saddle shape.

Both the nanodisc system and discoidal apo-A1 HDLs are considered to be monodisperse, from size exclusion data. However in paper 3, SAXS measurements of 5 fractions along the nanodisc SEC elution peak show more loaded and less elliptical nanodiscs in the first fractions and less loaded and more elliptic nanodiscs in the last fractions. However, none of the fractions are found to contain circular nanodiscs. This combination of SAXS and size exclusion shows that the discs are both elliptical and polydisperse.

The use of the molecular constraints in the model limits the possible solutions in such a way that that a breaking of symmetry is required in order to fit the data. In this context the finding of an elliptical cross section may be viewed as a first order deviation form a circular cross section. The model implicitly assumes a flat bilayer disc. It is possible that the deviation from a circular geometry, represented by the elliptic cross section, in fact represents an out of plane anisotropy, fluctuations in the shape of the nanodisc or saddle shape discussed by Maleki *et al.*.

Nanodiscs as a Model Systems for Lipid Bilayers

As one of the aims of nanodisc technology is to provide a solution based tool for studying membrane proteins in a native lipid environment, it is relevant to discuss how the nanodisc lipids compare to lipids in a membrane.

It is clear from the data presented in this work and by others, that the presence of the MSP perturbs the lipids in the nanodisc, compared to lipids in bilayer vesicles. In Paper 1 the perturbation is credited as a consequence of the hydrophobic mismatch between the MSP and the phospholipids. The adjustment of the lipid tail length to minimize the hydrophobic mismatch changes the lateral pressure in the lipid bilayer both positively and negatively, depending on the lipid tail length. This effect is in agreement with calorimetric studies of the phospholipids in nanodiscs, where the main transition temperature is shifted upwards by 2 to 5 degrees depending on the phospholipid, compared to vesicles[82]. Denisov *et. al* ascribe this increase in melting temperature to a domain of perturbed lipids along the of the edge of the bilayer, which is about 2 lipids wide. Additional studies on the lipid environment has been done by Nakano *et al.*[89]. Using fluorescent probes and neutron scattering, they study both the static and dynamic lipid environment in DMPC nanodiscs and compared to vesicles. They find that the acyl chains of DMPC on average are more ordered in the nanodisc compared to vesicles and that there is an increased contact frequency with water molecules for the acyl groups located at the bilayer interface of the nanodiscs. Both results again show that the presence of the scaffolding proteins perturb the phospholipid bilayer of the nanodisc compared to vesicles. Furthermore Nakano *et al.* also find an increased exchange rate of DMPC between the nanodiscs.

These results all show that the state of the lipids incorporated into the nanodisc is slightly altered compared to the lipids in vesicles, however they are still able to provide a functional lipid environment for many membrane proteins[104]. Depending the studied membrane protein the effect of the slightly perturbed lipid environment may have to be taken into account.

Self-Assembly

The self-assembly of nanodiscs has previously been studied in the form of a partial phase diagram for the formation of nanodisc and as molecular dynamics simulations[12][111]. In paper 3, the high level of information provided by the model from paper 2 is used to investigate the effect of different parameters on the nanodisc self-assembly. In contrast to studies of the effect of detergent removal speed on vesicles[71], no effect of the detergent removal speed was observed within the timescale accessible. However, the choice of detergent, in particular the size of the mixed lipid detergent micelle, and the lipid:MSP stochiometry both have a large impact on the assembled nanodisc.

Paper 3 reports that detergents with low packing parameters, forming small mixed micelles, containing relatively few lipids such as cholate or chaps, result in nanodiscs with ∼130 POPC molecules. On the other hand, detergents with higher packing parameters form nanodiscs with a decreasing number of lipids. One may speculate that there are two distinct self-assembly processes for nanodiscs made from mixed micelles with fewer than 130 lipids and mixed micelles with more than 130 lipids. In the first case the nanodiscs must be formed by micelle fusion and in the second case by breaking of micelles. These different pathways may have different parameters

for the formation of an "ideal" nanodisc. Additionally, adding a membrane protein to the pre-nanodisc lipid detergent micelles may change these parameters in an unforeseen manner.

As concluded in the nanodisc section, the basis of the nanodisc self-assembly process can be understood by principles of amphiphilic self-assembly. However the finer details are not yet in place. A more complete understanding of the self assembly process is a prerequisite for a predictable reconstitution of membrane proteins into nanodiscs.

4.2 Nanodiscs for Structural Studies of Membrane Proteins

Papers 4 and 5 show the final results of this thesis in the effort of using nanodiscs to study membrane protein structure. They also showcase the "hybrid" approach for combining discrete models with arbitrary shape with models using geometrical objects, which has not previously been done. Using the molecular constrained nanodisc model as a basis it is possible of extract structural information about a membrane protein incorporated into the lipid bilayer in the nanodisc.

The nanodisc system in combination with small-angle scattering, both may be used study the membrane protein as well as the interactions between the lipids and the incorporated membrane protein. This is evident from the analysis of small-angle scattering data from nanodiscs with incorporated bacteriorhodopsin presented in paper 4. Here it is observed that the area per lipid head group increases in nanodiscs with bacteriorhodopsin compared to "empty" nanodiscs (See figure 5 in paper 4). This information is obtained along with information on the localization of the membrane protein in the lipid bilayer, something that may be difficult to obtain using e.g. cryo-EM where the lipids are clearly resolved due to the low electron density.

Nanodiscs have successfully been used to observe the effect of the phospholipid environment on the activity of an incorporated membrane protein[99]. The results presented in paper 4 indicate that it may be possible to observe the underlying structural causes of this using small-angle scattering and nanodiscs. This was indeed the aim of the short, unfinished, study presented in chapter 10 where the effect of negatively charged phospholipids and calcium ions on the membrane protein tissue factor was investigated.

Finally the "hybrid" approach also opens up for doing *ab initio* molecular shape reconstructions of membrane proteins while incorporated into the nanodisc. In the form presented in paper 5, the analysis hinges on the assumption that the phospholipid bilayer is only perturbed minimally by the presence of the membrane protein. However, if more data is available e.g. in the form of neutron scattering data this assumption may not be needed.

Paper 5 also describes how the *ab initio* models resulting from "hybrid" approach can be plugged into the *ab initio* analysis methodologies already developed by Svergun *et al.* for interpreting *ab initio* models for soluble proteins.

4.3 Outlook and Conclusion

It is a hope that the knowledge and methodologies developed in this work will help to push the field of membrane protein structure and function forward. One potential obstacle for this goal is that the data analysis of small angle scattering data can be somewhat complicated requiring tailor made computer programs. Therefore most of the models used in this work are included in the "Will it Fit?" [95], hopefully making them more accessible.

In conclusion this work shows that nanodiscs can be used as a tool for small-angle scattering studies of membrane proteins and to gain additional information on the lipid protein interaction, not easily accessible with other techniques. Furthermore a methodology has been developed to combine geometrical modeling with discreet models, concluding in the option of preforming *ab initio* modeling of membrane proteins anchored in nanodiscs.

The next chapters present the scientific results of this thesis. Papers 1,2 and 4 are published whereas 3 and 5 are submitted here as paper drafts.

Bibliography

[1] Sligar lab homepage. http://sligarlab.life.uiuc.edu/nanodisc.html, 2013.

[2] A Abdul Ajees, G M Anantharamaiah, Vinod K Mishra, M Mahmood Hussain, and H M Krishna Murthy. Crystal structure of human apolipoprotein A-I: insights into its protective effect against cardiovascular diseases. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103(7):2126–2131, February 2006.

[3] G Allen and J S Higgins. Physico-chemical aspects of neutron studies of molecular motion. *Reports On Progress In Physics*, 36:1073, 1973.

[4] Heidi D Andersen, Chunhua Wang, Günther H Peters, Lise Arleth, and Peter Westh. Reconciliation of opposing views on membrane-sugar interactions. *PNAS*, 108(5):1874–1878, February 2011.

[5] Olaf S Andersen and Roger E Koeppe II. Bilayer Thickness and Membrane Protein Function: An Energetic Perspective. *Annual Review Of Biophysics And Biomolecular Structure*, 36(1):107–130, June 2007.

[6] Lise Arleth, D Posselt, D Gazeau, C Larpent, T Zemb, K Mortensen, and Jan Skov Pedersen. Small-angle scattering study of TAC8: A surfactant with cation complexing potential. *Langmuir*, 13(7):1887–1896, 1997.

[7] Lise Arleth and Jan Skov Pedersen. Droplet polydispersity and shape fluctuations in AOT [bis(2-ethylhexyl)sulfosuccinate sodium salt] microemulsions studied by contrast variation small-angle neutron scattering. *Physical Review E*, 63(10):061406, 2001.

[8] B J Baas, SG Sligar, and I G Denisov. ScienceDirect.com - Archives of Biochemistry and Biophysics - Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment. *Archives Of Biochemistry And Biophysics*, 2004.

[9] Sourabh Banerjee, Thomas Huber, and Thomas P Sakmar. Rapid incorporation of functional rhodopsin into nanoscale apolipoprotein bound bilayer (NABB) particles. *Journal Of Molecular Biology*, 377(4):1067–1081, 2008.

[10] D Bashtovyy, G M Anantharamaiah, J P Segrest, and M K Jones. Sequence conservation of apolipoprotein A-I affords novel insights into HDL structure-function. *Journal Of Lipid Research*, 52(3):435–450, February 2011.

[11] D Bashtovyy, Andrea Catte, James C Patterson, M K Jones, W Gray Jerome, Zhengchang Su, Feifei Gu, Jianguo Chen, Marcela P Aliste, Stephen C Harvey, Ling Li, Gilbert Weinstein, and J P Segrest. Novel Changes in Discoidal High Density Lipoprotein Morphology: A Molecular Dynamics Study. *sciencedirect.com*, 90(12):4345–4360, June 2006.

[12] Timothy H Bayburt, Yelena V Grinkova, and SG Sligar. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters*, 2(8):853–856, August 2002.

[13] Timothy H Bayburt, Yelena V Grinkova, and SG Sligar. Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs. *Archives Of Biochemistry And Biophysics*, 450(2):215–222, 2006.

[14] Timothy H Bayburt, Andrew J Leitz, Guifu Xie, Daniel D Oprian, and SG Sligar. Transducin activation by nanoscale lipid bilayers containing one and two rhodopsins. *Journal Of Biological Chemistry*, 282(20):14875–14881, 2007.

[15] Timothy H Bayburt and SG Sligar. Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. *Protein Science*, 12(11):2476–2481, 2003.

[16] Timothy H Bayburt and SG Sligar. Membrane protein assembly into Nanodiscs. *Febs Letters*, 584(9):1721–1727, 2010.

[17] A Bergmann, Otto Glatter, and D Orthaber. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal Of Applied Crystallography*, 33:218–225, 2000.

[18] Magnus Bergström and Jan Skov Pedersen. Structure of pure SDS and DTAB micelles in brine determined by small-angle neutron scattering (SANS). *Phys. Chem. Chem. Phys.*, 1(18):4437–4446, 1999.

[19] Alice Berthaud, Stéphanie Mangenot, John Manzi, and Javier Pérez. Modeling detergent organization around aquaporin-0 using small-angle X-ray scattering. *Journal Of The American Chemical Society*, 134(24):10080–10088, June 2012.

[20] D K Blumenthal, P A Seeger, D B Heidorn, H Crespi, A R Means, S E Rokop, and Jill Trewhella. Changes in the structure of calmodulin induced by a peptide based on the calmodulin-binding domain of myosin light chain kinase. *Biochemistry*, 28(16):6757–6764, 1989.

[21] T Boldog, S Grimme, Mingshan Li, SG Sligar, and G L Hazelbauer. Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *PNAS*, 103(31):11509–11514, 2006.

[22] T Boldog, Mingshan Li, and G L Hazelbauer. Using nanodiscs to create water-soluble transmembrane chemoreceptors inserted in lipid bilayers. *Two-Component Signaling Systems, Pt B*, 423:317–335, 2007.

[23] Jonas Borch and Thomas Hamann. The nanodisc: a novel tool for membrane protein studies. *Biological Chemistry*, 390(8):805–814, 2009.

[24] C G Brouillette, G M Anantharamaiah, JA Engler, and DW Borhani. Structural models of human apolipoprotein A-I: a critical analysis and review. *Biochimica Et Biophysica Acta-Molecular And Cell Biology Of Lipids*, 1531(1-2):4–46, 2001.

[25] B Cabane, R DUPLESSIX, and T Zemb. *Journal De Physique*, (12), 1985.

[26] E Caponetti, MA Floriano, E Didio, and R Triolo. On the shape of the radial-distribution function of an assembly of monodisperse ellipsoidal scatterers. *Journal Of Applied Crystallography*, 26:612–615, 1993.

[27] Elisabeth P Carpenter, Konstantinos Beis, Alexander D Cameron, and So Lwata. Overcoming the challenges of membrane protein crystallography. *Current Opinion In Structural Biology*, 18(5):581–586, 2008.

[28] David Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, September 2005.

[29] S H Chen. Small-Angle Neutron-Scattering Studies of the Structure and Interaction in Micellar and Microemulsion Systems. *Annual Review of Physical Chemistry*, 37:351–399, 1986.

[30] Brett Chromy, Jenny Cappuccio, Matthew Coleman, Craig Blanchette, Edward Kuhn, Todd Sulchek, Erin S Arroyo, Joel M Kralj, Angela K Hinz, Brent Segelke, Kenneth J Rothschild, Julia E Fletcher, Federico Katzen, Todd C Peterson, Wieslaw A Kudlicki, Graham Bench, and Paul D Hoeprich. Cell-free Co-expression of Functional Membrane Proteins and Apolipoprotein, Forming Soluble Nanolipoprotein Particles. *Molecular & Cellular Proteomics*, 7(11):2246–2253, 2008.

[31] B Hong Chung, J L Jones, C G Brouillette, T C Ng, Henri Kercret, and J P Segrest. Structural studies of apolipoprotein A-I/phosphatidylcholine recombinants by high-field proton NMR, nondenaturing gradient gel electrophoresis, and electron microscopy. 23(2):359–367, 1984.

[32] John H Clint. Micellization of mixed nonionic surface active agents. *Journal of the Chemical Society, Faraday Transactions 1*, 71:1327, 1975.

[33] Davide Comoletti, Alexander Grishaev, Andrew E Whitten, Igor Tsigelny, Palmer Taylor, and Jill Trewhella. Synaptic Arrangement of the Neuroligin/β-Neurexin Complex Revealed by X-Ray and Neutron Scattering. *Structure*, 15(6):693–705, June 2007.

[34] Paul Curnow, Annela M Seddon, and Paula J Booth. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666(1-2):105–117, November 2004.

[35] P Debye. X-ray dispersal. *Annalen Der Physik*, 46(6):809–823, 1915.

[36] P Debye. Molecular-weight determination by light scattering. *The Journal of physical and colloid chemistry*, 51(1):18–32, 1947.

[37] I G Denisov, Yelena V Grinkova, AA Lazarides, and SG Sligar. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *Journal Of The American Chemical Society*, 126(11):3477–3487, 2004.

[38] H Duan, Mary A Schuler, NR Civjan, and SG Sligar. Co-incorporation of heterologously expressed Arabidopsis cytochrome P450 and P450 reductase into soluble nanoscale lipid bilayers. *Archives Of Biochemistry And Biophysics*, 424(2):141–153, 2004.

[39] S Egelhaaf and P Schurtenberger. Micelle-to-Vesicle Transition: A Time-Resolved Structural Study. *Physical Review Letters*, 82(13):2804–2807, March 1999.

[40] S Eisenberg. High density lipoprotein metabolism. *Journal Of Lipid Research*, 25(10):1017, October 1984.

[41] Donald M Engelman, JF Hunt, PD McCrea, and G Zaccai. Assessment of the aggregation state of integral membrane proteins in reconstituted phospholipid vesicles using small angle neutron scattering. *Journal Of Molecular Biology*, 273(5):1004–1019, 1997.

[42] Salem Faham and James U Bowie. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *Journal Of Molecular Biology*, 316(1):1–6, February 2002.

[43] TM Forte, AV Nichols, EL Gong, S Lux, and RI Levy. Electron microscopic study on reassembly of plasma high density apoprotein with various lipids. *Biochimica Et Biophysica Acta-Biomembranes*, 248(2):381–386, 1971.

[44] Gérard Fournet. Fournet: Scattering functions for geometrical forms - Google Scholar. *Bull Soc Fr Minéral Cryst*, 74:39–113, 1951.

[45] Jens Frauenfeld, James Gumbart, Eli O van der Sluis, Soledad Funes, Marco Gartmann, Birgitta Beatrix, Thorsten Mielke, Otto Berninghausen, Thomas Becker, K Schulten, and Roland Beckmann. Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nature Structural & Molecular Biology*, 18(5):614–621, April 2011.

[46] T Gao, S Ly, S Fore, T A Laurence, J Voss, J Petrlova, and T Huser. Stoichiometry of Reconstituted High-Density Lipoproteins in the Hydrated State Determined by Photon Antibunching. *Biophysical Journal*, 101(4):970–975, 2011.

[47] R E Ghosh and A R Rennie. Assessment of detector calibration materials for SANS experiments. *Journal Of Applied Crystallography*, 32(6):1157–1163, 1999.

[48] Otto Glatter. A New Method for the Evaluation of Small-Angle Scattering Data. *Journal Of Applied Crystallography*, 10(OCT1):415–421, 1977.

[49] Otto Glatter and O Kratky. 1982.

[50] Yann Gohon, Tassadite Dahmane, Rob W H Ruigrok, Peter Schuck, Delphine Charvolin, Fabrice Rappaport, Peter Timmins, Donald M Engelman, Christophe Tribet, Jean-Luc Popot, and Christine Ebel. Bacteriorhodopsin/Amphipol Complexes: Structural and Functional Properties. *Biophysical Journal*, 94(9):3523–3537, May 2008.

[51] André Guinier and Gérard Fournet. *SMALL-ANGLE SCATTERING OF X-RAYS*. January 1955.

[52] Steen L Hansen. Calculation of small-angle scattering profiles using monte-carlo simulation. *Journal Of Applied Crystallography*, 23:344–346, 1990.

[53] Steen L Hansen. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *Journal Of Applied Crystallography*, 33:1415–1421, 2000.

[54] Steen L Hansen. BayesApp: a web site for indirect transformation of small-angle scattering data. *Journal Of Applied Crystallography*, 45(3):566–567, May 2012.

[55] SC Harvey, J P Segrest, M Hellinger, H De Loof, M K Jones, AE Klon, and CJ Sheldahl. A detailed molecular belt model for apolipoprotein A-I in discoidal high density lipoprotein. *Journal Of Biological Chemistry*, 274(45):31755–31758, 1999.

[56] William M Haynes. CRC Press, 2012.

[57] Heiko Heerklotz. Interactions of surfactants with lipid membranes. *Quarterly Reviews of Biophysics*, 41(3-4):205, December 2008.

[58] A Helenius and K Simons. Solubilization of membranes by detergents. *Biochimica Et Biophysica Acta*, 415(1):29–79, March 1975.

[59] Ralph Holl, Henry N Chapman, Petra Fromme, Anton Barty, Thomas A White, Richard Kirian, Andrew Aquila, Joachim Schulz, Daniel P Deponte, Uwe Weierstall, R Bruce Doak, Filipe R N C Maia, Andrew V Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L Shoeman, Sascha W Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sébastien Boutet, Michael J Bogan, Jacek Krzywinski, Christoph Bostedt, Saša Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, André Hömke, Christian Reich, Daniel Pietschner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kühnel, Marc Messerschmidt, John D Bozek, Stefan P Hau-Riege, Matthias Frank, Christina Y Hampton, Raymond G Sierra, Dmitri Starodub, Garth Williams, Janos Hajdu, Nicusor Timneanu, M Marvin Seibert, Jakob Andreasson, Andrea Rocker, Olof Jönsson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Kevin E Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M Holton, Thomas R M Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, and John C H Spence. Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332):73, February 2011.

[60] Ralph Holl, Richard Kirian, Garth Williams, M Marvin Seibert, Tomas Ekeberg, Filipe R N C Maia, Martin Svenda, Jakob Andreasson, Olof Jönsson, Duško Odić, Bianca Iwan, Andrea Rocker, Daniel Westphal, Max Hantke, Daniel P Deponte, Anton Barty, Joachim Schulz, Lars Gumprecht, Nicola Coppola, Andrew Aquila, Mengning Liang, Thomas A White, Andrew V Martin, Carl Caleman, Stephan Stern, Chantal Abergel, Virginie Seltzer, Jean-Michel Claverie, Christoph Bostedt, John D Bozek, Sébastien Boutet, A Alan Miahnahri, Marc Messerschmidt, Jacek Krzywinski, Keith O Hodgson, Michael J Bogan, Christina Y Hampton, Raymond G Sierra, Dmitri Starodub, Inger Andersson, Saša Bajt, Miriam Barthelmess, John C H Spence, Petra Fromme, Uwe Weierstall, Mark Hunter, R Bruce Doak, Stefano Marchesini, Stefan P Hau-Riege, Matthias Frank, Robert L Shoeman, Lukas Lomb, Sascha W Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Carlo Schmidt, Lutz Foucar, Nils Kimmel, Benedikt Rudek, Benjamin Erk, André Hömke, Christian Reich, Daniel Pietschner, Georg Weidenspointner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Ilme Schlichting, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kühnel, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, Henry N Chapman, and Janos Hajdu. Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature*, 470(7332):78, February 2011.

[61] Stephen Hyde, Z Blum, T Landh, S Lidin, B W Ninham, S Andersson, and K Larsson. The language of shape: the role of curvature in condensed matter: physics, chemistry and biology. 1996.

[62] Jacob N Israelachvili. *The Hydrophobic Effect*. Wiley, New York, second edition, 1980.

[63] Jacob N Israelachvili. *Intermolecular and Surface Forces*. Academic Press, London, 2011.

[64] B Jacrot. The study of biological structures by neutron scattering from solution. *Reports On Progress In Physics*, 1976.

[65] Morten Ø Jensen and OG Mouritsen. Lipids do influence protein function—the hydrophobic matching hypothesis revisited. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666(1-2):205–226, November 2004.

[66] Aleksandra Z Kijac, Ying Li, SG Sligar, and C M Rienstra. Magic-angle spinning solid-state NMR Spectroscopy of nanodisc-embedded human CYP3A4. *Biochemistry*, 46(48):13696–13703, 2007.

[67] B Kobilka, Søren G F Rasmussen, M R Whorton, Michael P Bokoch, Bo Huang, Richard N Zare, and R K Sunahara. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *PNAS*, 104(18):7682–7687, 2007.

[68] M Kotlarchyk and S H Chen. Analysis of small angle neutron scattering spectra from polydisperse interacting colloids — Browse - Journal of Chemical Physics. *The Journal of chemical physics*, 1983.

[69] Norbert Kucerka, Mu-Ping Nieh, and J Katsaras. Small-Angle Scattering from Homogenous and Heterogeneous Lipid Bilayers. pages 201–235. Elsevier, 2010.

[70] V V Kumar. Complementary molecular shapes and additivity of the packing parameter of lipids. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 88(2):444–448, January 1991.

[71] Daniel Lévy, Aline Bluzat, Michel Seigneuret, and J L Rigaud. A systematic study of liposome and proteoliposome reconstitution involving Bio-Bead-mediated Triton X-100 removal. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1025(2):179–190, June 1990.

[72] Ying Li, Aleksandra Z Kijac, SG Sligar, and C M Rienstra. Structural analysis of nanoscale self-assembled discoidal lipid bilayers by solid-state NMR spectroscopy. *Biophysical Journal*, 91(10):3819–3828, 2006.

[73] Z Li, JC Phillips, K Schulten, A Jonas, and W Wriggers. Predicting the structure of apolipoprotein A-I in reconstituted high-density lipoprotein disks. *Biophysical Journal*, 73(5):2337–2346, November 1997.

[74] P. Linder and Th. Zemb, editors. *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. Elsevier Science B. V., Amsterdam, 2002.

[75] Peter Lindner. Scatteringb experiments. In P. Linder and Th. Zemb, editors, *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, chapter 2. Elsevier Science B. V., Amsterdam, 2002.

[76] B Lorber, J B Bishop, and L J DeLucas. Purification of octyl β-D-glucopyranoside and re-estimation of its micellar size. *Biochimica Et Biophysica Acta-Biomembranes*, 1023(2):254–265, 1990.

[77] Mohsen Maleki, M Maleki, E Fried, and Eliot Fried. Stability of discoidal high-density lipoprotein particles. *Soft Matter*, 2013.

[78] Yves L Marcel and Robert S Kiss. Structure-function relationships of apolipoprotein A-I: a flexible protein with dynamic lipid associations. *Current opinion in lipidology*, 14(2):151–157, April 2003.

[79] Selma Maric, Nicholas Skar Gislinge, Sø ren Midtgaard, Mikkel Thygesen Boas, Jrgen Schiller, Henrich Frielinghaus, Martine Moulin, Michael Haertlein, Trevor Forsyth, Thomas Gnther Pomorski, and Lise Arleth. Stealth Carriers for Low-Resolution Structure Determination of Membrane Proteins in Solution. *Accepted in Acta Crystallographica D*, 2014.

[80] L O Martinez, S Jacquet, F Tercé, X Collet, B Perret, and R Barbaras. New insight on the molecular mechanisms of high-density lipoprotein cellular interactions. *Cellular and molecular life sciences : CMLS*, 61(18):2343–2360, September 2004.

[81] N A Mazer, W Känzig, and P Schurtenberger. Micelle to vesicle transition in aqueous solutions of bile salt and lecithin. *The Journal of Physical Chemistry*, 89(6):1042–1049, 1985.

[82] Mark A McLean, I G Denisov, AW Shaw, Yelena V Grinkova, and SG Sligar. Thermotropic phase transition in soluble nanoscale lipid bilayers. *Journal Of Physical Chemistry B*, 109(32):15580–15588, 2005.

[83] X Mei and D Atkinson. Crystal Structure of C-terminal Truncated Apolipoprotein A-I Reveals the Assembly of High Density Lipoprotein (HDL) by Dimerization. *Journal Of Biological Chemistry*, 286(44):38570, October 2011.

[84] P Mittelbach and Günter Porod. Zur Röntgenkleinwinkelstreuung verdünnter kolloider Systeme VI. *Acta Phys. Austriaca*, (14):405–439, 1961.

[85] M J Mohlenkamp. A fast transform for spherical harmonics. *Journal of Fourier analysis and applications*, 5(2):159–184, 1999.

[86] Christopher R Morgan, Christine M Hebling, Kasper D Rand, Darrel W Stafford, James W Jorgenson, and John R Engen. Conformational transitions in the membrane scaffold protein of phospholipid bilayer nanodiscs. *Molecular & Cellular Proteomics*, 10(9):M111.010876, September 2011.

[87] Martine Moulin, Joseph A Didonato, Matthew A Wagner, Zhiping Wu, Xavier Lee, Valentin Gogonea, X-M Li, Y Huang, A Undurti, Roland P May, M Haertlein, I Gutsche, G Zaccai, and S L Hazen. Double Superhelix Model of High Density Lipoprotein. *Journal Of Biological Chemistry*, 284(52):36605–36619, December 2009.

[88] Efstratios Mylonas and Dmitri I Svergun. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal Of Applied Crystallography*, 40:245–249, 2007.

[89] Minoru Nakano, Masakazu Fukuda, Takayuki Kudo, Masakazu Miyazaki, Yusuke Wada, Naoya Matsuzaki, Hitoshi Endo, and Tetsurou Handa. Static and Dynamic Properties of Phospholipid Bilayer Nanodiscs. *Journal Of The American Chemical Society*, 131(23):8308–8312, 2009.

[90] B W Ninham, Jacob N Israelachvili, and D J Mitchell. Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers. *Journal Of The Chemical Society-Faraday Transactions Ii*, 72:1525–1568, 1976.

[91] Olga V Nolandt, Torsten H Walther, Stephan L Grage, and Anne S Ulrich. Magnetically oriented dodecylphosphocholine bicelles for solid-state NMR structure analysis. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(5):1142–1147, May 2012.

[92] Julian Oberdisse, Peter Hine, and Wim Pyckhout-Hintzen. Structure of interacting aggregates of silica nanoparticles in a polymer matrix: small-angle scattering and reverse Monte Carlo simulations. *Soft Matter*, 3(4):476, 2007.

[93] Jan S Pedersen. Modelling of small-angle scattering from colloids and polymer systems. In P. Linder and Th. Zemb, editors, *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, chapter 16. Elsevier Science B. V., Amsterdam, 2002.

[94] Jan S Pedersen. Monte carlo simulation techniques applied in the analysis of small-angle scattering data from colloids and polymer systems. In P. Linder and Th. Zemb, editors, *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, chapter 15. Elsevier Science B. V., Amsterdam, 2002.

[95] M C Pedersen, L Arleth, and K Mortensen. WillItFit: a framework for fitting of constrained models to small-angle scattering data. *Journal Of Applied Crystallography*, 46(6):1894–1898, November 2013.

[96] Stephen J Perkins. Molecular modelling of human complement subcomponent C1q and its complex with C1r2C1s2 derived from neutron-scattering curves and hydrodynamic properties. *Biochemical Journal*, 228(1):13, May 1985.

[97] Jean-Luc Popot. Amphipols, Nanodiscs, and Fluorinated Surfactants: Three Nonconventional Approaches to Studying Membrane Proteins in Aqueous Solutions. *Annual Review of Biochemistry*, 79(1):737–775, June 2010.

[98] Gilbert G Privé. Detergents for the stabilization and crystallization of membrane proteins. *Methods*, 41(4):388–397, April 2007.

[99] V S Pureza, SG Sligar, J H Morrissey, and AW Shaw. The Local Phospholipid Environment Modulates the Activation of Blood Clotting. *Journal Of Biological Chemistry*, 282(9):6556–6563, December 2006.

[100] G Raffard, Gérard Raffard, Siegfried Steinbruckner, S Steinbruckner, A Arnold, Alexandre Arnold, James H Davis, J H Davis, and Erick J Dufourc. Temperature-composition diagram of dimyristoylphosphatidylcholine-dicaproylphosphatidylcholine "bicelles" self-orienting in the magnetic field. A solid state 2H and .... *Langmuir*, 2000.

[101] Salette Reis, Carla Guimarães Moutinho, Carla Matos, Baltazar de Castro, Paula Gameiro, and José L F C Lima. Noninvasive methods to determine the critical micelle concentration of some bile acid salts. *Analytical Biochemistry*, 334(1):117–126, November 2004.

[102] Amanda J Rice, Gerhard Wagner, Thomas Raschle, Sebastian Hiller, Thomas Walz, Tsyr-Yan Yu, and Tsyr-Yan Yu. Structural and Functional Characterization of the Integral Membrane Protein VDAC-1 in Lipid Bilayer Nanodiscs. *Journal Of The American Chemical Society*, 131(49):17777–17779, December 2009.

[103] J L Rigaud, Daniel Lévy, Gervaise Mosser, and Oliver Lambert. Detergent removal by non-polar polystyrene beads. *European Biophysics Journal*, 27(4):305–319, June 1998.

[104] TK Ritchie, J K Zolnerciks, SG Sligar, Yelena V Grinkova, Y V Grinkova, T H Bayburt, Timothy H Bayburt, I G Denisov, William M Atkins, and W M Atkins. Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. pages 211–231. Elsevier, 2009.

[105] Hiroyuki Saito. Alpha-Helix Formation Is Required for High Affinity Binding of Human Apolipoprotein A-I to Lipids. *Journal Of Biological Chemistry*, 279(20):20974–20981, March 2004.

[106] C R Sanders and R S Prosser. Bicelles: a model membrane system for all seasons? *Structure*, 6(10):1227–1234, October 1998.

[107] Dina Schneidman-Duhovny, Michal Hammel, and Andrej Sali. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic acids research*, 38(Web Server issue):W540–4, July 2010.

[108] J P Segrest, A Wlodawer, J N Weinstein, B Hong Chung, and R Chiovetti. High-density lipoprotein recombinants: evidence for a bicycle tire micelle structure obtained by neutron scattering and electron microscopy. *Febs Letters*, 104(2):231–235, August 1979.

[109] A Y Shih, I G Denisov, JC Phillips, SG Sligar, and K Schulten. Molecular dynamics simulations of discoidal bilayers assembled from truncated human lipoproteins. *Biophysical Journal*, 88(1):548–556, 2005.

[110] A Y Shih, Peter L Freddolino, Anton Arkhipov, and K Schulten. Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *Journal Of Structural Biology*, 157(3):579–592, 2007.

[111] A Y Shih, Peter L Freddolino, SG Sligar, and K Schulten. Disassembly of nanodiscs with cholate. *Nano Letters*, 7(6):1692–1696, 2007.

[112] A Y Shih, SG Sligar, and K Schulten. Maturation of high-density lipoproteins. *Journal of the Royal Society, Interface / the Royal Society*, 6(39):863–871, October 2009.

[113] Jan Skov Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Advances In Colloid And Interface Science*, 70:171–210, 1997.

[114] Jan Skov Pedersen, Cristiano L P Oliveira, Henriette Baun Hübschmann, Lise Arleth, Søren Manniche, Nicolai Kirkby, and Hanne Mørck Nielsen. Structure of immune stimulating complex matrices and immune stimulating complexes in suspension determined by small-angle x-ray scattering. *Biophysical Journal*, 102(10):2372–2380, May 2012.

[115] Jan Skov Pedersen, D Posselt, and K Mortensen. Analytical treatment of the resolution function for small-angle scattering. *Journal of applied ...*, 23(4):321–333, 1990.

[116] HM Smith, J Dickson, D Atkinson, and JP Austin. Interaction of apoprotein from porcine high-density lipoprotein with dimyristoyl lecithin .1. structure of complexes. *European Journal Of Biochemistry*, 64(2):541–547, 1976.

[117] Dmitri I Svergun. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal Of Applied Crystallography*, 25(4):495–503, 1992.

[118] Dmitri I Svergun. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal*, 76(6):2879–2886, June 1999.

[119] Dmitri I Svergun, C Barberato, and M H Koch. CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal Of Applied Crystallography*, 28:768–773, 1995.

[120] Dmitri I Svergun and L A Feĭgin. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Plenum Publishing Corporation, 1987.

[121] Dmitri I Svergun and M H Koch. Small-angle scattering studies of biological macromolecules in solution. *Reports On Progress In Physics*, 66(10):1782, 2003.

[122] Dmitri I Svergun, M H Koch, and Maxim V Petoukhov. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*, 80(6):2946–2953, June 2001.

[123] Dmitri I Svergun, Maxim V Petoukhov, P V Konarev, and A G Kikhney. ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied Crystallography*, 2007.

[124] Charles Tanford. Micelle shape and size. *Journal Of Physical Chemistry*, 76(21):3020–3024, 1972.

[125] Michael Thomas, Mary Sorci-Thomas, and Shaila Bhat. Three-dimensional models of HDL apoA-I: implications for its assembly and function. *Journal Of Lipid Research*, 49(9):1875–1883, September 2008.

[126] Elina Tjioe and William T Heller. ORNL_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes. *Journal Of Applied Crystallography*, 40(4):782–785, July 2007.

[127] F S Varley. Neutron scattering lengths and cross sectioirn. *Neutron News*, 3(3), 1992.

[128] B Vestergaard and Steen L Hansen. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *Journal Of Applied Crystallography*, 39:797–804, 2006.

[129] V V Volkov and Dmitri I Svergun. Uniqueness of ab initio shape determination in small-angle scattering. *Journal Of Applied Crystallography*, 36(3):860–864, 2003.

[130] M Wadsater, R Barker, Dimitrios Stamou, M Cardenas, Biger Lindberg Møller, A Singha, R Feidenhans'l, Nikos S Hatzakis, T Laursen, and K Mortensen. Monitoring Shifts in the Conformation Equilibrium of the Membrane Protein Cytochrome P450 Reductase (POR) in Nanodiscs. *Journal Of Biological Chemistry*, 287(41):34596–34603, October 2012.

[131] Gerhard Wagner, Franz Hagn, Thomas Raschle, and Manuel Etzkorn. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal Of The American Chemical Society*, page 130108043754005, January 2013.

[132] Hui Wang, Johannes Elferich, and Eric Gouaux. Structures of LeuT in bicelles define conformation and substrate binding in a membrane-like context. *Nature Publishing Group*, 19(2):212–219, January 2012.

[133] Minghan Wang and Michael R Briggs. HDL: the metabolism, function, and therapeutic importance. *Chemical Reviews*, 104(1):119–137, January 2004.

[134] H Wennerström and Björn Lindman. Micelles. Physical chemistry of surfactant association. *Physics Reports*, 52(1):1–86, 1979.

[135] Gerwin H Westfield, Søren G F Rasmussen, Min Su, Somnath Dutta, Brian T DeVree, Ka Young Chung, Diane Calinski, Gisselle Velez-Ruiz, Austin N Oleskie, Els Pardon, Pil Seok Chae, Tong Liu, Sheng Li, Virgil L Woods, Jan Steyaert, B Kobilka, R K Sunahara, and Georgios Skiniotis. Structural flexibility of the G alpha s alpha-helical domain in the beta2-adrenoceptor Gs complex. *Proceedings of the National Academy of Sciences*, 108(38):16086–16091, September 2011.

[136] SH White and WC Wimley. Membrane protein folding and stability: Physical principles. *Annual Review Of Biophysics And Biomolecular Structure*, 28:319–365, 1999.

[137] Muhammed A Yıldırım, Kwang-Il Goh, Michael E Cusick, Albert-László Barabási, and Marc Vidal. Drug—target network. *Nature Methods*, 25(10):1119–1126, October 2007.

Chapter 5

Paper 1

J|A|C|S
ARTICLES

Published on Web 09/09/2010

Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein

Nicholas Skar-Gislinge,[†] Jens Bæk Simonsen,[†] Kell Mortensen,[†] Robert Feidenhans'l,[‡] Stephen G. Sligar,[¶] Birger Lindberg Møller,[§] Thomas Bjørnholm,[‡] and Lise Arleth*,[†]

Biophysics, Faculty of Life Sciences, Nano Science Center, Faculty of Sciences, and Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, DK-1165 Copenhagen, Denmark, and Department of Biochemistry, University of Illinois at Urbana−Champaign, Champaign, Illinois 61820

Received April 12, 2010; E-mail: lia@life.ku.dk

Abstract: Phospholipid bilayers host and support the function of membrane proteins and may be stabilized in disc-like nanostructures, allowing for unprecedented solution studies of the assembly, structure, and function of membrane proteins (Bayburt et al. *Nano Lett.* **2002**, *2*, 853−856). Based on small-angle neutron scattering in combination with variable-temperature studies of synchrotron small-angle X-ray scattering on nanodiscs in solution, we show that the fundamental nanodisc unit, consisting of a lipid bilayer surrounded by amphiphilic scaffold proteins, possesses intrinsically an elliptical shape. The temperature dependence of the curvature of the nanodiscs prepared with two different phospholipid types (DLPC and POPC) shows that it is the scaffold protein that determines the overall elliptical shape and that the nanodiscs become more circular with increasing temperature. Our data also show that the hydrophobic bilayer thickness is, to a large extent, dictated by the scaffolding protein and adjusted to minimize the hydrophobic mismatch between protein and phospholipid. Our conclusions result from a new comprehensive and molecular-based model of the nanodisc structure and the use of this to analyze the experimental scattering profile from nanodiscs. The model paves the way for future detailed structural studies of functional membrane proteins encapsulated in nanodiscs.

Introduction

With the emergence of the field of synthetic biology,[2−5] which combines science and engineering to design and build novel biological functions and systems, it has become increasingly important to identify biological modules that can be assembled into new functional units at the nanoscale. In this context, membrane proteins, being exceptionally advanced nanoscale "machines", constitute a most important group of biomolecules because they carry important functions such as signal transduction, ion pumping, scaffolding, and photosynthetic capabilities. Traditional studies of membrane proteins are based on reconstitution in vesicles or surfactants, and entire pathways may be assembled into a single membrane anchored metabolon.[6,7] Recently, it has furthermore been shown how reconstitution is also possible in so-called nanodiscs.[8−17] Since the size of nanodiscs is compatible with that of typical membrane proteins, nanodiscs hold a very large potential as nanosample environments and allow for a whole new range of

[†] Biophysics, Faculty of Life Sciences, University of Copenhagen.
[‡] Nano Science Center, Faculty of Sciences, University of Copenhagen.
[¶] Department of Biochemistry, University of Illinois at Urbana−Champaign.
[§] Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen.

(1) Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G. *Nano Lett.* **2002**, *2*, 853–856.
(2) Purnick, P. E. M.; Weiss, R. *Nat. Rev. Mol. Cell. Bio.* **2009**, *10*, 410–422.
(3) Skerker, J. M.; Lucks, J. B.; Arkin, A. P. *Genome Biol.* **2009**, *10*, 114.
(4) Ghim, C.-M.; Kim, T.; Mitchell, R. J.; Lee, S. K. *Biotechnol. Bioprocess Eng.* **2010**, *15*, 11–21.
(5) Kwok, R. *Nature News* **2010**, *463*, 288–290.
(6) Jorgensen, K.; Rasmussen, A.; Morant, M.; Nielsen, A.; Bjarnholt, N.; Zagrobelny, M.; Bak, S.; Moller, B. L. *Curr. Opin. Plant Biol.* **2005**, *8*, 280–291.
(7) Nielsen, K. A.; Tattersall, D. B.; Jones, P. R.; Moller, B. L. *Phytochemistry* **2008**, *69*, 88–98.
(8) Bayburt, T. H.; Sligar, S. G. *Protein Sci.* **2003**, *12*, 2476–2481.
(9) Civjan, N.; Bayburt, T. H.; Schuler, M.; Sligar, S. G. *Biotechniques* **2003**, *35*, 548–554.
(10) Duan, H.; Civjan, N.; Sligar, S. G.; Schuler, M. *Arch. Biochem. Biophys.* **2004**, *424*, 141–153.
(11) Boldog, T.; Grimme, S.; Li, M.; Sligar, S. G.; Hazelbauer, G. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11509–11514.
(12) Leitz, A. J.; Bayburt, T. H.; Barnakov, A.; Springer, B.; Sligar, S. G. *Biotechniques* **2006**, *40*, 601–612.
(13) Nath, A.; Atkins, W. M.; Sligar, S. G. *Biochemistry* **2007**, *46*, 2059–2069.
(14) Whorton, M. R.; Bokoch, M. P.; Rasmussena, S. G. F.; Huang, B.; Zare, R. N.; Kobilka, B.; Sunahara, R. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7682–7687.
(15) Banerjee, S.; Huber, T.; Sakmar, T. P. *J. Mol. Biol.* **2008**, *377*, 1067–1081.
(16) Glueck, J. M.; Wittlich, M.; Feuerstein, S.; Hoffmann, S.; Willbold, D.; Koenig, B. W. *J. Am. Chem. Soc.* **2009**, *131*, 12060–12061.
(17) Yao, X. J.; Ruiz, G. V.; Whorton, M. R.; Rasmussena, S. G. F.; Devree, B. T.; Deupi, X.; Sunahara, R. K.; Kobilka, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 9501–9506.

10.1021/ja1030613

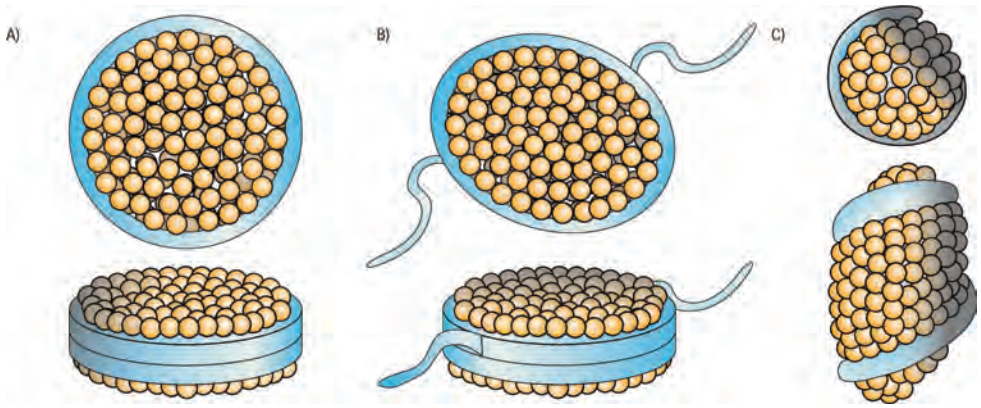


Figure 1. Different suggested nanodisc and HDL model structures, shown from the top and side. (A) Discoidal nanodisc with circular cross section.[24,25,39] (B) Discoidal nanodisc with elliptical cross section and protruding His-tags found in the present study. (C) Double superhelical Apo A1 with a prolate lipid core.[32]

functional and structural studies of membrane proteins in solution and at surfaces.[14,17] Nanodiscs are hence emerging as a versatile system for new fundamental studies of membrane proteins[18] and as potential new building blocks for chemical assembly of new supramolecular synthetic biological nanosystems and entire pathways. Following the original investigations of high density lipoproteins (HDLs),[19−22] the term "nanodisc" was coined by Sligar et al. in this millennium to underscore the possibility to exploit genetic engineering to synthesize and tailor-make such nanoscale objects.[1,23] Structural studies of the nanodiscs comprise small-angle X-ray (SAXS) and neutron scattering (SANS),[24−27] NMR,[16,28,29] atomic force microscopy,[1,30] and electron microscopy.[22,31] Based on these data, the current consensus—still under debate, however[32,33]—is that nanodiscs consists of a phospholipid bilayer surrounded by two amphipatic proteins both encircling the interior bilayer and resulting in a circular disc-like structure (Figure 1). The driving force for the assembly of the unit is the amphiphilic nature of the protein, favoring hydrophobic interactions between the inside of the protein and the alkyl chains of the phospholipid bilayer, resulting in an essentially all-hydrophilic surface exposed to the aqueous environment.[27] Depending on the structure of the phospholipid, the phospholipid bilayer is composed of 120−160 molecules. The current debate[32,33] about the different folding motifs of the amphipatic protein indicates, however, that structural methods so far are not good enough to unambiguously address the detailed structure of neither native HDL particles or engineered APO-A1-based systems including the nanodiscs. This in turn provides a serious obstacle for exploiting the nanodisc systems as a carrier of functional membrane proteins and as a platform to facilitate solution-based structural studies of them.

This has prompted us to embark on a comprehensive experimental and modeling study using SAXS from the native nanodiscs with the goal to enable SAXS and its sister technique, small-angle neutron scattering (SANS), to more routinely provide sufficiently accurate structural information to reveal structural details of membrane proteins embedded in the nanodiscs. In the following we show how a model that properly takes into account the contributions from the amphiphilic protein and from the hydrophobic and hydrophilic parts of the phospholipids provides new insights and improved structural accuracy. This study documents a pronounced elliptical shape of the nanodiscs not previously reported for fully loaded nanodiscs. The so-called twisted-belt model[33] is the only one suggested in the literature that allows an elliptical shape. Contrary to previous models used to analyze SAXS data from nanodiscs, our model design is completely based on the use of the molecular constituents as explicit building blocks of the nanodiscs. This implies that molecular constraints imposed by the phospholipids and the membrane scaffolding protein (MSP) belts, such as partial specific molecular volumes, molecular scattering lengths, and macroscopic sample concentrations, are implicit in the model. Thus, despite the relatively complex structure and potentially large structural parameter space, the theoretical solutions to the model fitting problem are automatically constrained to the much smaller parameter space which is spanned by the physically realistic structures. A similar approach

(18) Borch, J.; Hamann, T. *Biol. Chem.* **2009**, *390*, 805–814.
(19) Matz, C.; Jonas, A. *J. Biol. Chem.* **1982**, *257*, 4535–4540.
(20) Jonas, A. *Methods Enzymol.* **1986**, *128*, 553–582.
(21) Atkinson, D.; Smith, H.; Dickson, J.; Austin, J. *Eur. J. Biochem.* **1976**, *64*, 541–547.
(22) Forte, T.; Nichols, A.; Gong, E.; Lux, S.; Levy, R. *Biochim. Biophys. Acta: Biomembranes* **1971**, *248*, 381–386.
(23) Service, R. *Science* **2004**, *304*, 674–674.
(24) Nakano, M.; Fukuda, M.; Kudo, T.; Miyazaki, M.; Wada, Y.; Matsuzaki, N.; Endo, H.; Handa, T. *J. Am. Chem. Soc.* **2009**, *131*, 8308–8312.
(25) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A.; Sligar, S. G. *J. Am. Chem. Soc.* **2004**, *126*, 3477–3487.
(26) Denisov, I. G.; McLean, M.; Shaw, A.; Grinkova, Y. V.; Sligar, S. G. *J. Phys. Chem. B* **2005**, *109*, 15580–15588.
(27) Shih, A. Y.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. *Nano Lett.* **2007**, *7*, 1692–1696.
(28) Li, Y.; Kijac, A. Z.; Sligar, S. G.; Rienstra, C. M. *Biophys. J.* **2006**, *91*, 3819–3828.
(29) Gu, F.; Jones, M. K.; Chen, J.; Patterson, J. C.; Catte, A.; Jerome, W. G.; Li, L.; Segrest, J. P. *J. Biol. Chem.* **2010**, *285*, 4652–4665.
(30) Blanchette, C. D.; Cappuccio, J. A.; Kuhn, E. A.; Segelke, B. W.; Benner, W. H.; Chromy, B. A.; Coleman, M. A.; Bench, G.; Hoeprich, P. D.; Sulchek, T. A. *Biochim. Biophys. Acta: Biomembranes* **2009**, *1788*, 724–731.
(31) Ye, F.; Hu, G.; Taylor, D.; Ratnikov, B.; Bobkov, A. A.; Mclean, M. A.; Sligar, S. G.; Taylor, K. A.; Ginsberg, M. H. *J. Cell Biol.* **2010**, *188*, 157–173.
(32) Wu, Z.; Gogonea, V.; Lee, X.; Wagner, M. A.; Li, X.-M.; Huang, Y.; Undurti, A.; May, R. P.; Haertlein, M.; Moulin, M.; Gutsche, I.; Zaccai, G.; Didonato, J. A.; Hazen, S. L. *J. Biol. Chem.* **2009**, *284*, 36605–36619.
(33) Thomas, M.; Bhat, S.; Sorci-Thomas, M. *J. Lipid Res.* **2008**, *49*, 1875–1883.

has been applied previously by some of us for simpler self-assembling systems.[34,35]

To identify the roles of the phospholipids and the MSPs in the nanodiscs and to test the accuracy of the model, two different lipid systems were studied and compared. The first, dilaurylphosphatidyl choline (DLPC), has a secondary melting transition and therefore a relatively large intrinsic thermal expansion of the area per headgroup between 0 and 20 °C,[36,37] whereas the second, palmityloleylphosphatidyl choline (POPC), is in the fluid phase throughout the studied temperature range.[38] This gives it a constant and weak intrinsic expansion of its area per headgroup. Our temperature studies of these two nanodisc systems reveal a close agreement with the intrinsic thermal behavior of the lipids, in terms of both the overall expansion of the partial specific volumes of the lipids and the specific expansion of the area per headgroup. However, when confined in the discs, the DLPC is more laterally compressed and POPC is more laterally expanded than in large bilayer liposomes. Our data furthermore suggest that the MSP has an intrinsic tendency to form elliptically shaped discs. The shape of these discs is apparently determined by the temperature.

On the basis of our study we propose that, once the nanodisc is formed, the roles of the lipids and proteins are cast as follows: At any given temperature, the MSP determines the overall elliptical shape of the nanodisc and expands or compresses the area per headgroup of the phospholipids to minimize the hydrophobic mismatch between the bilayer alkyl chains and the hydrophobic inside of the MSPs. The MSP furthermore has sufficient flexibility to slightly adapt the circumference of the nanodiscs such that the lipids in the bilayer leaflet can expand laterally as a response to their intrinsic properties. While our findings are in good qualitative agreement with the consensus model for nanodiscs,[25,26,39] our proposed models are in contrast to the consensus model in that we suggest that maximally loaded nanodiscs do not necessarily have a circular cross section.

Theory and Mathematical Modeling

Our approach for modeling the nanodiscs takes a previously proposed model as its starting point.[25] However, the consequent use of molecular constraints in our approach allows for a much more detailed description than previously obtained. In general, the small-angle scattering intensity of a dilute isotropic suspension of particles can be expressed by

$$I(q) = nb^2P(q) \quad (1)$$

where n is the particle number density and b is the particle excess scattering length, which may also be expressed by $b = V\Delta\rho$. V is the single-particle volume, and $\Delta\rho$ is the average excess scattering length density of the particle as compared to the solvent. $P(q)$, the single-particle form factor intensity, is normalized such that $P(0) = 1$.

For the modeling, we describe the nanodiscs as elliptical discs, with a hydrophobic core region composed of the alkyl chains of the phospholipid bilayer sandwiched between the two hydrophilic

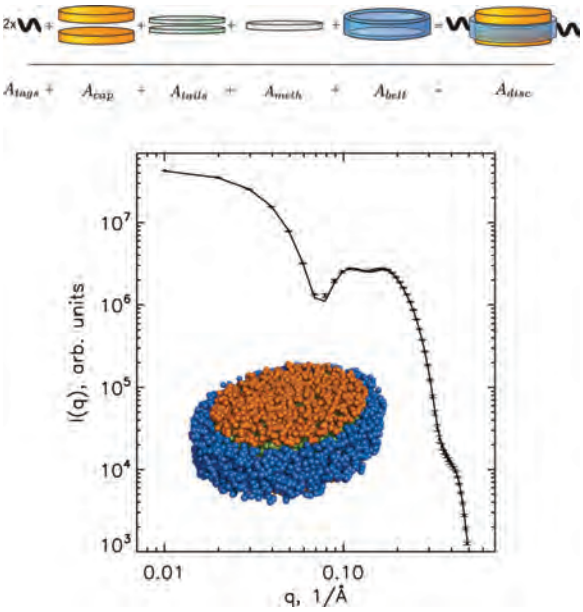


Figure 2. Top: Diagram of the principle for calculating the form factor amplitude of the His-tagged nanodiscs. Bottom: Comparison of the analytical expression for the form factor intensity as calculated from the direct-space model (shown) of the nanodisc structure (points) and from the analytical expression (full line).

headgroup regions. The hydrophilic headgroup regions are composed of the phosphatidyl choline headgroups and hydration water. In agreement with previous models for phospholipid bilayers and nanodiscs, the hydrophobic interior is separated into a central low-density methyl group region, sandwiched between two higher density leaflets composed of the $(CH_2)_n$ chains. The interior phospholipid disc is stabilized by a surrounding belt formed by the two amphiphilic MSPs. The MSP1D1 used in this experiment furthermore features a poly-histidine tag linked to the N-terminal region via a TEV cleavage site. In the following we refer to this 22 amino acid sequence as the His-tag. In the model we assume that this His-tag is protruding out from the belt and that we can describe it as a flexible chain in Gaussian random coil conformation (see Figure 2). Two coils are attached, one for each MSP belt. As the two His-tags have an unknown orientation with respect to each other, we have chosen to place the coils randomly along the rim of the hollow cylinder describing the MSPs.

The model form factor from the nanodiscs is constructed using the approach illustrated in Figure 2 and via the analytical form factor amplitude, $A_{disc}(a,b,h)$, of a short elliptical cylinder of height h and cross-sectional minor and major axes, a and b, as the basic building block.[40]

The short cylinder/disc form factor amplitude, $A_{disc}(a,b,h)$, is normalized such that $|A_{disc}(0)| = 1$. The normalized form factor amplitude of the entire disc (without His-tags) then becomes

$$A_{nanodisc}(q,\phi,\alpha) = (\Delta\rho_{belt}V_{belt}A_{belt} + \Delta\rho_{meth}V_{meth}A_{meth} + \Delta\rho_{tails}V_{tails}A_{tails} + \Delta V_{cap}\rho_{cap}A_{cap})/(\Delta\rho_{belt}V_{belt} + \Delta\rho_{meth}V_{meth} + \Delta\rho_{tails}V_{tails} + \Delta V_{cap}\rho_{cap}) \quad (2)$$

where the $\Delta\rho_i$ values denote the excess scattering length densities of the different constituents, respectively, and the V_i values denote

(34) Arleth, L.; Posselt, D.; Gazeau, D.; Larpent, C.; Zemb, T.; Mortensen, K.; Pedersen, J. S. *Langmuir* **1997**, *13*, 1887–1896.
(35) Arleth, L.; Ashok, B.; Onyuksel, H.; Thiyagarajan, P.; Jacob, J.; Hjelm, R. *Langmuir* **2005**, *21*, 3279–3290.
(36) Finegold, L.; Shaw, W.; Singer, M. *Chem. Phys. Lipids* **1990**, *53*, 177–184.
(37) Tada, K.; Goto, M.; Tamai, N.; Matsuki, H.; Kaneshina, S. *Chem. Phys. Lipids* **2008**, *153*, 138–143.
(38) Jimenez-Monreal, A.; Villalain, J.; Aranda, F.; Gomez-Fernandez, J. *Biochim. Biophys. Acta: Biomembranes* **1998**, *1373*, 209–219.
(39) Shih, A. Y.; Denisov, I. G.; Phillips, J.; Sligar, S. G.; Schulten, K. *Biophys. J.* **2005**, *88*, 548–556.
(40) Pedersen, J. S. Modelling of Small-Angle Scattering from Colloids and Polymer Systems. In *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*; Linder, P., Zemb, T., Eds.; Elsevier Science B.V.: Amsterdam, 2002; Chapter 16.

the volumes of the different building blocks as outlined in detail in the following section.

The different amplitude terms in eq 2 are derived using the standard approach.[40] As an example, the elliptical hollow cylinder, representing the MSP belt in eq 2, is given by the difference between the form factor amplitudes of two cylinders with different radii and axis ratios of the minor and major axes, a_i and b_i, but the same belt height h:

$$A_{\text{belt}} = A_{\text{disc}}(q, a_{\text{outer}}, b_{\text{outer}}, h) - A_{\text{disc}}(q, a_{\text{inner}}, b_{\text{inner}}, h) \quad (3)$$

The total nanodisc form factor amplitude in eq 2 is absolute squared and orientally averaged to yield the form factor intensity, $P(q)$, using the standard procedure.[40]

As indicated in Figure 2, the model is refined further by incorporating the His-tags described as flexible chains with Gaussian statistics. The final analytical form factor is derived following the same principles as outlined in refs 35, 41, and 42.

In order to independently verify the Fortran implementation of the relatively long and complex equations for the analytical form factor, we compared the outcome of the analytical form factor calculations for a given set of parameters to the small-angle scattering that would be obtained for a geometrical, direct-space model with exactly the same parameters. This small-angle scattering from the direct-space model was calculated as described in ref 43. The direct-space models, along with the corresponding small-angle scattering, calculated via these two different approaches are plotted and compared in Figure 2, and we observe a perfect agreement between the two.

Incorporation of Molecular Constraints. The mathematical model outlined in the previous section is described by a large number of fitting parameters. In order to reduce the degrees of freedom in the model and hence the number of free fitting parameters, we systematically incorporate molecular constraints in the model following the same principles as outlined in ref 34. As a part of this, we exploit the known sample concentration to analyze the data on an absolute scale.

Stoichiometry of the Nanodiscs. Each nanodisc is composed of two MSP belts (the MSP1D1s) and N phospholipids.[1,25] N is a fit parameter in the model, and due to the extremely low critical micellar concentration of the two phospholipids,[44] it is reasonable to assume that the stoichiometry remains constant once the nanodiscs are formed. Consequently, we use the same constant stoichiometry throughout the temperature scans.

Phospholipid Core. Our approach for modeling the phospholipid core is similar in spirit to the models used in the recent structural work on phospholipid bilayers.[45,46] However, we have chosen to use box models instead of Gaussians to model the scattering length density profile. For the modeling, the interior phospholipid bilayer of the nanodiscs is assumed to be separated into a hydrophilic headgroup region and a hydrophobic tail region. The hydrophobic tail region is further divided into a central electron-poor methyl group layer sandwiched between two hydrocarbon-chain-containing layers. The scattering lengths of these three parts of the molecule are determined from the molecular structure and partial specific volumes of the phospholipids.

Initially, we estimate the partial specific molecular volumes of POPC and DLPC to be $\nu_{\text{POPC}} = 1246$ Å^3 and $\nu_{\text{DLPC}} = 985$ Å^3. These values are determined from our densitometry measurements of MSP-free phospholipids in water taken at 20°C. (See further details in Supporting Information.) For the hydrophilic headgroups, a volume of $\nu_{\text{head}} = 319$ Å^3 is used,[47] and for the two methyl groups, a total volume of $\nu_{\text{methyl}} = 108.6$ Å^3 is used,[48] such that the volume of the hydrophobic tails becomes $\nu_{\text{pc}} - \nu_{\text{head}} - \nu_{\text{methyl}}$.

The model furthermore allows for including hydration water in the headgroup region, and we have the number of water molecules per PC headgroup, N_{w}, as an explicit fit parameter.

The fitted area per headgroup, A_{head}, combined with the determined molecular volume of the phospholipids and the fitted hydration number, N_{w}, fixes the total height of the phospholipids including the hydrophobic and hydrophilic heights. The total area of the disc is determined by the stoichiometry N. This area, combined with the fitted axis ratio ε, allows for determining the minor and major axes of the cross section of the lipid core; i.e., the lipid core is modeled with N, N_{w}, ε, and A_{head} as the only free parameters.

Membrane Scaffolding Protein Belt. The MSP1D1 consists of 211 amino acids. For the modeling we assume that 22 of these constitute an intrinsically unfolded His-tag and the remaining 189 constitute the belt, which is mainly folded into α-helical coils.[1]

The total scattering length of the MSP1D1 belt, b_{belt}, is proportional to the number of electrons in the belt via the Thompson scattering length. The number of electrons is counted from the amino acid composition of the MSP1D1. The molecular volume of the MSP1D1, ν_{belt}, is calculated using an average value for the partial specific mass density of proteins of 1.35 cm^3/g.[49] The scattering length density of the protein is then $b_{\text{belt}}/\nu_{\text{belt}}$. The inner minor and major axes are equal to those of the phospholipid core, so the only additional parameters required to describe the belt are the belt height, h_{belt}, and the radius of gyration, R_{g}, of the His-tag.

Resulting Model Fit Parameters. As a result of the molecular constraints, the resulting effective fit parameters are the number of phospholipids per nanodisc, N; the phospholipid area per headgroup, A_{head}; the axis ratio of the phospholipid disc, ε; the number of hydration water molecules, N_{w}; the height of the MSP belt, h_{belt}; and the radius of gyration of the His-tag, R_{g}. In addition to these six primary fit parameters, an effective Gaussian-type roughness term was included to account for the fact that the interfaces between the different constituents of the discs are not infinitely smooth. This term was included by multiplying the total form factor amplitude by $\exp(-(qS_{\text{R}})^2/2)$, where S_{R} is the surface roughness.

Three adjustment parameters were furthermore introduced: a small constant background to be added or subtracted from the data (this parameter should end up close to zero) and (2) scale factors for the partial specific volumes of, respectively, the MSP belt and the phospholipids. These two scale factors should be very close to unity and will account for small differences between the a priori estimated values and the actual values in the nanodiscs in a self-consistent molecular constrained way.

It should be emphasized that we, throughout the mathematical modeling process, have been very cautious to not include more fit parameters than necessary to obtain good and realistic model fits.

Results and Discussion

Initial Data Treatment and Visual Inspection. Figure 3 plots the experimental small-angle scattering from aqueous solutions of nanodiscs made from POPC and DLPC. The two sets of scattering data each have their characteristic features but share an overall behavior with an initial minimum at 0.05−0.07 1/Å followed by a broad peak. This characteristic scattering intensity is due to the complex scattering contrast situation in the

(41) Pedersen, J. S.; Gerstenberg, M. *Macromolecules* **1996**, *29*, 1363–1365.
(42) Pedersen, J. S. *J. Appl. Crystallogr.* **2000**, *33*, 637–640.
(43) Hansen, S. L. *J. Appl. Crystallogr.* **1990**, *23*, 344–346.
(44) Israelachvili, J. N. *Intermolecular and Surface Forces*; Academic Press: London, 1991.
(45) Kucerka, N.; Tristram-Nagle, S.; Nagle, J. F. *J. Membr. Biol.* **2005**, *208*, 193–202.
(46) Nagle, J.; Tristram-Nagle, S. *Biochim. Biophys. Acta: Biomembranes* **2000**, *1469*, 159–195.
(47) Sun, W.; Suter, R.; Knewtson, M.; Worthington, C.; Tristram-Nagle, S.; Zhang, R.; Nagle, J. F. *Phys. Rev. E* **1994**, *49*, 4665–4676.
(48) Tanford, C. *J. Phys. Chem.* **1972**, *76*, 3020–3024.
(49) Mylonas, E.; Svergun, D. I. *J. Appl. Crystallogr.* **2007**, *40*, 245–249.

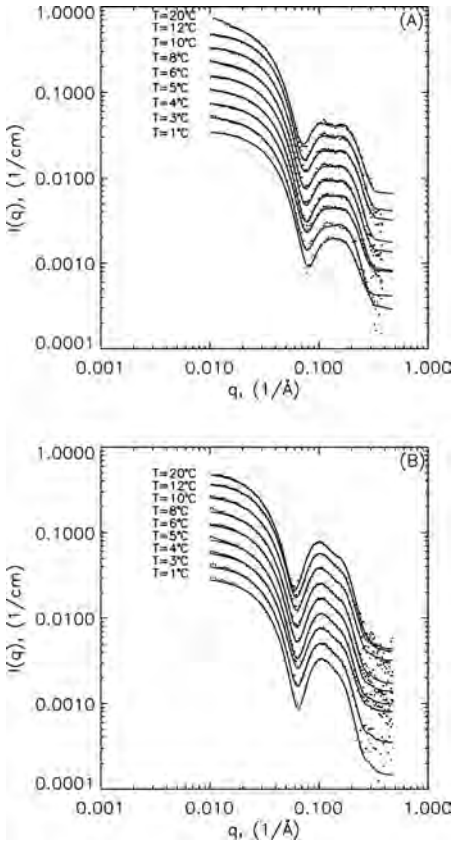


Figure 3. SAXS temperature scans for DLPC nanodiscs (A) and POPC nanodiscs (B). Experimental data points and model fits using the molecular constrained model illustrated in Figure 2. Starting from the lowermost temperature, the data are rescaled by 1.5^n, where n runs from 0 to 8.

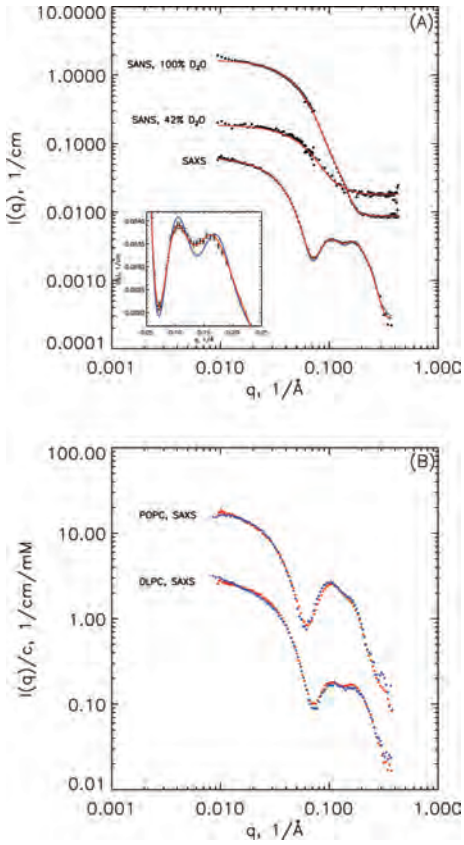


Figure 4. (A) DLPC nanodiscs ($T = 20$°C) measured by SANS and SAXS (black points) and fitted simultaneously with the same structural model for respectively circular (blue) and elliptical (red) nanodiscs. Inset: Zoom-in on the SAXS data with the two model fits. (B) Test of reproducibility of preparation and SAXS measurements of POPC and DLPC nanodiscs. Red points: data measured at Synchrotron SOLEIL. Blue points: data measured at the ESRF synchrotron on a another preparation of the same nanodiscs.

nanodiscs, with negative excess scattering length density for the hydrophobic alkyl chains and positive excess scattering length densities in both the hydrophilic phosphatidyl headgroup regions and the surrounding MSP belt. In both nanodisc systems, the broad peak at 0.1−0.2 1/Å splits into two distinct oscillations as temperature is increased. This shows that the nanodisc structure, including the nanodisc contrast situation, changes with temperature.

Figure 4A plots SAXS and SANS data measured on the same DLPC nanodisc sample. Due to the very different contrast situations for neutrons and X-rays, the same DLPC nanodiscs give rise to very different small-angle scattering curves in the two contrasts, thus emphasizing different features of the nanodisc structure. The DLPC nanodiscs are measured by SANS in the so-called bulk contrast obtained in 100% D_2O, as well as in 42% D_2O/58% H_2O (v/v), where the protein belts and the hydrophilic headgroups are, in practice, matched. These two SANS contrasts complement the SAXS data, which are, on the other hand, very sensitive to the fine balance between the electron-poor hydrophobic regions and the electron-rich hydrophilic regions. The use of three contrasts reduces the number of possible solutions to the model fitting problem significantly, increases the certainty of the structural model, and allows, in practice, for extracting even more detailed information.

In order to investigate the reproducibility of the nanodisc preparations, we measured SAXS data on several preparations of nanodisc samples. Typical examples of data obtained this way are shown in Figure 4B. It is clearly seen that the structure of the nanodiscs formed, as represented in terms of the SAXS data, is very reproducible.

Figure 5 shows examples of real-space representations of the small-angle scattering data in terms of the pair-distance distribution functions, $p(r)$, as determined via indirect Fourier transform of the scattering data.[50,51] The overall similarities between the POPC and DLPC nanodisc systems are obvious. The strong temperature dependence of the SAXS data, now represented in terms of the $p(r)$ functions, is also evident. In both cases, the D_{max} of the $p(r)$, i.e., the maximum dimension of the particles, reaches zero around 120 Å. The maximum dimensions of the nanodiscs formed in the two systems are therefore, as expected, quite similar and in both cases close to 120 Å. Furthermore, both systems exhibit a clearly visible tail starting at ∼100 Å and extending to 120 Å. This is fully consistent with our expectations of how a His-tag connected to the outside of the belt should appear in the $p(r)$ representation. We also observe that the broad peak initially centered around 60 Å splits into two broad peaks as temperature is increased from 1 to 20 °C;

(50) Glatter, O. *J. Appl. Crystallogr.* **1977**, *10*, 415–421.
(51) Pedersen, J. S.; Hansen, S. L.; Bauer, R. *Eur. Biophys. J.* **1994**, *22*, 379–389.

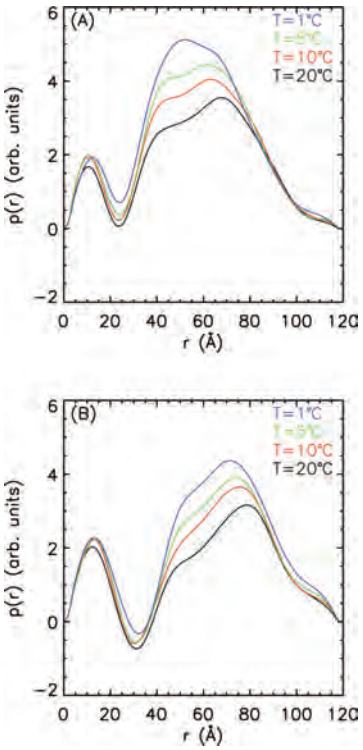


Figure 5. Examples of pair-distance distribution functions determined from the SAXS temperature scans of DLPC nanodiscs (A) and POPC nanodiscs (B).

i.e., the temperature-dependent "peak-splitting" is visible in both direct and reciprocal space. However, the detailed nanodisc structure is far too complex to allow for a model-free interpretation of this behavior based on a pure inspection of the obtained $I(q)$ and $p(r)$ data.

Model Fit Results. a. Testing the Hypotheses of the Structure of the Nanodiscs. So far, there has been consensus in the literature that fully loaded nanodiscs are circular,[14,25,26,39] and most small-angle scattering (SAXS and SANS) data from nanodiscs have been interpreted with this implicit assumption. Hence, the process of analyzing the scattering data in the present project naturally started with a model for circular discs. Using this approach, it was, however, quickly found that the scattering data could not be described satisfactorily. Figure 4A (blue curve) shows an example of a simultaneous fit of the circular model to a set of SANS and SAXS data obtained on the DLPC nanodisc system. While the general fit quality is decent, with a reduced χ^2 of 2.36, some clear and systematic deviations between model and experimental data are clearly observed. Because of the contrast situation, the deviations are most clearly observed in the SAXS data (see inset in Figure 4A), but they are also significant in the SANS data. These systematic deviations are a clear indication that the nanodiscs are either not circular or not monodisperse, or both.

In special cases, small-angle scattering data allow for separating dispersity in size and shape.[52,53] However, the small-angle scattering data obtained on the nanodisc system do not allow for such a separation; the model for monodisperse elliptical nanodiscs already gives perfect simultaneous fits to the SANS and SAXS data on the DLPC nanodiscs. This implies that a more complex model with a higher number of fit parameters cannot be used for extracting more detailed information from the data.

The structure of the MSP and the expected constant length of the MSP speak against a significant polydispersity and suggest that the shape deformation is the leading term. Consequently, the small-angle scattering data were analyzed in terms of a model for shape deformations in terms of elliptical nanodiscs. The red curve in Figure 4A is the best simultaneous fit obtained using this approach. The reduced χ^2 is 1.85, and there are no systematic deviations between model and data. The resulting model-fit parameters are listed in Table 1.

b. Fixing the Stoichiometry of the Nanodiscs. The absolute scaled SANS data (100% D_2O) are very sensitive to the stoichiometry of the nanodiscs, once the sample concentration is known. We have therefore determined the stoichiometry of the DLPC nanodiscs from the above-discussed simultaneous fit. This gives a stoichiometry of 76 DLPCs per MSP, corresponding to $N = 152 \pm 1$. This number lies close to the value obtained for DMPC nanodiscs,[54] and N was fixed to this value throughout the analysis of the DLPC system. For the POPC nanodiscs, the SAXS data were consistent with a stoichiometry of 61.5 POPCs per MSP, i.e., $N = 123 \pm 8$. This value is in good accordance the value of $N = 124 \pm 5$ that has previously been obtained,[1,25] and $N = 124$ was used throughout the analysis of the POPC data.

c. Model Fits to the SAXS Temperature Scans. The model fit results are plotted along with the experimental data in Figure 3. The obtained χ^2 values are all at or below unity, and as seen from the plots, the model fit quality is excellent and the model fits captures all significant features of the data.

The resulting fit parameters as well as deduced structural parameters of the nanodiscs at 1 and 20 °C are listed in Table 1. Furthermore, Figure 6 plots the temperature dependence of (A) the partial specific molecular volume, (B) the area per PC headgroup, (C) the axis ratios of the DLPC and POPC nanodiscs, and (D) the circumference of the MSP belts. All parameters are obtained from the model fits.

The observed partial specific molecular volumes for DLPC and POPC (see Figure 6A) are seen to increase slowly with temperature. This is the case both when determined directly from the densitometry measurements on plain and MSP-free samples of DLPC and POPC (multilamellar vesicle systems) and when determined from the model fits to the nanodisc samples. As clearly seen in the plot, the molecular volumes measured in these two ways are fully proportional to each other. For both DLPC and POPC, the values for partial specific molecular volumes are 2−3% higher in the MSP nanodiscs than in the MSP-free bilayer. With only two data series, it is not possible to tell whether the offset is a systematic result of a slightly different packing of the phospholipids when localized in multilamellar vesicles and nanodiscs, respectively, or whether the offset is simply an unsystematic result of small errors in the determination of the sample concentrations for the densitometry samples. However, in either case, the excellent agreement between the results obtained by these two independent approaches is a very strong proof of the self-consistency of the SAXS data analysis.

(52) Caponetti, E.; Floriano, M.; Didio, E.; Triolo, R. *J. Appl. Crystallogr.* **1993**, *26*, 612–615.
(53) Arleth, L.; Pedersen, J. S. *Phys. Rev. E* **2001**, *63*, 061406.
(54) Bayburt, T. H.; Sligar, S. G. *FEBS Lett.* **2010**, *584*, 1721–1727.

Table 1. Results from the Simultaneous Fits to SANS and SAXS Data on DLPC Nanodiscs at 20 °C (Marked N/X) and from the Fits to the SAXS Temperature Scan on DLPC and POPC Nanodiscs (Marked X) at 20 and 1 °C[a]

	DLPC, N/X	DLPC, X		POPC, X	
	20 °C	20 °C	1 °C	20 °C	1 °C
		Fit Parameters			
N	152 ± 1	152[b]	152[b]	124[b]	124[b]
ε	1.29 ± 0.02	1.26 ± 0.03	1.34 ± 0.04	1.31 ± 0.02	1.39 ± 0.03
A_{head}, Å^2	58.6 ± 0.2	58.8 ± 0.3	52.2 ± 0.3	70.6 ± 0.6	67.4 ± 1.2
N_w	0.0[b]	0.0[b]	0.0[b]	7.2 ± 1.0	13.6 ± 2.0
ν_{PC}, Å^3	997.5 ± 0.4	994.6 ± 0.7	965.8 ± 0.9	1277.3 ± 1.0	1254.5 ± 2.5
ν_{MSP}, Å^3	30217 ± 33	30658 ± 53	30954 ± 65	29851 ± 39	29410 ± 99
H_b, Å	24.0[b]	24.0[b]	24.0[b]	24.0[b]	24.0[b]
R_g, Å	12.7 ± 0.7	12.7[b]	12.7[b]	12.7[b]	12.7[b]
		Deduced Parameters			
H_{total}, Å	34.0 ± 0.1	33.8 ± 0.2	37.0 ± 0.2	42.7 ± 0.9	49.4 ± 1.9
$H_{hydrophobic}$, Å	23.0 ± 0.1	22.8 ± 0.2	24.2 ± 0.3	27.2 ± 0.2	27.7 ± 0.4
A_{disc}, Å^2	4408 ± 13	4428 ± 23	3924 ± 65	4314 ± 39	4193 ± 73
D_{minor}, Å	65.9 ± 0.7	66.9 ± 1.0	61.0 ± 1.6	65.1 ± 0.8	62.0 ± 1.4
D_{major}, Å	85.1 ± 1.4	84.3 ± 2.5	81.9 ± 2.7	85.1 ± 1.5	86.1 ± 2.3
d_{belt}, Å	8.5 ± 0.8	8.7 ± 1.4	9.1 ± 1.5	8.5 ± 0.9	8.5 ± 1.6

[a] Fitting parameters are those determined directly by the fits. Deduced parameters are derived from the fitted parameters as a consequence of the molecular constraints imposed on the model. N, number of phospholipids per nanodisc; ε, axis ratio of lipid core; A_{head}, area of the PC headgroup; N_w, hydration number; ν_{PC} and ν_{MSP}, partial specific molecular volumes of the phospholipid and MSP, respectively; H_b, height of the MSP belt; R_g, radius of gyration of the His-tags; A_{disc}, total area of the phospholipid part of the disc; H_{total}, total height of the phospholipid bilayer; $H_{hydrophobic}$, height of hydrophobic bilayer; D_{minor} and D_{major}, minor and major diameters of the phospholipid bilayer, respectively; and d_{belt}, thickness of the MSP belt. [b] Parameter not fitted.

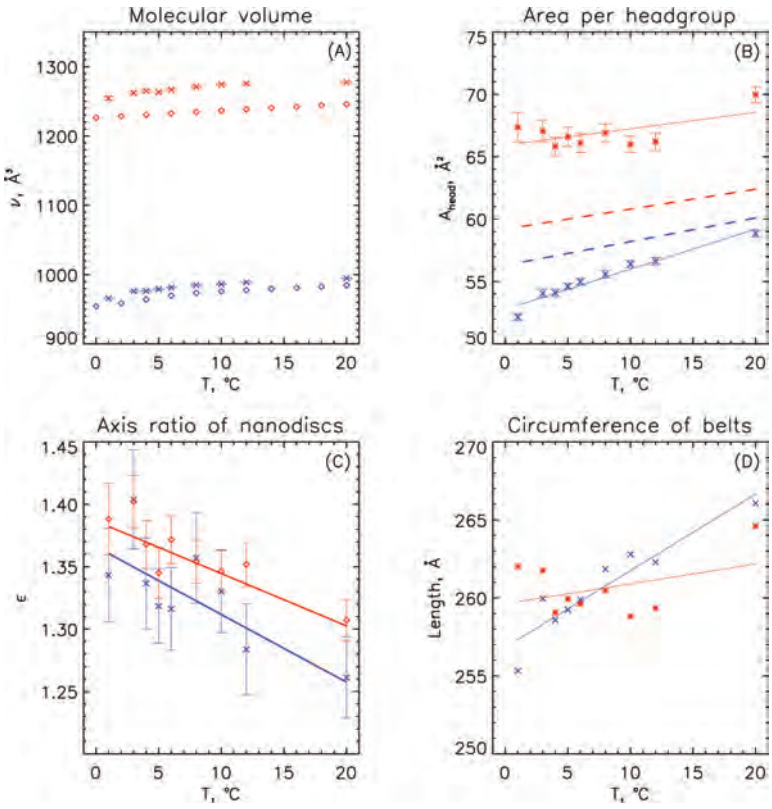


Figure 6. (A) Partial specific molecular volume, ν, plotted as a function of temperature for DLPC (blue) and POPC (red). Diamonds: data obtained from densitometry measurements. Crosses: data obtained from the SAXS analysis. (B) Temperature dependence of the molecular area per phospholipid headgroup. Same colors as in (A). Crosses: values obtained from the SAXS analysis. Full lines: fit of straight lines to the experimental data. Dashed lines: plots of the temperature dependence of the same lipids when localized in large bilayer vesicles. (C) Temperature dependence of the axis ratios of the nanodiscs. Same colors as in (A). Points: data determined from the SAXS analysis. Full lines: fit of straight lines to the experimental data. (D) Temperature dependence of the circumference of the nanodiscs. Same colors as in (A). Points: data determined from the SAXS analysis. Full lines: fit of straight lines to the experimental data.

The observed areas per headgroup of DLPC and POPC when localized in the nanodiscs are plotted in Figure 6B. In both cases, a temperature-dependent expansion is clearly observed. Straight lines fitted to the two sets of data give the following empirical

expressions for the area per headgroup as a function of temperature, T, measured in °C: $A_{DLPC, ND} = (52.8 + 0.32T)$ Å^2 and $A_{POPC, ND} = (65.9 + 0.13T)$ Å^2. It is relevant to compare the temperature-dependent expansion in the nanodiscs to that observed for phospholipids in "free" bilayers. Kucerka and coworkers reported the following temperature dependence:[55,56] $A_{DLPC} = (56.3 + 0.19T)$Å^2 and $A_{POPC} = (59.7 + 0.16T)$Å^2 (see plots in Figure 6B). The equations were obtained via simultaneous analysis of SANS and SAXS data on DLPC and POPC bilayers using a fully molecular constrained approach similar in spirit to the one used in the present article.[55] It should be emphasized that the equations were extrapolated from the temperature dependence observed around $T = 30$ °C and are only expected to be valid in the fluid phase, i.e., for $T > \sim 0°$ C in the case of POPC and for $T > \sim 10$ °C in the case of DLPC. It is clearly seen from the plots that DLPC has a smaller area per headgroup in nanodiscs than it has in free bilayers, whereas the POPC behaves oppositely and has a *larger* area per headgroup in nanodiscs than in free bilayers. Regarding the temperature-dependent lateral expansion, it is seen that POPC has a similar expansion of the areas per headgroup in nanodiscs and in "free" bilayers, while DLPC, where the lipids move through a secondary melting transition between 0 and 10 °C,[36,37] has, as expected, on average a larger expansion coefficient in nanodiscs in the studied temperature range than in the "free" bilayer data from the fully fluid phase above this melting transition. The total bilayer areas in the nanodiscs (see Table 1) are in very close agreement with results obtained using atomic force microscopy.[1]

Figure 6C plots the temperature dependence of the axis ratios, ε, of the two types of nanodiscs. In both systems ε decreases significantly with increasing temperature, implying that the discs become more circular with increasing temperature. Interestingly, the data suggest that there is no dependence between lipid type and disc shape; i.e., the larger temperature expansion of the DLPC than of POPC is not reflected in a corresponding larger decrease of the axis ratio.

Figure 6D plots the circumference of the belts as estimated by

$$P = 2\pi\sqrt{[(r_{minor} + d_{belt})^2 + (r_{major} + d_{belt})^2]/2}$$

In both systems the perimeter increases with temperature. However, the increase is significantly stronger for the DLPC nanodiscs than for the POPC nanodiscs. The stronger increase of the perimeter of DLPC nanodiscs is consistent with the larger area-expansion in this system in the studied temperature range. Assuming that the 189 amino acids in the belt part of the MSP make a chain of α-helical coils with a helical pitch of 1.5 Å per residue, the maximal extension of the belt becomes ∼284 Å. For both the DLPC and the POPC nanodiscs, the circumference is slightly lower than this value. This indicates that the belt is not fully stretched in the nanodiscs, which leaves some flexibility in the system for adapting to the increasing disc area.

The model fits are not very sensitive to the belt height when it is varied between 20 and 30 Å. Therefore, in the fitting process, it was decided to fix the belt height to 24 Å, which is compatible with the thickness of two α-helical chains. This poorly defined height is most likely a result of our attempt to model the curved α-helical belts with an effective edged hollow cylinder structure. Along this line, it is also relevant to address the questions about the conformation of the MSP belt and whether both belts surround the nanodiscs or they form half-circles or ellipses in a picket-fence-like structure.[13] Unfortunately, our present data do not contain information to answer this question.

The hydrophobic thickness values (see Table 1) are comparable for the DLPC and POPC nanodiscs, and both of them are relatively close to the belt height. At 20 °C we find a hydrophobic bilayer thickness of 24 Å for DLPC and 28 Å for POPC. Both systems show a slightly decreasing thickness for increasing temperature, directly related to the increasing area per headgroup. This finding suggests that the phospholipid bilayer adapts to the MSP belt to minimize the membrane/belt hydrophobic mismatch.[44,57]

The His-tags are clearly observed. They are revealed in the $p(r)$ functions as well as in the reciprocal space SAXS data, and a model term was included to explicitly take them into consideration such that sufficiently good model fits could be obtained. The simultaneous fit (Figure 4) gave a radius of gyration of $R_g = 12.7 \pm 0.7$ Å for the His-tags. No significant deviations from this value were found throughout the two temperature scans, and the final fits were performed with fixed $R_g = 12.7$ Å. This value, which is at the higher end of what we expected, most likely results from a combination of the relatively high rigidity of the His-tag and the low number of effective Kuhn lengths, such that the tag protrudes more or less perpendicular to the belt surface.

As mentioned previously, the number of hydration water molecules per PC headgroup is an explicit fit parameter in the model. When the data from the DLPC nanodiscs were fitted, we obtained, much to our surprise, values for the hydration number very close zero and sometimes even slightly negative. The latter situation is obviously unphysical, and in order to constrain the fits to physically realistic models, the hydration number was fixed to zero for the entire DLPC series. The very low hydration number may be a consequence of the lateral compression of the DLPC when localized in nanodiscs and may even suggest that the DLPC hydration water molecules are more densely packed (i.e., with a lower partial specific molecular volume) than bulk water molecules. However, the present data do not, unfortunately, allow us to investigate this effect in further detail.

In the POPC system we observe a decreasing hydration with increasing temperature. At 1 °C, we find that the POPC headgroups have 14 water molecules per PC headgroup, whereas at 20 °C, each POPC is associated with 7 water molecules. These numbers are in good accordance with the numbers reported in previous work,[45] and the temperature-dependent dehydration of the hydrophilic headgroups is fully consistent with what is observed in many other amphiphilic systems.[58]

General Discussion. The experimental SAXS data obtained from the nanodiscs are fully consistent with a model for elliptical disc-shaped nanodiscs. The SAXS-determined phospholipid partial specific volumes, including their thermal expansions, are in perfect agreement with the results of densitometry measurements and provide an internal self-consistency check of the SAXS analysis. While our observations do not explicitly rule

(55) Kucerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. *Biophys. J.* **2008**, *95*, 2356–2367.
(56) Personal communication cited with permission from N. Kucerka, Canadian Neutron Beam Centre, National Research Council of Canada, Chalk River Laboratories, Chalk River, Canada.
(57) Mouritsen, O.; Bloom, M. *Biophys. J.* **1984**, *46*, 141–153.
(58) Strey, R. *Colloid Polym. Sci.* **1994**, *272*, 1005–1019.

out a structure like the double-helical structure recently proposed by Wu and co-workers,[32] we find the elliptical disc model much more realistic because it allows the phospholipids to form a flat bilayer, in accordance with their close-to-unity packing parameter[44,59] and the relatively high bending rigidity of phospholipid bilayers, typically on the order of $10-20\ k_BT$.[44,60] Our finding of an elliptical lipid core is nevertheless in qualitatively good agreement with the results of Wu and co-workers.[32]

a. Elliptical versus Circular Shape of the Nanodiscs. So far, there has been consensus in the literature that nanodiscs are circular when they are fully loaded[14,25,26,39] but become elliptical/deformed when they are less than fully loaded.[39] This consensus follows from an intuitively tempting underlying model with a fixed disc circumference, defined by the MSP, which gives rise to an increasingly circular shape as more and more phospholipids are loaded into the nanodiscs. This underlying model implicitly defines the "fully loaded state" as the state that gives circular nanodiscs.

On this basis, it is natural to question to what extent the nanodiscs studied in the present project have been fully loaded. In the above-discussed intuitive model, our finding of elliptical discs would definitely imply that the discs are not fully loaded. However, the nanodiscs studied in the present work were carefully prepared through protocols that are optimized for maximal loading of the nanodiscs, the sample preparations appear very reproducible, as judged both by the size exclusion chromatograms and in particular by the much more sensitive SAXS analysis (Figure 4B).

Regarding the disc self-assembly from a statistical mechanics point of view, however, it is clear that the fully loaded state with a circular disc and a fully stretched belt is very unlikely. The argument is that there is only one way of loading the disc maximally to M phospholipids per nanodisc, while there will be M ways of loading the disc with $M-1$ phosholipids, $M(M-1)/2!$ ways of loading the disc with $M-2$ phosholipids, and, generally, $M!/N!$ ways of loading the disc with $M-N$ phospholipids. Thus, entropy will always favor the less than fully loaded conformations, and fully loaded discs are possible only in the limit of infinitely strong cooperativity. Another consequence of this line of argument is that, if the cooperativity is very strong, then the system will spontaneously microphase-separate into fully loaded and minimally loaded nanodiscs, such that fully loaded (nearly) circular discs will be formed independently of the initial ratio between phospholipids and belts, and such that optimizing the ratio will optimize the nanodisc yield but not affect the shape of the fully loaded fraction of assembled nanodiscs. This hypothesis lends experimental support to the original results in ref 1 that, by means of a combination of size exclusion chromatography and radioactively labeled phopspholipids, showed that the stoichiometry of the formed nanodiscs depends only weakly on the initial phospholipid-to-MSP ratio.

On the basis of these points and our experimental observations, we propose that, in experimental situations, maximally loaded nanodiscs along with APO-A1 discs will in practice always have an elliptical cross section. We therefore propose that our observation of elliptical nanodiscs is the rule rather than the exception to the rule. We plan to subject this hypothesis to a more careful test in an upcoming study.

b. Implications for Incorporation of Membrane Proteins into Nanodiscs. One of the central findings presented in this work is that the phospholipids localized in nanodiscs are slightly perturbed as compared to when they are localized in bilayer liposomes. We propose that this is driven by the MSP belt and the resulting minimization of the hydrophobic mismatch between the hydrophobic interiors of the MSP belt and the phospholipd bilayer.[57]

As the partial specific molecular volume of the phospholipid is virtually incompressible, the molecular area and hydrophobic height are inversely proportional to each other. Consequently, the molecular area decreases as the hydrophobic height increases as a response to the hydrophobic mismatch minimization and vice versa. For the DLPC, where the hydrophobic height is increased by the MSP, this results in a relatively smaller area per headgroup and also a very small value for the hydration number, which is apparently close to zero. This very small hydration number may affect the lateral packing of the PC headgroups. For the POPC, the hydrophobic height is decreased by the MSP, and this results in a significantly larger area per headgroup, whereas the observed hydration numbers (see Table 1) are comparable to those reported from phospholipid bilayer liposome systems.[45] A consequence of this finding is that the lateral pressure within the phospholipid bilayer environment will be different in nanodiscs and in bilayer liposomes. For DLPC, a higher lateral pressure will be obtained in nanodiscs as compared to in bilayer liposomes, whereas the opposite is the case for POPC.

That the phospholipids in nanodiscs are in a different state compared to bilayer liposomes is also supported by other experimental reports. From DSC measurements, the melting transition temperature of DMPC and DPPC is observed to shift toward higher temperatures,[26] whereas in a SANS study, the activation energy required to move one lipid into solution is decreased in nanodiscs.[24]

This effect may be relevant when using nanodiscs as a platform for investigating transmembrane proteins. It implies a relatively simple mechanism for the interplay between phospholipid type and functioning of the membrane protein and suggests that a detailed study should be conducted where the effect of phospholipid type is directly correlated with the membrane protein function.

Conclusion

A detailed and fully molecular constrained model for the small-angle scattering from nanodisc particles has been derived, and the model has been used to investigate the temperature dependence of DLPC and POPC nanodiscs and to identify the roles of phospholipids and MSP belts in nanodiscs. DLPC and POPC were chosen because they have very different intrinsic expansion coefficients of their areas per headgroup in the studied temperature range from 1 to 20 °C.

The nanodiscs were concluded to have an elliptical shape with the His-tags protruding out from the outer rim of the belts. The shape of the nanodiscs becomes more circular as temperature increases. Despite the different intrinsic expansion coefficients of DLPC and POPC, we do not observe significant differences in the temperature dependence of the shapes in the two systems. However, as a response to the relatively larger expansion of the DLPC, the belt perimeter increases more in the case of DLPC nanodiscs than in the case of POPC discs.

(59) Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 1525–1568.
(60) Yi, Z.; Nagao, M.; Bossev, D. P. *J. Phys.-Condens. Matter* **2009**, *21*, 155104.

The areas per headgroup of POPC and DLPC in nanodiscs differ significantly from the values observed in liposomes and "free" bilayers. POPC is expanded by the nanodiscs, while DLPC is squeezed. The POPC expansion and the DLPC squeezing drive the hydrophobic thicknesses of both lipids toward the height of the MSP belt. This shows that the phospholipids adapt their hydrophobic bilayer thickness and consequently their areas per headgroup to minimize the hydrophobic mismatch at the protein−lipid interface. This minimization of the hydrophobic mismatch may be the driving force for the effective squeezing of the DLPC and expansion of the POPC.

Altogether, our data and analysis clearly document that it is the protein belt that determines the shape and temperature dependence of the disc.

An important biological consequence of our findings of both the elliptical shape and the temperature dependence of the system is that it shows that the HDL particles are rather flexible particles that may adapt their circumference to increased or decreased load of phospholipids and cholesterol. In applications of nanodiscs as a platform for investigating membrane proteins, our findings suggest that the lateral pressure within the nanodiscs will depend strongly on the chosen phospholipid type. This will most likely feed back on the activity of the inserted membrane proteins. In a wider perspective, our detailed determination of the structure of the nanodiscs is a crucial step on the way to using nanodiscs as a platform for SAXS and SANS determinations of the low-resolution structure of membrane proteins.

Acknowledgment. The authors thank Dr. J. Perez and Dr. G. David at the SWING beamline at SOLEIL, the national French synchrotron facility, and we gratefully acknowledge the access to beamtime at SOLEIL. The authors thank Dr. Y. Cerenius and Dr. T. Plivelic at the I711 beamline at Max-lab, the national Swedish synchrotron facility, and we gratefully acknowledge the access to beamtime for pilot experiments at Max-Lab. The authors thank Dr. A. Round at the ID14-3 Bio-SAXS beamline at ESRF, the European Synchrotron Radiation Facility, and Dr. R. Schweins at the SANS instrument at beamline D11 at ILL, Institute Laue Langevin. We gratefully acknowledge the access to "simultaneous" SAXS and SANS beamtime at ILL and ESRF. The authors also thank Dr. T. Hamann for introducing us to the preparation of nanodiscs, Dr. N. Kucerka for sharing his unpublished results on POPC and DLPC bilayer membranes with us, and Dr. I. Denisov for enlightening discussion during the writing of the manuscript. Finally we thank the Danish Government funded UNIK Synthetic Biology program for co-funding of the project.

Supporting Information Available: Detailed account of the applied procedure for preparing the nanodiscs, report of the densitometry experiments, and experimental report of the SAXS/SANS experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

JA1030613

Supporting Information: Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein

Nicholas Skar-Gislinge,[†] Jens Bæk Simonsen,[†] Kell Mortensen,[†] Robert Feidenhans'l,[‡] Stephen G. Sligar,[¶] Birger Lindberg Møller,[§] Thomas Bjørnholm,[‡] and Lise Arleth[*,†]

Biophysics, Faculty of Life Sciences, University of Copenhagen, Denmark, Nano Science Center, Faculty of Sciences, University of Copenhagen, Denmark, Department of Biochemistry, University of Illinois at Urbana-Champaign, Illinois USA, and Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, Denmark

E-mail: lia@life.ku.dk

August 4, 2010

[*]To whom correspondence should be addressed
[†]Biophysics, Faculty of Life Sciences, University of Copenhagen, Denmark
[‡]Nano Science Center, Faculty of Sciences, University of Copenhagen, Denmark
[¶]Department of Biochemistry, University of Illinois at Urbana-Champaign, Illinois USA
[§]Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, Denmark

Experimental

Preparation of Nanodiscs:

Nanodiscs were prepared with the his-tagged membrane scaffolding protein MSP1D1 and, respectively, POPC (palmityl-oleoyl-phosphatidyl choline) and DLPC (Di-lauryl-phosphatidyl choline). The DLPC and POPC were acquired from Avanti Polar Lipids (AL, USA), whereas the MSP1D1 was provided by the Sligar lab.[1] The nanodiscs were prepared as previously described by Sligar et al:[2] An amount of phospholipid was dried and then quantified gravimetrically and finally resuspended in a 20mM Tris buffer at $pH = 7.4$ containing 100mM sodium cholate and 100mM NaCl to a final concentration of 50mM. The phospholipid stock solution was mixed with a MSP stock solution to reach the relevant lipid to MSP molar ratios. For DLPC this is 80:1 and for POPC it is 65:1 plus 5 % excess. Finally, a buffer with 100mM NaCl and 20mM Tris was added in order to achieve a final phospholipid concentration of $\sim$10mM. The final mixture was left to incubate for 1h at 4°C (just above t_m in both cases). Cholate was removed by addition 1g bio-beads (Bio-Rad) per ml of mixture and shaking gently for 3-8 h. The nanodiscs were purified using a Superdex-200 300/10 SEC column on an Äkta purifier. The chromatogram showed a peak at 12.5 ml characteristic for nanodiscs.[2] The peak was collected in 8 fractions. It was tested by SAXS that the 4 most concentrated fractions were similar and they were combined to maximize the yield. When needed, the samples were concentrated using spin-filters with a 100kDa cutoff.

Densitometry:

The partial specific molar volume of the phospholipids was measured using a vibrating tube densitometer (DMA 5000 from Anton Paar, Graz Austria). A stock of each phospholipid in milliQ water at 1.5% mass percent lipid was prepared. An amount of phospholipid was quantified gravimetrically (drying and weighing) and a small amount of radioactivity in the form of tritium labeled DMPC (DLPC samples) or POPC (POPC samples) was added. The initial ratio between lipid and β-radioactivity was then used to quantify the exact concentration of the further dilutions. The

stocks were then resuspended in milliQ water and heated and cooled between 30°C to 4°C ten times over night while gently shaken. To dissolve the last lipids, the solution was placed in an ultra sonic bath until the mixture was clear. The stock was then used to prepare three samples of each lipid at concentrations: 0.5%, 1% and 1.5% (mass). The density of the samples was measured in a temperature range from 20°C to 0°C with a two degree step. Before each measurement the sample was sonicated for additional 2 min. The measurements were repeated three times.
The three measurements together with a blank pure water sample were then used to extrapolate the density at a 0% lipid concentration and the partial specific molar volumes of DLPC and POPC were calculated. The result is shown as open diamonds in Figure 6 (A).

SAXS and SANS measurements:

SAXS measurements for the POPC and DLPC nanodisc temperature scans (Figure 3 and Figure 4 (B)) were performed at the Synchotron Soléil (Paris, France) at the SWING beamline. Each nanodisc sample was measured at temperatures between 20°C and 1°C and each measurement was performed by exposing the sample to 10 $\times$ 500 ms of radiation while flowing in order to minimize radiation damage. The samples were delivered to the measuring point using an adapted HPLC setup, enabling the samples to be stored at the measurement temperature. A combination of the wavelength λ of the incoming X-ray beam and the sample detector distance was chosen such that a q-range from 0.007 1/Å to 0.5 1/Å was covered. The scattering vector q is defined by $q = 4\pi \sin\theta/\lambda$, where θ is the half scattering angle and λ is the wavelength of the incoming X-ray beam.

The measured 10 $\times$ 500 ms blocks data were dezingered, averaged and background subtracted with the appropriate buffer backgrounds using the program BioXTAS RAW.[3] The absolute scale calibration was performed using H_2O and double checked against a lysozyme sample with a well-known concentration. This allows for absolute scale calibrating the data to an accuracy better than 10%. A dilution series of the two kinds of nanodisc was measured to check that concentration dependent effects could be neglected.

SANS and SAXS data sets for the simultaneous analysis (Figure 4) were acquired at, respectively, the SANS instrument D11 at Institut Laue Langevin (ILL) and at the Bio-SAXS instrument at beamline ID14-3 at European Synchrotron Radiation Facility (ESRF). Both facilities are located in Grenoble, France. For the SANS measurements a combination of two instrumental settings were used to obtain a sufficiently wide *q*-range. Absolute scale calibration was performed using H_2O according to the standard procedures at the facility. The samples were placed in cylindrical Hellma quartz cells during the measurements. For the SAXS measurements we used the fixed instrument setup including the HPLC-based automatic sample loading robot standard to the beamline ID14-3. Absolute scale calibration was performed using Bovine Serum Albumin as the external reference.

The SANS measurements were performed two days before the SAXS measurements. Ahead of the SANS measurements the usual H_2O-based buffer used in the nanodisc preparation, had been substituted for corresponding buffers with, respectively 100% D_2O and 42% D_2O/58% H_2O. In the latter contrast, the membrane scaffolding protein is contrast matched and in practice invisible in the experiments. The isotope substitution of the buffers was performed using centrifugal spin-filters with a cut-off of 100 kDa.

Notes and References

(1) http://sligarlab.life.uiuc.edu/nanodisc.html, **2010**.

(2) Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G. *Nano Lett.* **2002**, *2*, 853–856.

(3) Nielsen, S. S.; Toft, K. N.; Snakenborg, D.; Jeppesen, M. G.; Jacobsen, J. K.; Vestergaard, B.; Kutter, J. P.; Arleth, L. *J Appl Crystallogr* **2009**, *42*, 959–964.

Chapter 6

Paper 2

PCCP

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 3161–3170

www.rsc.org/pccp

PAPER

Small-angle scattering from phospholipid nanodiscs: derivation and refinement of a molecular constrained analytical model form factor†

Nicholas Skar-Gislinge and Lise Arleth*

Received 3rd July 2010, Accepted 1st November 2010
DOI: 10.1039/c0cp01074j

Nanodiscs™ consist of small phospholipid bilayer discs surrounded and stabilized by amphiphilic protein belts. Nanodiscs and their confinement and stabilization of nanometer sized pieces of phospholipid bilayer are highly interesting from a membrane physics point of view. We demonstrate how the detailed structure of Di-Lauroyl-Phosphatidyl Choline (DLPC) nanodiscs may be determined by simultaneous fitting of a structural model to small-angle scattering data from the nanodiscs as investigated in three different contrast situations, respectively two SANS contrasts and one SAXS contrast. The article gives a detailed account of the underlying structural model for the nanodiscs and describe how additional chemical and biophysical information can be incorporated in the model in terms of molecular constraints. We discuss and quantify the contribution from the different elements of the structural model and provide very strong experimental support for the nanodiscs as having an elliptical cross-section and with poly-histidine tags protruding out from the rim of the protein belt. The analysis also provides unprecedented information about the structural conformation of the phospholipids when these are localized in the nanodiscs. The model paves the first part of the way in order to reach our long term goal of using the nanodiscs as a platform for small-angle scattering based structural investigations of membrane proteins in solution.

Introduction

The quality of the experimental data obtained at modern SAXS and SANS facilities has improved steadily throughout the last decades. This is a natural result of the increase in both neutron and, in particular, X-ray flux as well as a result of a continuous specialization of both instruments and beam-line scientists. One of the consequences of this is that so-called "Bio-SAXS" instruments are being established at more and more large scale synchrotron X-ray facilities. A Bio-SAXS is characterized by having a relatively fixed experimental configuration which is optimized for weakly scattering biomolecules in solution. This includes for example samples placed in a fixed flow-through cuvette and unbroken vacuum from the beginning of the collimation section to the detector. The high quality data that may be obtained from these instruments opens up for extracting much more detailed information from SAXS and SANS data than previously possible. In order to make the most of these data, a similar development is therefore required for the development of the data analysis methods. With this article we wish to provide an example of such a method development.

Biophysics, Faculty of Life Sciences, University of Copenhagen, Denmark. E-mail: lia@life.ku.dk
† This article was submitted as part of a special collection on scattering methods applied to soft matter, marking the 65th birthday of Professor Otto Glatter.

We demonstrate that by combining data from state of the art SAXS and SANS instruments with a state of the art approach to the data analysis we can determine the solution structure of a complex biomolecular system in unprecedented detail. Besides from combining SAXS and SANS, the approach builds on the inclusion of so-called molecular constraints; *i.e.* the inclusion of information available from other sources, *e.g.* information about the chemical composition of the building blocks of the sample along with biophysical information such as sample concentration, partial specific densities of the single components *etc.*

The test case of this presented model development is the so-called phospholipid nanodisc system. Nanodisc's[1] are amphilic self-assembled particles consisting of an phospholipid bilayer disc which is stabilized by a "belt" constituted by two amphiphilic membrane scaffold proteins (MSP's). The amino-acid sequence of the MSP defines the length of the belt, which again fully controls the outer diameter of the formed nanodisc. Typical diameters of the nanodiscs are from 10 nm to 14 nm.[2,3] While the nanodisc system is highly interesting on its own right from a membrane physics point-of-view, the major broader interest in the system is related to the applicability and large potential of the system as a custom-fitted, nanometer-sized sample holder for membrane proteins.[4–6] The MSP, *i.e.* the protein belt, is inspired from the high density lipoprotein (HDL)-system, Apo-A1, and may be seen as a version of this that is genetically optimised towards the formation of circular discs.[2,3,7]

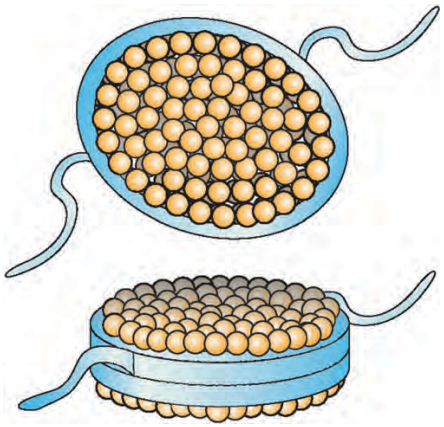


Fig. 1 Illustration of the nanodisc structure according to ref. 8. The cross-section is elliptical and the histidine-tags of the MSP's are visible on the outer rim of the belt.

In a recent publication[8] we have reported a detailed small-angle scattering study of nanodiscs formed with, respectively, Di-Lauroyl Phosphatidyl Choline (DLPC) and Palmitoyl-Oleoyl Phosphatidyl Choline (POPC) and in the temperature range from 1 °C to 20 °C. In the study we combined small-angle neutron scattering (SANS) and small-angle X-ray scattering (SAXS). We found an elliptical shape of the nanodiscs and showed that the histidine-tags of the MSP-belts are clearly visible in the nanodisc structure and protruding out from the outer rim of the MSP-belt that stabilizes the nanodiscs. The obtained structure for the nanodiscs is illustrated in Fig. 1. We showed that the MSP-belt and its temperature dependence is the main decisive factor for both the overall shape of the nanodiscs and for the packing of the phospholipids. We found that POPC becomes laterally expanded when localized in nanodiscs, leading to a relative thinning of the phospholipid membrane as compared to their structural state in unilamellar liposomes. Whereas the opposite is the case for the relatively short-chain DLPC, which becomes laterally compressed and have a slightly more thick hydrophobic bilayer in the nanodiscs. We propose that this, respectively the lateral expansion of POPC and the lateral compression of DLPC, is a consequence of the minimization of the hydrophobic mismatch[9] between the hydrophobic inside of the MSP-belt and the hydrophic sides of the encircled phospholipid bilayer disc.

The combined SANS and SAXS structural study of the nanodisc system is totally dependent on a sufficiently good and detailed structural model for the nanodiscs system. This includes the careful incorporation of molecular information about the molecular building blocks as obtained by complementary biophysical techniques, such as high precision densitometry and determination of sample concentrations. In the present article, we will therefore take the opportunity to explain the underlying model in full detail and evaluate the quantitative effect of the different elements of the model, both in terms of the differential scattering cross-sections, that are directly determined in a SAXS or SANS experiment, and in terms of the direct space pair-distance distribution functions that may be determined by indirect Fourier transform of the scattering data.[10]

We will show how the systematic incorporation of molecular information, in terms of so-called, molecular constraints, along with the simultaneous fitting of three small-angle scattering contrasts on the same sample, two SANS contrast and one SAXS contrast, allow us to minimize the number of free parameters in the structural model and use the combined set of experimental data to determine the structure of the nanodiscs in a so far, unprecedented detail. While cryo-TEM shows potential for providing structural information at comparable resolution,[11] small-angle scattering has the great advantage of being a solution based technique which requires very little prior sample preparation and which allows for easy accessible systematic investigations under a range of variable external conditions, including variable temperature, ion-strength and hydrostatic pressure, and which, in the case of synchrotron-SAXS, allows for ultra fast time-resolved studies.

Theory and mathematical modeling

General introduction to small-angle scattering from particles in solution

For an isotropic sample of particles in solution, the small-angle scattering experiment produces a centrosymetric scattering pattern. The elastically scattered intensity is usually described as a function of the scattering vector q, which is again related to the scattering angle, 2θ *via* the wave length of the incoming photon-beam (SAXS) or the neutron beam (SANS) the following way: $q = 4\pi \sin(\theta)/\lambda$. The small-angle scattering from particles in solution may then be described as the product of a form factor and a structure factor. According to commonly accepted nomenclature within small-angle scattering,[12] the form factor describes the intra-particle interference of the scattered photons or neutrons, and therefore contain information on the shape of the single particles, whereas the structure factor describes the inter-particle interference and contains information on particle–particle interactions within in the sample. For very dilute samples, the particle–particle interactions are negligible, the structure factor may be set to unity, and the q-dependent differential scattering cross section per unit volume becomes:

$$\frac{d\Sigma\ (q)}{d\Omega} = n(\Delta b)^2 P(q) \qquad (1)$$

where n is the particle number density, *i.e.* the number of particles per unit volume and Δb is the excess scattering length of the single particles. This entity is generally defined as $\Delta b = V \times (\rho - \rho_0)$, where V is the volume of the particle, ρ is the average scattering length density of the particle calculated as $\rho = \sum b_i / \sum v_i$, where the v_i's denote the volumes of the different parts of the particle and the b_i's denote the corresponding scattering lengths. ρ_0 is the average scattering length density of the solvent. More details on the calculation of Δb will be provided later. $P(q)$ is the form factor and describes the q-dependence of the scattering. In this presentation, $P(q)$ is normalized so that $P(0) = 1$. The form factor can in principle be calculated for any given shape *via*

the Fourier transformation of the excess scattering length density function, $\Delta\rho(\mathbf{r})$.

$$P(\mathbf{q}) = \frac{1}{\Delta b^2}\left|\int_v \rho(\mathbf{r})\mathrm{e}^{-i\mathbf{q}\cdot\mathbf{r}}\mathrm{d}\mathbf{r}\right|^2 = |\psi(\mathbf{q})|^2 \qquad (2)$$

where $\psi(\mathbf{q})$ denotes the so-called form factor amplitude. A large number of analytical and semi analytical form factors, for a wide range of geometrical objects are available in the literature.[13,14]

The above mentioned structure factor, on the other hand is somewhat more complicated to calculate and only exists as an analytical expression in a few cases. In this study, care has been taken to work in the dilute regime where the inter-particle interactions are weak and the structure factor effects become negligible.

The basic building block for the nanodisc form factor

Nanodiscs are expected to have a local cylindrical geometry and the form factor for nanodiscs was constructed using the form factor for a cylinder with an elliptical cross section as the basic building block:

$$\psi_{\mathrm{cyl}}(q,\alpha,\phi,r,L) = \frac{2J_1(qr(\phi)\sin(\alpha))}{qr(\phi)\sin(\alpha)}\frac{\sin(qL\cos(\alpha)/2)}{qL\cos(\alpha)/2} \qquad (3)$$

where J_1 is the first order Bessel function, α the azimuthal angle, L is the cylinder height and r the cylinder radius as a function of horizontal angle ϕ and the long and short half axis of the elliptical cross section, a and b: $r(\phi) = \sqrt{a^2\sin^2(\phi) + b^2\cos^2(\phi)}$ (see Fig. 2).

As illustrated in Fig. 3 the form factor of different objects may be combined to yield form factors of even more complicated objects, this is the basic principle used in this work.

The phospholipid interior. The phospholipid bilayer interior of the nanodisc is represented by a stack of cylinders with elliptical cross section where the top and bottom cylinders represent the hydrophilic phospholipid head group region. In between these two head group layers, the hydrophobic alkyl chain region is represented by a pair of cylinders corresponding to the top and bottom alkyl chain layer. A third layer representing the methyl end groups of the lipid tails is sandwiched in between these two alkyl chain layers. The scattering amplitude can be written as the sum of the form factor amplitudes for each cylinder weighted by their respective scattering lengths (see Fig. 4). The total scattering amplitude of this stack, A_{stack}, becomes:

$$A_{\mathrm{stack}} = A_{\mathrm{heads}} + A_{\mathrm{alkyl}} + A_{\mathrm{methyl}} = \Delta b_{\mathrm{heads}}\psi_{\mathrm{heads}} + \Delta b_{\mathrm{alkyl}}\psi_{\mathrm{alkyl}} + \Delta b_{\mathrm{methyl}}\psi_{\mathrm{methyl}} \qquad (4)$$

where the ψ's are the form factor amplitudes and the Δb's are the scattering lengths for the different units taken with respect to the aqueous solvent.

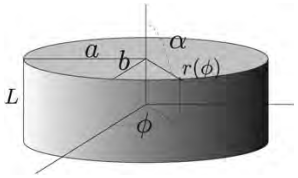


Fig. 2 Cylinder with an elliptical cross section.

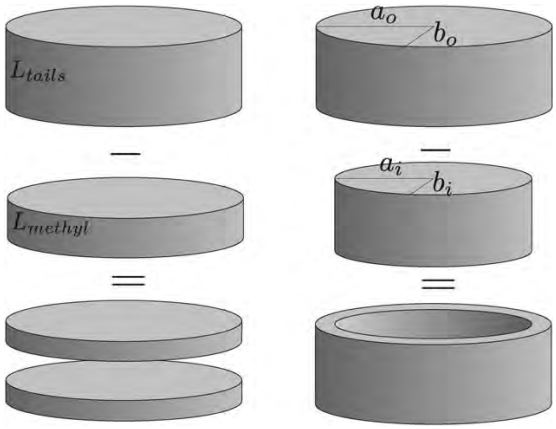


Fig. 3 Principle of combining form factors to yield more complex form factors.

The methyl groups in the center are represented by a single elliptical cylinder, with the form factor amplitude, ψ_{methyl}, calculated using eqn (3), where a and b is now the long and short axis of the lipid bilayer and L_{methyl} is the height of the methyl groups:

$$\psi_{\mathrm{methyl}} = \psi_{\mathrm{cyl}}(q,\alpha,\phi,r,L_{\mathrm{methyl}}) \qquad (5)$$

The alkyl chains of the lipids are represented by a cylinder with the same height as the tail groups, however with the volume corresponding to the methyl groups subtracted, this is sketched in Fig. 3. The form factor amplitude, ψ_{alkyl}, becomes:

$$\psi_{\mathrm{alkyl}} = \frac{(L_{\mathrm{tails}}\psi_{\mathrm{cyl}}(q,\alpha,\phi,r,L_{\mathrm{tails}}) - L_{\mathrm{methyl}}\psi_{\mathrm{methyl}})}{L_{\mathrm{alkyl}}} \qquad (6)$$

where $L_{\mathrm{alkyl}} = L_{\mathrm{tails}} - L_{\mathrm{methyl}}$ and where L_{tails} is the total height of the hydrophobic tails including the methyl group layer. The same approach is used to create form factor amplitude from the pair of cylinders representing the phospholipid head groups, ψ_{heads}: These are represented by a cylinder of the total height of the bilayer minus the combined volume taken up by the tails including the methyl groups:

$$\psi_{\mathrm{heads}} = \frac{(L_{\mathrm{total}}\psi_{\mathrm{cyl}}(q,\alpha,\phi,r,L_{\mathrm{total}}) - L_{\mathrm{tails}}\psi_{\mathrm{cyl}}(q,\alpha,\phi,r,L_{\mathrm{tails}}))}{L_{\mathrm{heads}}} \qquad (7)$$

where, $L_{\mathrm{heads}} = L_{\mathrm{total}} - L_{\mathrm{tails}}$. L_{total} is the total height of the bilayer. Note that r should be taken as a function of a, b and ϕ in eqn (7), eqn (6) and (7) as described in the comments to eqn (3).

The protein belt. The protein belt stabilizing the phospholipid bilayer is represented as a hollow cylinder. In order to calculate the corresponding scattering amplitude, A_{belt}, we combine the form factor amplitudes from two elliptical cylinders with a_o and b_o as the outer, and a_i and b_i as the inner half axes (See Fig. 3 (right)):

$$A_{\mathrm{belt}} = \Delta b_{\mathrm{belt}}\frac{L_{\mathrm{belt}}\pi(R_o^2\psi_{\mathrm{cyl}}(q,\alpha,\phi,R_o,L_{\mathrm{belt}}) - R_i^2\psi_{\mathrm{cyl}}(q,\alpha,\phi,R_i,L_{\mathrm{belt}}))}{V_{\mathrm{belt}}} = \Delta b_{\mathrm{belt}}\psi_{\mathrm{belt}} \qquad (8)$$

where R_o and R_i are the outer and inner radii. Again both are functions of the horizontal angle ϕ so that: $R_i(\phi) = \sqrt{a_i^2 \sin^2(\phi) + b_i^2 \cos^2(\phi)}$ and $R_o(\phi) = \sqrt{a_o^2 \sin^2(\phi) + b_o^2 \cos^2(\phi)}$. The pairs of axis ratios (a_i,b_i) and (a_o,b_o) are chosen such that $a_o - a_i = b_o - b_i$, such that we have approximately the same belt thickness for different ϕ-values.

Assembling the form factor for the nanodisc. As illustrated in Fig 4, the scattering amplitudes from the phospholipid bilayer and the protein belt can be combined to produce the total scattering amplitude of the nanodisc, A_{nd}, which using eqn (8) and (4) becomes:

$$A_{nd} = \Delta b_{nd}\psi_{nd} = \Delta b_{nd} \times \left(\frac{\Delta b_{\text{belt}}\psi_{\text{belt}} + \Delta b_{\text{heads}}\psi_{\text{heads}} + \Delta b_{\text{alkyl}}\psi_{\text{alkyl}} + \Delta b_{\text{methyl}}\psi_{\text{methyl}}}{\Delta b_{nd}}\right) \quad (9)$$

Referring to eqn (1), the differential scattering cross-section per unit volume is calculated as the absolute square of the scattering amplitude:

$$\frac{d\Sigma\ (q,\alpha,\phi)}{d\Omega} = n|A_{nd}|^2 = n\Delta b_{nd}^2 P_{nd}(q,\alpha,\phi) \quad (10)$$

where $P_{nd}(q,\alpha,\phi)$ is written as an explicit function of α and ϕ as this is the form factor of a model nanodisc with a certain orientation. In order to calculate the form factor of randomly oriented particles an orientational average of $P_{nd}(q,\alpha,\phi)$ has to be done:

$$P_{nd}(q) = \frac{2}{\pi}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} |\psi_{nd}|^2 d\phi \sin(\alpha) d\alpha \quad (11)$$

This orientational averaging has to be performed numerically.

Refinement of the form factor

As it will be discussed later, this basic model presented above does not describe the experimental data satisfactorily. Therefore the model is refined two ways: First, the protein belts that surround the phospholipid bilayer are known to have a poly histidine tag (his-tag) each. This his-tag is constituted by 22 amino acid (7 histidines and an amino acid spacer domain) and may in a first approximation be modeled as Gaussian random coils attached to the outer rim of the MSP-belts. Secondly, the interface roughness observed in real experimental systems is incorporated into the model, so that the infinitely sharp interfaces of the cylinder model are smoothened according to a Gaussian distribution for the surface roughness.

Inclusion of the scattering from the His-tags. As will be discussed later and as demonstrated in Fig. 5 and 7 it is clear that the protruding his tags are visible in the data and have to be taken explicitly into account. The normalized form factor of the basic nanodisc model with two his-tags in the form of Gaussian random coils with excess scattering lengths Δb_c attached to the outer rim of the belt can be written as the sum of the following four terms:[15–17]

$$P(q) = \frac{(\Delta b_{nd}^2 P_{nd} + 2^2 \Delta b_c \Delta b_{nd} S_{nc} + 2\Delta b_c^2 S_{cc} + 2\Delta b_c^2 P_c)}{\Delta b_{\text{total}}^2} \quad (12)$$

where P_{nd} is the basic nanodisc model form factor (disc-disc self correlation), S_{nc} describe the interference between the nanodisc and the tags (disc-chain cross correlation), S_{cc} describes the interference between the tags (chain-chain cross

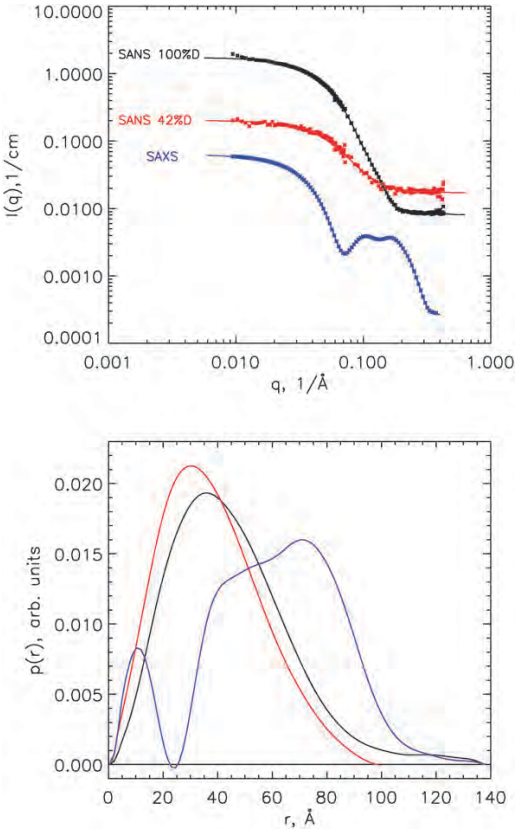


Fig. 5 (Top): SANS and SAXS data from the DLPC nanodiscs. SANS data are measured both in 100% D_2O and in 42% D_2O. (Bottom): The corresponding $p(r)$-functions.

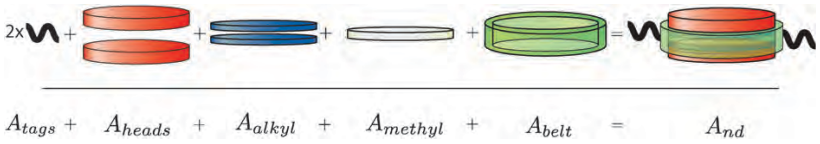


Fig. 4 Principle of assembling the entire nanodisc form factor.

correlation) and P_c describes the contributions from the single chain (chain-chain self correlation).

P_c is the form factor for a Gaussian random chain:[18]

$$P_c(q) = \frac{2(e^{-R_g^2q^2} + R_g^2q^2 - 1)}{R_g^2q^2} \quad (13)$$

where R_g is the radius of gyration of the Gaussian random chains.

Since the exact positions of the his-tags are unknown, and since we do not expect to have the structural resolution to locate them, we simply assume that they are uniformly distributed on the outer rim of he nanodisc belts. Under this assumption, the chain–chain and the disc-chain correlation terms of the Gaussian random chains attached to the disc become:[16]

$$P_{cc}(q,\alpha,\phi) = \left[\frac{1 - e^{-R_g^2q^2}}{R_g^2q^2}\right]^2 \times \left[J_0(q(R_o(\phi) + R_g)\sin(\alpha))\frac{\sin\left(q\frac{L}{2}\cos(\alpha)\right)}{q\frac{L}{2}\cos(\alpha)}\right]^2 \quad (14)$$

$$S_{nc}(q,\alpha,\phi) = \left[\frac{1 - e^{-R_g^2q^2}}{R_g^2q^2}\right] \times \left[J_0(q(R_o(\phi) + R_g)\sin(\alpha))\frac{\sin\left(q\frac{L}{2}\cos(\alpha)\right)}{q\frac{L}{2}\cos(\alpha)}\right]\psi_{nd} \quad (15)$$

where $R_o(a,b,\phi) = \sqrt{a_o^2\sin^2(\phi) + b_o^2\cos^2(\phi)}$ and L is the height of the belt. Note that, in eqn (14) and eqn (14), we have chosen to displace the center of mass of the his-tags a distance of R_g away from the outer rim of the belt. This is done in order to minimize the nonphysical overlap-clashes between the his-tags and the belt, *i.e.* to secure that space is only filled once.

As usual, the final form factor has to be orientational averaged over ϕ and α:

$$P(q) = \frac{1}{\Delta b_{\text{total}}^2}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}(\Delta b_{nd}^2P_{nd} + 4\Delta b_c\Delta b_{nd}S_{nc} + 2\Delta b_c^2S_{cc} + 2\Delta b_c^2P_c)\mathrm{d}\phi\sin(\alpha)\mathrm{d}\alpha \quad (16)$$

Inclusion of the interfacial roughness of the nanodiscs. The infinitely sharp interfaces of the analytical model results in some nonphysical artifacts at the high q part of the scattering curve. In order to remedy this, an average Gaussian distributed interface roughness is included in the model.[19] This is done by multiplying the obtained the final differential scattering cross-section by a Gaussian function: $P_{\text{final}}(q) = P(q)e^{-(q\sigma_i)^2/2}$ where σ_i is a measure for the interface roughness. It is important to note that this roughness parameter is an average roughness of all interfaces in the model and therefore can not be directly related to a specific interface. The importance of this term is clearly illustrated in Fig. 6(c) and turned out to be absolutely crucial to obtain good simultaneous fits to the SANS and SAXS data.

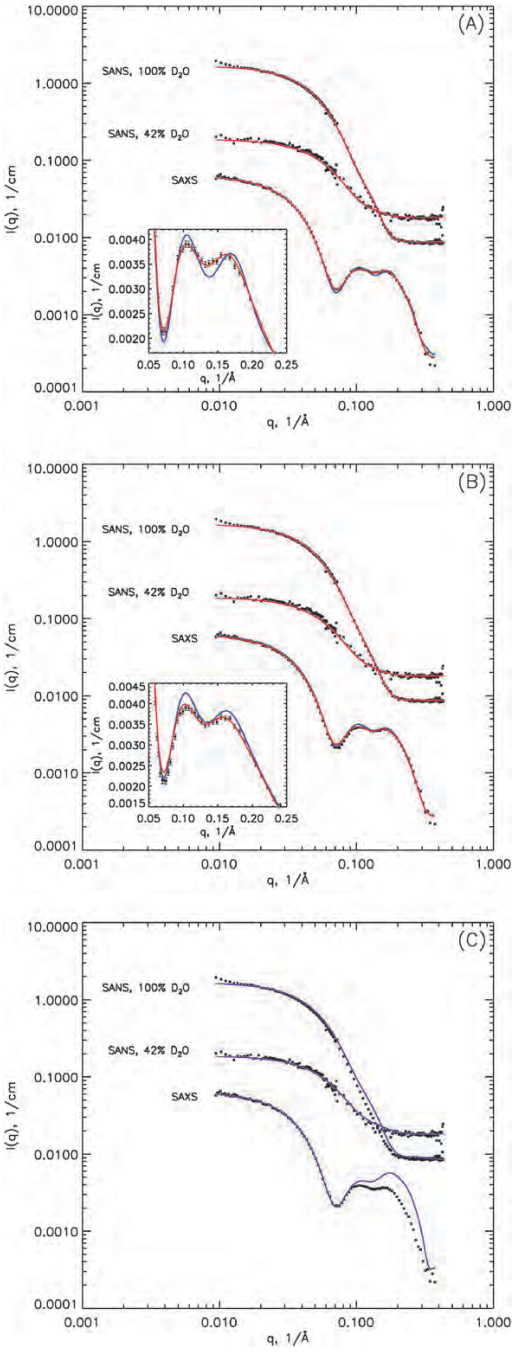


Fig. 6 Refinement of the modelling for the DLPC nanodiscs. Points: experimental data (see caption to Fig. 5). Full curves: The different iterations of the model for the DLPC-nanodiscs. (A) Effect of elliptical (red) *versus* circular (blue) nanodiscs. (B) Effect of not including the protruding His-tags. Blue: No his-tag but otherwise same model parameters as (red) elliptical model in (A). Red: Best fit with model without His-tag. (C) Effect of not including the surface roughness in the model.

Approach for the incorporation of molecular constraints

The underlying idea of including molecular constraints is to minimize the number of free parameters in the structural model and constrain the model fits to the relatively small subset of models that are physically realizable from the known molecular building blocks and which are furthermore consistent with the combined knowledge about the absolute scale of the different sets of scattering data along with other biophysical information, such as *e.g.* sample concentration and composition.

The particle number density. The particle number density is directly proportional to the protein concentration. As mentioned previously, nanodiscs consist of a protein belt and an interior phospholipid bilayer. It is generally accepted that the nanodisc protein belt consists of 2 MSP proteins[2] and a number of phospholipids which varies with the phospholipid type. The MSP concentration, in terms of moles per liter, M, may be accurately determined by a simple UV-absorption measurement for a given sample. Assuming two MSP's per nanodisc, the particle number density, n, is then easily calculated as $n = M/2 \cdot N_{\mathrm{lipid}}$, where N_{lipid} is the number of phospholipids per nanodisc. The lipid-concentration is notoriously more elaborate to determine with high accuracy from standard biophysical methods. In the present work this value has therefore been taken as a fit parameter *via* the fitting of the lipid-to-belt stoichiometry, N_{lipid}, of the single nanodiscs.

The phospholipid bilayer. In our model the phospholipid bilayer is modeled as a vertically symmetric stack of pairs of elliptically shaped cylindrical discs, where each cylindrical disc represents a different region of the phospholipid bilayer. The total volume of each pair of cylinders may be directly related to the partial specific molecular volume of the corresponding part of the phospholipid *via* N_{lipid}. Likewise, the total scattering length of each pair of cylindrical discs is directly calculated from the scattering length of the corresponding part of the phospholipid molecule times N_{lipid}. The scattering lengths and partial specific molecular volumes required for this calculation are listed in Table 1.

For example, the two methyl groups at the end of the hydrophobic phospholipid tails are estimated to have a partial specific molecular volume of $v_{\mathrm{meth}} = 108$ Å^3,[20] giving a total methyl volume of the methyl bilayer of $V_{\mathrm{methyls}} = N_{\mathrm{lipid}} v_{\mathrm{meth}}$. As each methyl group contributes to the scattering with 9 electrons, the total scattering length (for SAXS) of the cylinder representing the methyls is $b_{\mathrm{methyls}} = N_{\mathrm{lipid}} \cdot 9 r_0$, where r_0 is the Thompson scattering length. The excess scattering length of the methyl groups, $\Delta b_{\mathrm{methyls}}$, is $(b_{\mathrm{methyls}}/V_{\mathrm{methyls}} - \rho_0) \times V_{\mathrm{methyls}}$ which may also be calculated on the molecular level: $(9r_0/v_{\mathrm{methyls}} - \rho_0) \times v_{\mathrm{methyls}}$. From this we note that the excess scattering length density is dependent of the partial specific molecular volumes.

As a first estimate, the total partial specific molecular volume of the DLPC phospholipids in the bilayer is set to $v_{\mathrm{dlpc}} = 985$ Å^3. This value was obtained from high precision densitometry measurements involving radioactively labelled phospholipids for the accurate determination of sample concentration.[8] In order to quantify the molecular volume of the different parts of the DLPC we used a previously reported molecular volume of the PC head group of $v_{\mathrm{pc}} = 319$ Å^3.[21] The molecular volume of the alkyl chains then becomes: $v_{\mathrm{alkyl}} = v_{\mathrm{dlpc}} - v_{\mathrm{pc}} - v_{\mathrm{meth}}$.

The phospholipid head groups are hydrophilic and the model allows for the incorporation of a number of hydration water molecules associated with each PC group, N_w. These hydration water molecules will contribute to the volume of the cylinders representing the phospholipid heads, but not to the excess scattering length as they are assumed to have the same scattering length density as the solvent.

The DLPC area per head group, A_{head}, (*i.e.* the mean surface area taken up by one phospholipid molecule in the bilayer plane) is also taken as a fit parameter. From the combination of A_{head} and the partial specific molecular volume of the different molecular constituents of the phospholipid, the height of each part of the phospholipid lipid bilayer disc is easily calculated.

The axis ratio of the bilayer disc, $\varepsilon = a/b$ (see Fig. 2), is also taken as a fit-parameter.

The size and shape of the phospholipid bilayer is then determined by only four parameters: The aggregation number N_{agg}, the number of hydration water molecules per lipid N_w, the area per phospholipid A_{lip} and the cross-section axis ratio ε.

The membrane scaffolding protein belt. Our *a priori* estimate for the molecular volumes of the two scaffolding protein belts surrounding the phospholipid bilayer, is calculated using an average mass density of proteins of 1.356 cm^2/g.[22] The inner shape of the hollow cylinder describing the protein belt is constrained to the same shape as the phospholipid bilayer it surrounds. As the volume is fixed, the height of the belt, h_{belt}, is the only other parameter needed to describe the protein belt. The volumes of the two his-tags (22 amino acids each) is calculated using the same mass density and only their radius of gyration, R_g is a free fit parameter. The scattering length of the protein belt including the his-tags is determined from the MSP amino acid composition. And, using a similar approach as described above, the excess scattering length of the MSP becomes $\Delta b_{\mathrm{MSP}} = (b_{\mathrm{MSP}}/v_{\mathrm{mSP}} - \rho_0) - v_{\mathrm{msp}}$.

Adjustments to the molecular constraints. Because of the complex contrast situation in the SAXS experiment, with positive and negative excess scattering length densities in different domains of the nanodiscs, the model is extremely sensitive to minor deviation of the partial specific molecular volumes from the *a priori* values. Therefore adjustment parameters to be multiplied to the partial specific molecular volume of the DLPC and MSP were introduced. These parameters also allows the study of fine but experimentally anticipated thermal expansion or contraction when heating or cooling the sample. In any case, however, these volume correction-factors should be close to unity.

As a final free parameter, a constant can be added to the model scattering in order to correct for errors in the background subtraction, this parameter should be close zero.

Table 1 List of initial scattering lengths per molecule and partial molecular specific volumes used as molecular constraints

	Formula	b_{SAXS} [cm]	b_{SANS} [cm] in 100% D_2O	b_{SANS} [cm] in 42% D_2O	v [Å^3]
MSP belt	—	3.34×10^{-09}	8.27×10^{-10}	6.50×10^{-10}	27 588
MSP tag	—	3.87×10^{-10}	9.58×10^{-11}	7.52×10^{-11}	3194
DLPC head	$C_{10}H_{18}NO_8P$	4.62×10^{-11}	7.05×10^{-12}	7.05×10^{-12}	319
DLPC alkyl	$C_{20}H_{40}$	4.51×10^{-11}	-1.67×10^{-12}	-1.67×10^{-12}	557
DLPC methyl	C_2H_6	5.08×10^{-12}	-9.15×10^{-13}	-9.15×10^{-13}	108.6
H_2O	H_2O	2.82×10^{-12}	-1.68×10^{-13}	-1.68×10^{-13}	30.0
D_2O	D_2O	2.82×10^{-12}	1.91×10^{-12}	1.91×10^{-12}	30.0

Results and discussion

Visual inspection of the experimental scattering intensities and pair-distance distribution functions

DLPC-nanodiscs were measured by SANS and SAXS in two SANS contrast situations, respectively, 100% D_2O buffer and 42% D_2O/58% H_2O buffer, as well as in the ordinary SAXS contrast obtained in 100% aqueous buffer. An example of the obtained scattering data along with the corresponding pair-distance distribution functions, $p(r)$, are plotted in Fig. 5. The sample concentrations for the SAXS sample and SANS sample in 100% D_2O sample was 3.2 mg/ml. This corresponds to a nanodisc concentration of 23.0 μM. The SANS sample in 42% D_2O had a sample concentration of 5.5 mg/ml, corresponding to 38.5 μM nanodiscs.

The 100% D_2O SANS contrast is a bulk contrast with significant negative excess scattering length densities for both protein belt and phospholipid core. The protein belt is matched out to zero excess scattering length density in the 42% D_2O SANS contrast, making only the phospholipid core visible. Both SANS data set exhibit a monotonically decaying differential scattering cross-section. In the SAXS contrast situation, the hydrophobic part of the phospholipid core has negative excess scattering length density, whereas both the phospholipid headgroups and the protein belt has positive excess scattering length densities. This inhomogeneous contrast with both positive and negative contributions to the scattering amplitudes gives rise to the very significant oscillating behavior of the SAXS scattering intensity comparable to what is *e.g.* observed in surfactant micelle systems.

The $p(r)$ functions were calculated using an implementation of Glatters original IFT-procedure.[10] We used a Fortran implementation of the routine (originally programmed by Prof. Jan Skov Pedersen) that had a slightly modified smoothness constraint.[23] In the $p(r)$ plot in Fig. 5, the $p(r)$-functions are normalized such that the area under the $p(r)$-function becomes unity.

The $p(r)$-function of the SANS 100% D_2O contrast has a slightly skewed, but otherwise bell-shaped $p(r)$-function. A tail is clearly visible at r-values in the interval from 100 Å to 140 Å. This behaviour is consistent with our expectations to the shape of the nanodiscs as being small compact discs with his-tags protruding out from the outer rim of the belts that stabilize the sides of the discs. The $p(r)$-function of the SANS 42% D_2O contrast has a similar shape as that of the SANS 100% D_2O contrast. However, in 42% D_2O, both the maximum value of the $p(r)$ and the D_{max} are clearly shifted to lower values, and in particular the tail at high r-values is not present. This is all a result of the contrast matching of the protein belt, which is then no longer visible. Instead the smaller hydrophobic part of the lipid bilayer almost exclusively dominates the scattering pattern and the $p(r)$-function. In the SAXS contrast, where, as mentioned above, an oscillating differential scattering cross-section is obtained as a result of the combination of the negative excess scattering length density in the hydrophobic core and positive excess scattering length density in both the hydrophilic head-groups and the protein-belt, we also obtain a non-monotonic behavior of the $p(r)$-function. The D_{max} is similar to that obtained in the SANS 100% D_2O contrast and again we observe a tail at high r-values which is most likely a signature of the protruding his-tags. Due to the totally different SAXS contrast situation, the remaining part of the $p(r)$-curve differs significantly from those obtained in the other two contrast situations. The comparison between the three illustrates very clearly the value of combining SANS and SAXS.

Refinement of the structural model for the nanodiscs

The data from the two SANS contrasts and the single SAXS contrast were fitted simultaneously by minimizing the reduced χ^2 using the model for nanodiscs with elliptical cross-section with protruding his-tags in a Gaussian random coil conformation described in section "Theory and mathematical Modeling". As shown in Fig. 6(A) an excellent agreement ($\chi^2 = 1.85$) is obtained between model and experimental data when using a model for nanodiscs with an elliptical cross-section and protruding his-tags in a Gaussian random coil conformation, whereas significant systematic deviations between model and data are obtained along with a higher reduced χ^2 of 2.36 is obtained when the nanodiscs are forced into the previously suggested consensus model for nanodiscs with a circular cross-section.[2,3,24]

As a part of the model-refinement process several more simple models were investigated. The effect of including the his-tags explicitly as protruding Gaussian random coils attached to the outer rim of the MSP-belts is investigated in Fig. 6(B). The blue curve is a model calculation using the same model parameters as in the elliptical model fit (red curve) shown in Fig. 6(A) except that the his-tag is now omitted and the corresponding mass instead assumed evenly distributed in the protein belt. Without any fitting this gives rise to large and systematic deviations and an increase of the χ^2 to 5.94. After fitting of the remaining parameters we obtained the red curve of Fig. 6(B) which is the best fit that may be obtained with the model without his-tag. The over-all fit is good ($\chi^2 = 2.71$), however, significant

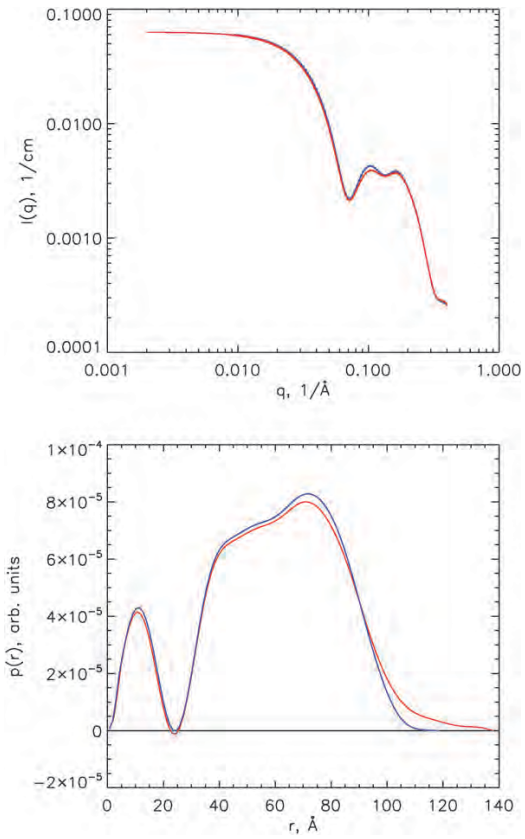


Fig. 7 (Top): Analytical SAXS differential scattering cross-sections for the DLPC nanodiscs. Red: Model for nanodiscs with protruding his-tags. Blue: Model for nanodiscs without his-tags. (Bottom): The corresponding $p(r)$-functions (same colour-code).

systematic deviations around the first minimum in the SAXS curve are clearly visible.

To further elucidate the effect of the his-tags on the differential scattering cross-sections, the model calculations for the nanodiscs, with and without his-tag, are plotted in Fig. 7 without the experimental data. Only the model calculation corresponding to the SAXS contrast is plotted here. As seen, the effect of removing the his-tag is visible in the scattering data. However, the effect is difficult to interpret from the differential scattering cross-sections alone. This situation changes in the $p(r)$-representation of the analytical scattering cross-section (also plotted in Fig. 7), where the effect of the protruding his-tags is very clearly seen as a significant tail at high r-values. This clearly illustrates the advantage of plotting small-angle scattering data both in their reciprocal space and direct space representations. It also gives strong support to the interpretation of the comparable tails observed in the $p(r)$ functions determined from our experimental data Fig. 5.

In Fig. 6(C) the effect of *not* including the roughness term is shown. The blue curve is the small-angle scattering calculated for the same model parameters as for the elliptical nanodiscs (red curve) in Fig. 6(A) however the surface roughness is set to zero corresponding to infinitely smooth interfaces between the different inner and outer layers of the nanodisc. While this model is clearly nonphysical it also gives rise to a really poor agreement between model and data. Clear deviations between model and data already sets in before $q = 0.001$ 1/Å and the over-all χ^2 increases to 100. It was attempted to obtain a better fit. However, while a significantly lower χ^2 could be obtained (lowest χ^2 was about 25), large systematic deviations remained clearly visible and the fit did not converge satisfactorily.

Table 2 Results from the simultaneous fits to SANS and SAXS data on DLPC nanodiscs at 20 °C. The "Fitting parameters" are the parameters determined directly by the fits. "Deduced parameters" are derived from the fitted parameters as a consequence of the molecular constraints imposed on the model

Fit parameters:	DLPC, N/X T = 20 °C
N_{DLPC}	150 ± 1
ε	1.29 ± 0.02
A_{head}, Å^2	58.6 ± 0.2
N_w	0.0[a]
ν_{DLPC}, Å^3	997.5 ± 0.4
ν_{MSP}, Å^3	30217 ± 33
H_b, Å	24.0[a]
R_g, Å	12.7 ± 0.7
$\sigma_{roughness,X}$, Å	4.24 ± 0.04
$\sigma_{roughness,N}$, Å	4.58 ± 0.18
Deduced parameters:	
H_{total}, Å	34.0 ± 0.1
$H_{Hydrophobic}$, Å	23.0 ± 0.1
A_{disc}, Å^2	4408 ± 13
D_{minor}, Å	65.9 ± 0.7
D_{major}, Å	85.1 ± 1.4
d_{belt}, Å	8.5 ± 0.8

[a] Parameter not fitted. N_{DLPC}: The number of DLPC's per nanodisc, ε: Axis ratio of lipid core, A_{head}: Area of the PC head group, N_w: Hydration number, ν_{PC}: Partial specific molecular volumes of the phospholipid. ν_{MSP}: Partial specific molecular volume of the MSP (*i.e.* the sum MSP belt and MSP his-tag). Note that H_b: Height of the MSP-belt, R_g: Radius of gyration of the his tags, $\sigma_{roughness,X}$: Average interface roughness as observed in the SAXS contrast, $\sigma_{roughness,N}$: Average interface roughness as observed in the SANS contrast. A_{disc}: Total area of the phospholipid part of the disc. H_{Total}: Total height of the phospholipid bilayer. $H_{Hydrophobic}$: Height of hydrophobic bilayer. D_{minor} and D_{major}: Minor and major diameter of the phospholipid bilayer. d_{belt}: Thickness of the MSP-belt.

The fit parameters corresponding to the best fit, plotted in Fig. 6a (red), are summarized in Table 2. In the upper half of the table, the actual fit parameters are listed. Whereas in the lower half, those structural parameters are listed that, due the systematic use of molecular constraints, are deduced from the fit parameters.

Over-all shape of the nanodiscs. We find that the nanodiscs have a significant elliptical shape with an axis ratio of 1.29 and an average number of DLPC's per nanodisc of 150. One of the central dogma in small-angle scattering is that the effect of polydispersity and the effect of deviating from sphero-symmetry can not be distinguished. While there is already both theoretical[25] and experimental[19] evidence that speaks against this dogma, one should definitely be careful when trying to distinguish between the two effects and quantify them separately. In the nanodisc system, however, the case is somewhat more simple as the constant circumference of the MSP-belt gives strong restrictions to how polydisperse the nanodiscs may be. To first order and assuming a fixed area

per headgroup of the phospholipids, only a dispersity in terms of the number of DLPC molecules per nanodisc, N_{DLPC}, is allowed. This has the consequence that perfectly circular discs are automatically also monodisperse, whereas a dispersion in N_{DLPC} will give rise to an, on average elliptical, shape of the nanodisc-cross-section. It may however be questioned to what extent the maximally loaded circular discs are favoured entropically. Simple statistical mechanical arguments[8] along with experimental evidence for much faster lipid exchange in nanodiscs as compared to liposomes[24] strongly supports that the thermodynamically favoured state of the nanodiscs will be a state where the belt is not maximally expanded and the disc has an elliptical shape. As this elliptical shape will always provide the leading contribution to the smearing of the differential scattering cross-section, an elliptical rather than a polydisperse (spherical) shape was implemented already in the early stages of the model refinement-process. In the process of data analysis and model refinement, a dispersion in the N_{DLPC} was also implemented in terms of a (truncated) Gaussian distribution number of DLPC's per nanodisc around the average N_{DLPC}. This introduced the width of the Gaussian, $\sigma_{N\text{-DLPC}}$, as an extra fit-parameter, however, when $\sigma_{N\text{-DLPC}}$ was fitted, it very quickly decreased towards a value corresponding to zero dispersion. While this is not a proof that the nanodiscs are elliptical and fully monodisperse, it definitely suggests, that the dispersion in N_{DLPC} is rather small.

The packing of the molecular constituents of the nanodiscs. The obtained A_{head} of 58.6 Å^2 for DLPC in nanodiscs is comparable to the value observed in phospholipid bilayer liposomes.[26] However, as it is being discussed and documented in more detail in ref. 8, the area is slightly smaller than the liposome value, and consistent with a lateral compression of the DLPC bilayer when this is localized in a nanodisc. The data suggest that the number of water molecules per PC headgroup, N_w, is very low in the system. Indeed, when the N_w was fitted, it ended up very close to zero and in practice we decided to fix the N_w to zero in the simultaneous fit in order to avoid nonphysical minima with a slightly negative N_w-value. As mentioned previously, we have very good estimates for the partial specific molecular volumes of the DLPC and the MSP, but do not know these values in the nanodiscs with 100% certainty. For this reason these values were taken as adjustment parameters. While the effect of slightly adjusting the partial specific molecular volumes is general non-significant in the fitting of the bulk-like SANS-contrasts, the fine-tuning of these play a remarkably central role for obtaining a good fit to the SAXS data. In the nanodiscs we observe a ν_{DLPC} of 998 Å^3 and a ν_{MSP} of 30220 Å^3. For the DLPC, the obtained value should be compared to the value of 985 Å^3 that we have observed by densitometry measurements in multilamellar liposome systems (see ref. 8), whereas for the MSP the observed molecular volume corresponds to a mass-density of 1.362 g/cm^3 in very close agreement with the value of 1.347 g/cm^3 reported as a representative average value for a wide range of different sized protein systems.[22]

Interface roughness of the nanodiscs. As mentioned above and illustrated in Fig. 6(C), the effect of explicitly including the interfacial roughness of the nanodiscs is very significant. The data from the three contrast could not be fitted simultaneously without explicitly including the roughness term, while excellent fits was obtained when including this term. In both the SAXS and the SANS contrasts we find a surface roughness slightly above 4 Å. This roughness is slightly higher than what we have previously observed in a droplet microemulsion system,[19] but not surprising when taking into account the relatively rough peptide interface as defined by the MSP-belts and when considering that we are describing the alphahelical MSP-belt by an edged hollow cylinder.

Conclusion

A detailed structural model for nanodiscs is derived and refined against experimental SANS and SAXS data obtained on a DLPC-nanodisc sample measured at 20 °C. The systematic incorporation of additional experimental information of molecular or biophysical character allow us to determine the solution structure of the nanodiscs with a so far, unprecedented structural resolution. Several smaller elements have to be combined in order to obtain a good and self-consistent structural model. One of the remarkable effects is the surface roughness term which is absolutely crucial to take into account, in particular when analysing high-resolution small-angle X-scattering data ($q_{max} \approx 0.5$ Å) that are becoming more and more widely accessible at Synchrotron SAXS beamlines optimized for solution scattering. In the present work, we succeed in giving a satisfactory description of the scattering data using a model based on "geometrical" building blocks. However, this approach has a limited resolution and we expect that the steadily improving high-q resolution at synchrotron SAXS instruments, will create a demand for more realistic descriptions of the molecular structures. This could be descriptions that explicitly include the molecular structure of the building blocks such that the short length scale density fluctuations and interface roughness present in alpha-helices, phospholipid bilayers *etc.* are automatically taken into account. This approach will also potentially enable the inclusion of non-symmetric objects such as membrane proteins, which can only in a very crude approximation be described *via* a geometrical object.

Experimental

Preparation of nanodiscs

The preparation of studied nanodisc samples are fully described in ref. 8 and follows the standard procedures previously documented.[1]

SAXS and SANS measurements

SANS and SAXS data sets for the simultaneous analysis (Fig. 6) were acquired at, respectively, the SANS instrument D11 at Institut Laue Langevin (ILL) and at the Bio-SAXS instrument at beamline ID14-3 at European Synchrotron Radiation Facility (ESRF). Both facilities are located in Grenoble, France. For the SANS measurements a combination of two instrumental settings were used to obtain

a sufficiently wide q-range: A setting for covering the high q-values with a sample-detector distance of 1.5 m, and a collimation length of 4 m and a setting for covering the medium q-values with a sample-detector distance of 4 m, and a collimation length of 10.5 m. In both cases a neutron wavelength of 6.0 Å was used with a wavelength spread, $\Delta\lambda/\lambda$ of 10% FWHM. Both the SANS and SAXS measurements were performed at 20 °C. Absolute scale calibration was performed using H_2O as external reference and following the standard procedures at the facility. All samples were measured in flat rectangular Hellma quartz cells during the measurements. In order to optimise signal-to-noise and minimize incoherent background and multiple scattering effects, the samples in 100% D_2O were measured in cells with a path length of 2 mm, whereas the samples in 42% D_2O/58% H_2O was measured in cells with a path length of 1 mm. For the SAXS measurements we used the fixed instrument setup including the HPLC-based automatic sample loading robot standard to the beamline ID14-3. Here the path length of the sample cell was 1 mm. Absolute scale calibration was performed using Bovine Serum Albumin as the external reference.

The SANS measurements were performed two days prior to the SAXS measurements. Ahead of the SANS measurements, the usual H_2O-based buffer used in the nanodisc preparation, had been substituted for corresponding buffers with, respectively 100% D_2O and 42% D_2O/58% H_2O. In the latter contrast, the membrane scaffolding protein is contrast matched and in practice invisible in the experiments. The isotope substitution of the buffers was performed using centrifugal spin-filters with a cut-off of 100 kDa.

Small resolution effects are present in the SANS data, mainly due to the non-negligible spread of the wavelengths of the incoming neutrons. In the synchrotron SAXS data, resolution effects are in practice negligible due to a very high monochromaticity of the incoming beam, a very small beam-diameter and a close to perfect collimation. The resolution effects may easily be included in the data analysis following the principles described in ref. 27 by smearing the model function by the appropriate resolution function before comparing model to experimental data. The inclusion of the resolution function typically affects the fit results, however, in the present case, and with the good resolution obtained at D11 at ILL, the effects were in practice fully negligible; neither the reduced χ^2 or the fit-results were affected. As an exception to usual procedures, the resolution effects were therefore not included in order to gain computational speed.

Acknowledgements

The authors thank Dr A. Round at the ID14-3 Bio-SAXS beamline at ESRF, the European Synchrotron Radiation Facility and Dr R. Schweins at the SANS instrument at beamline D11 at ILL, Institute Laue Langevin. In particular, we gratefully acknowledge the access to "simultaneous" SAXS and SANS beamtime at ILL and ESRF. The authors also wish to thank Prof. Steven Sligar and the Sligar-lab for providing us with MSP1D1 for the nanodiscs. Finally we thank the Danish Government funded UNIK Synthetic Biology program for co-funding of the project and the Danish Research Council funded Danscatt-organization for providing economical support to travels to large facility SAXS and SANS beam time.

References

1 T. H. Bayburt, Y. V. Grinkova and S. G. Sligar, *Nano Lett.*, 2002, **2**, 853–856.
2 I. G. Denisov, Y. V. Grinkova, A. Lazarides and S. G. Sligar, *J. Am. Chem. Soc.*, 2004, **126**, 3477–3487.
3 I. G. Denisov, M. McLean, A. Shaw, Y. V. Grinkova and S. G. Sligar, *J. Phys. Chem. B*, 2005, **109**, 15580–15588.
4 T. H. Bayburt, Y. V. Grinkova and S. G. Sligar, *Arch. Biochem. Biophys.*, 2006, **450**, 215–222.
5 M. R. Whorton, M. P. Bokoch, S. G. F. Rasmussena, B. Huang, R. N. Zare, B. Kobilka and R. K. Sunahara, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7682–7687.
6 T. H. Bayburt and S. G. Sligar, *FEBS Lett.*, 2010, **584**, 1721–1727.
7 A. Y. Shih, I. G. Denisov, J. Phillips, S. G. Sligar and K. Schulten, *Biophys. J.*, 2005, **88**, 548–556.
8 N. Skar-Gislinge, J. B. Simonsen, K. Mortensen, R. Feidenhans'l, S. G. Sligar, B. L. Møller, T. Bjørnholm and L. Arleth, *J. Am. Chem. Soc.*, 2010, **132**, 13713–13722.
9 O. Mouritsen and M. Bloom, *Biophys. J.*, 1984, **46**, 141–153.
10 O. Glatter, *J. Appl. Crystallogr.*, 1977, **10**, 415–421.
11 F. Ye, G. Hu, D. Taylor, B. Ratnikov, A. A. Bobkov, M. A. Mclean, S. G. Sligar, K. A. Taylor and M. H. Ginsberg, *J. Cell Biol.*, 2010, **188**, 157–173.
12 J. S. Pedersen and C. Svaneborg, *Curr. Opin. Colloid Interface Sci.*, 2002, **7**, 158–166.
13 J. S. Pedersen, *Adv. Colloid Interface Sci.*, 1997, **70**, 171–210.
14 J. S. Pedersen, in *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, ed. P. Linder and T. Zemb, Elsevier Science B. V., Amsterdam, 2002, ch. 16.
15 J. S. Pedersen and M. Gerstenberg, *Macromolecules*, 1996, **29**, 1363–1365.
16 J. S. Pedersen, *J. Appl. Crystallogr.*, 2000, **33**, 637–640.
17 L. Arleth, B. Ashok, H. Onyuksel, P. Thiyagarajan, J. Jacob and R. Hjelm, *Langmuir*, 2005, **21**, 3279–3290.
18 P. Debye, *J. Phys. Chem.*, 1947, **51**, 18–32.
19 L. Arleth and J. S. Pedersen, *Phys. Rev. E*, 2001, **63**, 061406.
20 C. Tanford, *J. Phys. Chem.*, 1972, **76**, 3020–3024.
21 W. Sun, R. Suter, M. Knewtson, C. Worthington, S. Tristram-Nagle, R. Zhang and J. F. Nagle, *Phys. Rev. E*, 1994, **49**, 4665–4676.
22 E. Mylonas and D. I. Svergun, *J. Appl. Crystallogr.*, 2007, **40**, 245–249.
23 J. S. Pedersen, S. L. Hansen and R. Bauer, *Eur. Biophys. J.*, 1994, **22**, 379–389.
24 M. Nakano, M. Fukuda, T. Kudo, M. Miyazaki, Y. Wada, N. Matsuzaki, H. Endo and T. Handa, *J. Am. Chem. Soc.*, 2009, **131**, 8308–8312.
25 E. Caponetti, M. Floriano, E. Didio and R. Triolo, *J. Appl. Crystallogr.*, 1993, **26**, 612–615.
26 Dr N. Kucerka, Personal communication cited with permission from N. K. Canadian Neutron Beam Centre, National Research Council of Canada Chalk River Laboratories, Chalk River, Canada.
27 J. S. Pedersen, D. Posselt and K. Mortensen, *J. Appl. Crystallogr.*, 1990, 321–333.

Chapter 7

Paper 3

What determines the shape and stoichiometry of self-assembled phospholipid nanodics: Speed of self-assembly process, initial lipid:MSP stoichiometry or detergent type?

Nicholas Skar-Gislinge, Rasmus Høiberg-Nielsen, and Lise Arleth*

Biophysics, Faculty of Life Sciences, University of Copenhagen, Denmark

E-mail: lia@life.ku.dk

Abstract

Phospholipid nanodiscs have quickly become a very widely used platform for functional studies of membrane protein studies, and several research groups are now investigating the system with the aim of developing it into a platform for structural studies of membrane proteins. Poor control of the molecular self-assembly process that ultimately should place a membrane protein inside a nanodisc, pose a challenge for a successful high yield reconstitution of membrane proteins in to nanodisc. In a recent study, a combined small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) allowed for extracting detailed structural information about the nanodiscs in terms of e.g. their over-all shape - this was shown to be elliptical - and the packing of the lipids within the nanodiscs in terms of the local areas per head group. In an attempt to further understand the nanodisc self-assembly process and this way contribute to obtaining better control of the membrane protein incorporation, the self-assembly process of POPC:MSP1D1 nanodiscs have been systematically modulated and the

*To whom correspondence should be addressed

detailed structure of the resulting nanodiscs have been determined using the previously developed small-angle scattering based approach. In direct contrast to what is found in bilayer liposome systems, we find that the removal speed of the reconstitution detergent and thereby the overall speed of the self-assembly process does not play a significant role for the shape and stoichiometry of the formed nanodiscs. The POPC:MSP1D1 stoichiometry of the reconstitution mixture has often been suggested to be a determining factor for the stoichiometry of the formed nanodiscs. In the present study we show that a POPC:MSP1D1 reconstitution ratio of less than 75:1 results lipid poor nanodiscs, whereas larger POPC to MSP1D1 reconstitution ratios produce nanodiscs with a POPC:MSP1D1 ratio of 65:1. The detergent type used in the reconstitution process is found to be a central determining factor for the shape and stoichiometry of the formed nanodiscs. We find a significantly increasing POPC:MSP1D1 stoichiometry of the nanodiscs as the reconstitution detergent type is changed in the order: Tween80, DDM, Tween40, Triton, OG, Chaps, Tween20 and Na-Cholate. Thus the detergents with smaller molecular packing parameters give rise to nanodiscs with higher stochiometries. This clearly shows that the detergents optimal for solution storage and crystallization of membrane proteins are not the same as the detergents optimal for the nanodisc reconstitution. The observation furthermore suggests that reconstitution strategies based on mixed micellar systems are not only convenient but also optimal when reconstituting membrane proteins into nanodiscs. Finally we observe that the prepared nanodiscs are not completely homogenous with respect to the POPC:MSP1D1.

Introduction

Phospholipid nanodiscs[1] , have been suggested as a platform for the investigation of membrane proteins by Sligar *et al.*[2] and have quickly found a broad range of applications within a variety studies of membrane proteins.[3,4,5,6,7,8] When studying this quickly expanding literature it is striking that the majority of the studies focus on functional investigations of membrane proteins inside nanodiscs, while only a minor fraction of the literature report from structural studies.[2] This observation is even more astounding when taking into account the general lack of structural information

about membrane proteins combined with the widely accepted biological and pharmaceutical importance of the systems. About 30% of the proteins coded for by the human genome are believed to be membrane proteins. However, at present, only 381[1] of the about 35000[2] unique protein structures published in the Protein Data Bank correspond to membrane protein structures. The commonly accepted explanation for the low number of membrane protein crystal structures is that membrane proteins are notoriously difficult to crystallize.

The nanodisc platform for reconstituting membrane proteins has the advantage over the traditionally used detergent systems of providing a more native-like phospholipid environment. Phospholipid liposomes are also widely used as a very relevant platform for investigations of membrane proteins. However, while liposomes certainly has the advantage of providing a functionally very relevant environment for the membrane proteins, with the possibility of having an inside and outside environment. They are, however, less well-suited for structural studies due to the low overall ratio of membrane protein to phospholipid.

Despite that there have been attempts [3] none of the present membrane protein crystal structures have made use of the nanodisc system to obtain membrane protein crystals. However, solution based techniques giving intrinsically lower structural resolution, such as CryoTEM and small-angle scattering have been able to elucidate the over-all structure of membrane proteins based on the nanodisc platform.[9,10] In addition, high resolution NMR techniques, are able to provide crucial structural information about small membrane proteins reconstituted into nanodiscs.[11] The number and quality of the structural results based on these techniques are however, also still limited. We hypothesize, based on our own experience, that an explanation for this is that it is much more challenging to obtain a structurally relevant reconstitution of membrane proteins into nanodiscs, than

[1] According to the data base of membrane protein structures maintained by Stephen White as of Jan. 16'th, 2013. Counting only entries solved by diffraction methods. http://blanco.biomol.uci.edu.

[2] According to a search in the Protein Data Bank Jan. 16'th, 2013. Counting only protein entries solved by diffraction method and removing similar sequences at 100% identity. http://www.rcsb.org/pdb/search/advSearch.

[3] Personal commmunication with Prof. Sligar

it is to obtain a functionally relevant reconstitution. While functional studies may be performed at very low concentrations and even on samples where only a minor population of membrane proteins are ultimately successfully reconstituted, the structural studies rely on a much stronger control of the self-assembly process and the availability structurally homogeneous reconstitutions of membrane proteins into nanodiscs.

In order to establish the knowledge necessary to develop a better control of the nanodisc self-assembly process, we have, in three cases, systematically varied the nanodisc self-assembly conditions and correlated this with the detailed structure of the formed nanodiscs. In a fourth case we study the individual fractions collected from the main SEC elution peak of a nanodisc preparation.

Inspired by results from liposome self-assembly by detergent removal, we first speculated that the removal speed of the reconstitution detergent would affect the self-assembly of the formed nanodiscs, analogous to what is observed in liposome systems where faster detergent removal leads to smaller and more uniform liposomes.[12] To investigate whether a similar effect was present for the Nanodisc-self-assembly, a series of nanodiscs were prepared from the same reconstitution mixture but the reconstitution detergent was removed by the addition of increasing amounts of Biobeads, where the amounts of Biobeads, and thus the detergent removal speed, was increased over an order of magnitude. To our surprise this variation of the detergent removal speed had no significant effect on the nanodisc self assembly, in contrast to the results from the the liposome self-assembly, where a significant increase of both average size and inhomogeneity is observed with decreasing detergent removal speed.

In protocols for nanodisc preparation it is commonly emphasized that a sufficiently high lipid-to-MSP ratio in the initial stock solution is crucial for the formation of nanodiscs fully loaded with phospholipids. While there is clearly a correlation between the lipid-to-MSP stoichiometry during the reconstitusion and the final lipid-to-MSP stoichiometry in the purified nanodiscs,[1] our SAXS

data show that for POPC:MSP1D1 this correlation is only present at reconstitusion ratios below 75:1, whereas above there is no significant dependence.

Transmembrane proteins can not be stabilized in aqueous solutions without the help of a storage detergent that encapsulates the hydrophobic transmembrane part. In this context, it is well documented that the conservation of both function and structure of membrane proteins in solution depend critically on the actual choice of storage detergent. A few non-ionic detergent types such as DDM and octyl glucoside appear to be compatible with conservation of membrane protein function and are often used in the crystallization of membrane proteins, while Na-cholate, which is used in the original nanodisc reconstitution protocol,[1] is generally not expected to be a good membrane protein storage detergent. This incompatability creates a general necessity for developing reconstitution protocols based on other detergents than Na-cholate.

Keeping in mind that the nanodisc self-assembly process is highly dynamic and takes place from a starting point of mixed micelles of MSP, membrane lipids and reconstitution detergent, during the removal of the reconstitution detergent, it may intuitively be expected that the choice of reconstitution detergent will affect the structure of the formed nanodiscs. Both when the desired end-product is membrane protein loaded nanodiscs and when it is "empty" nanodiscs. That this is actually the case is clearly demonstrated by the SAXS data presented in this article. The lipid-to-MSP ratio may be regarded as a simple measure for the structure of the formed nanodiscs and we find that this ratio and consequently the shape of the nanodiscs in terms of their geometrical ellipticity, depend critically on the choice of reconstitution detergent.

We have previously found that the nanodiscs prepared with either POPC or DLPC have an elliptical cross section, however this apparent elliptical shape can in principle also be explained by an in-homogenous sample of circular nanodiscs with a distribution of radii.[13] We have investigated this further by combining SAXS and Size Exclusion Chromatography (SEC). Due to the high flux

at synchotron bioSAXS instruments, it is becomming possible to preform SAXS measurements on each collected fraction from the SEC without the need to pool several fractions and then concentrate. From these measurements we find that the nanodiscs are present as a distribution of sizes (and hence lipid-to-MSP ratios) with a varying axis ratio. However none of the fractions are found to contain Nanodiscs with a circular cross section.

Materials and methods

SAXS experiments

SAXS experiments were performed at ID13-4 BioSAXS beamline[14] at the ESRF, Grenoble France, except for the measurements presented in case B which were measured at the I911saxs Beamline at MAXLAB, Lund Sweden. The data were reduced according to standard procedures using the beamline software, and scaled to absolute units using a secondary protein standard.[15] The Indirect Fourier Transforms[16] (IFT) were performed using a Baeysian of the IFT implementation capable of estimating both the smoothness parameter and the maximum distance in the sample.[17]

The data were analyzed further using the model described in the Theory and SAXS data Analysis section.

Procedures for the reconstitution of nanodiscs

All nanodiscs were prepared using the standard procedure[1] with the following modifications:

Case A: Nanodiscs were prepared using the standard procedure. After purification the histidine tag (HIS tag) present on the MSPs were cleaved by incubation with Tobacco Ecth Virus (TEV) and both TEV and cleaved HIS tag was removed via binding to a nickel column.

Case B: The initial POPC:MSP1D1 reconstitusion ratio was varied from 35:1 to 110:1 otherwise the standard procedure was followed.

Case C: Preparation volumes were half of the standard preparation and the amount BioBeads added was varied from 0.11 g/ml to 1.06 g/ml.

Case D: Each preparation was modified in order to take into account the different critical micelle concentrations of the detergents and so that the solubilised lipids and detergents were in a mixed micelle phase.

Phosphorus analysis

The total phosphorus analysis was done by chemically digestion of all phospholipids in H_2SO_4 [18] and then and then quantification quantifying the total phosphorus. From previous experiments this method is accurate to, at least, within $\pm$ 5 phospholipids.

Theory and SAXS Data Analysis

The analysis of the measured SAXS data from nanodisc samples are performed as described in our previous article.[19] Generally, the small-angle x-ray scattering intensity of a dilute suspension of particles can be expressed by:

$$I(q) = nb^2P(q) \tag{1}$$

Where n is the particle number density, b is the particle excess scattering length which may also be expressed by $b = V\Delta\rho$. V is the single particle volume and $\Delta\rho$ is the average excess scattering length density of the particle as compared to the solvent. q is related to the scattering angle, θ, by $q = 4\pi sin(\theta)/\lambda$ where λ is the wavelength of the x-rays. $P(q)$ is the single particle form factor intensity and contain the information on the particle shape, it is normalized such that $P(0) = 1$. The amplitude form factor is related to the form factor by $P(q) = |A(q)|^2$.

For SAXS, the amplitude form factor is the Fourier transform of the excess electron density distribution of the particle. Because of this, the form factor for various geometric objects can readily be calculated in a semi analytical manner[20] and be combined to yield even more complex objects. This principle is used to construct a model of the small angle scattering from nanodiscs where it is represented as an assembly of geometrical shapes: The central phospholipid bilayer is represented as a stack of five cylinders with an elliptic cross section. Each cylinder represent a different molecular part of the phospholipid bilayer. The top cylinder represents the head groups and associated hydration water, the next cylinder the alkyl tails of the top phospholipids, the middle cylinder the methyl end groups of both the top and bottom phospholipids, the fourth layer represents the alkyl tails of the lower phospholipids and the bottom cylinder represents the head groups and associated water molecules of the bottom half of the bilayer. The two MSP1D1s spanning the rim of the bilayer are represented by a hollow cylinder.

In order to minimize the number of free parameters in the model, constraints based on the molecular composition of the nanodisc has systematically been incorporated. For example the the volume of the cylindrical stack representing the phospholipid bilayer is constrained to the phospholipid aggregation number times the partial molecular volume of a phospholipid molecule. By fitting the model to the measured data a number of structural parameters of the nanodisc can be obtained.[21] A detailed description of the model can be found in *Skar-Gislinge & Arleth 2010*[19]

Results and discussion

Case A: Nanodiscs from Different SEC Fractions

Figure 1 show the SEC elution profile of POPC nanodisc, the collected peak fraction for the scattering experiments as well as the forward scattering of the collected fractions. The resulting SAXS curves are shown in figure 3 A) and the corresponding pair distance distributions functions (PDDF)

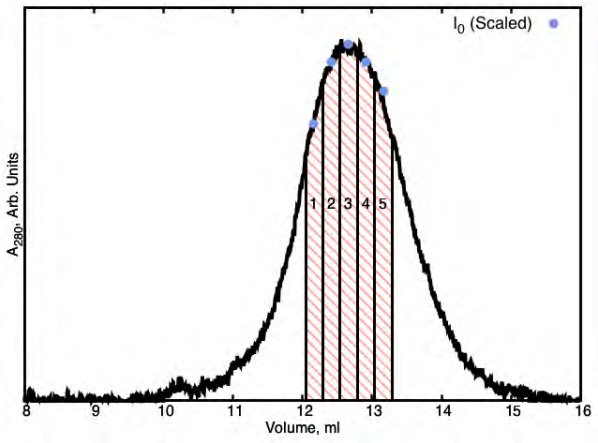


Figure 1: Size exclusion chromatography elution profile of POPC nanodiscs. The collected peak fraction are shown in red and are numbered 1 to 5. Forward scattering, proportional to the sample concentration, is shown as blue dots scaled to the UV absorption.

are shown in figure 3 B). The maximum distance present in the sample show no systematic dependence on the faction number, however the the PDDF does show a gradual change implying that the structure change. However, from the PDDF alone, the nature of this change is not clear.

The structural difference between the nanodiscs in the collected fractions, can be interpreted by fitting the model for small angle scattering of nanodiscs to the scattering data. The resulting model fits to the measured data, are shown in black in figure 3 A). Form the model analysis the first fraction is found to have a smaller axis ratio and a higher lipid aggregation number per nanodisc, then the nanodiscs in the later fractions. These parameters are shown in figure 3. Otherwise the structural parameters remain the same.

Case B: Nanodiscs Produced at Different Lipid-to-MSP Stochiometries

Figure 4 shows size exclusion data from nanodiscs prepared with different lipid-to-MSP reconstitusion ratios, as well as the collected peak fractions. From a lipid-to-MSP ratio of 110:1 to 75:1 there is only a little change in the retention volume. However, below a ratio of 75:1 the retention time increase with decreasing lipid-to-MSP, indicating that the hydrodynamic radius of the self

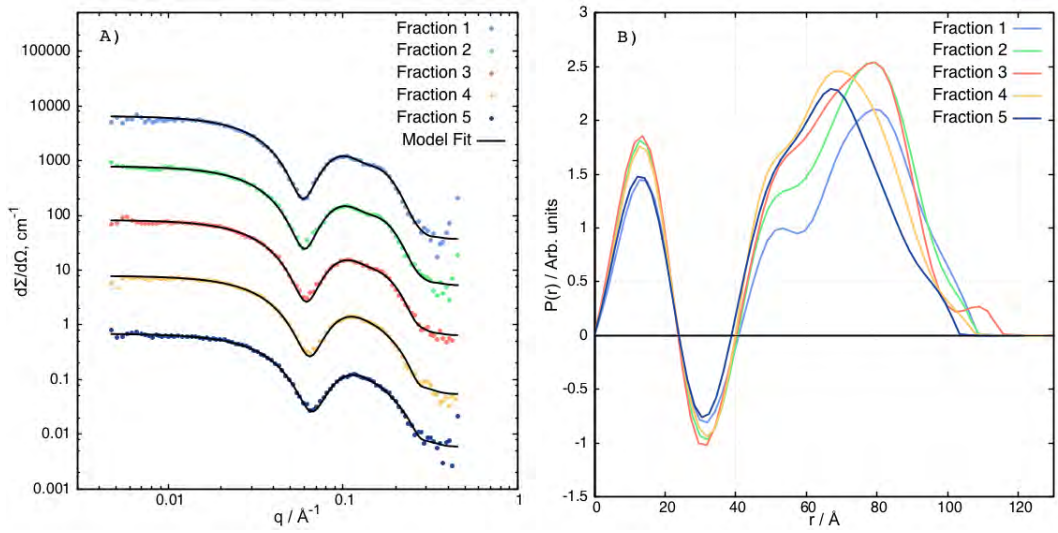


Figure 2: A): SAXS curves from the 5 collected frations in figure 1 along with the corresponding nanodisc model fits shown in black. B) Indirect Fourier Transforms of the data shown in A)

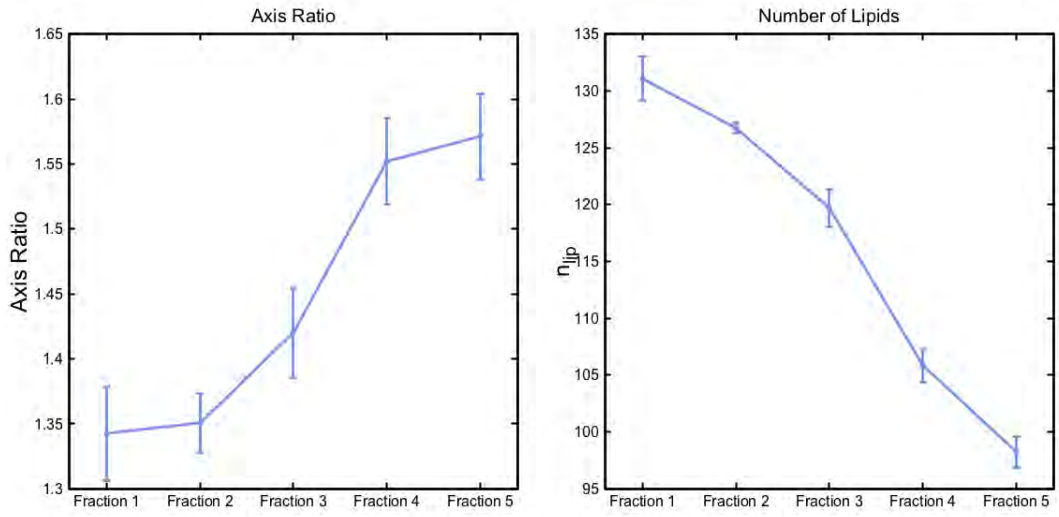


Figure 3: Axis ratio (A) and lipid aggregation number (B) obtained from the model fits shown in figure 1

assembled particles decrease with a decreasing initial lipid-to-MSP ratio.

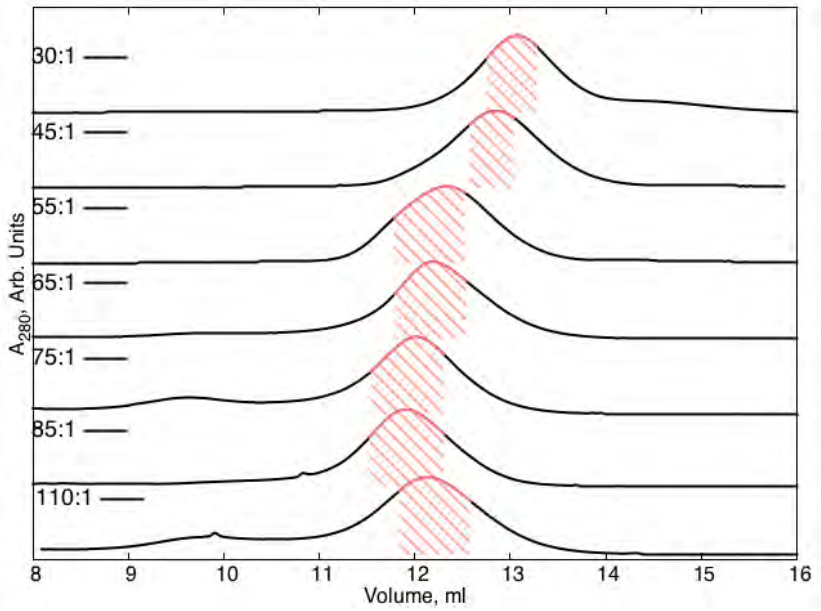


Figure 4: Size exclusion data of nanodiscs prepared with an increasing lipid to MSP ratio. The red areas indicate the collected peak fractions used for x-ray scattering.

Figure 5 A) show the scattering data and figure 5 B) show the corresponding indirect Fourier transforms, of the scattering from the collected peak fractions in figure 4. The scattering data show a displacement of the characteristic minimum from 0.08 q^{-1} to 0.058 q^{-1} and the splitting of the central peak as the lipid stoichiometry increases until a ratio of 65:1, above which the data are indistinguishable. Form the IFTs in figure 5 B) the maximum distance present in the sample, D_{max}, found to increases as a function of the increasing lipid-to-MSP ratio until 75:1, in good agreement with the size exclusion data (Table 1).

The scattering data was fitted using the nanodisc model and the resulting model fits are shown in black in figure 5 A.

The lipid aggregation number per disc was calculated by measuring the total phosphorus concentration and the protein concentration in each sample. The initial mixing ratio, the size of the

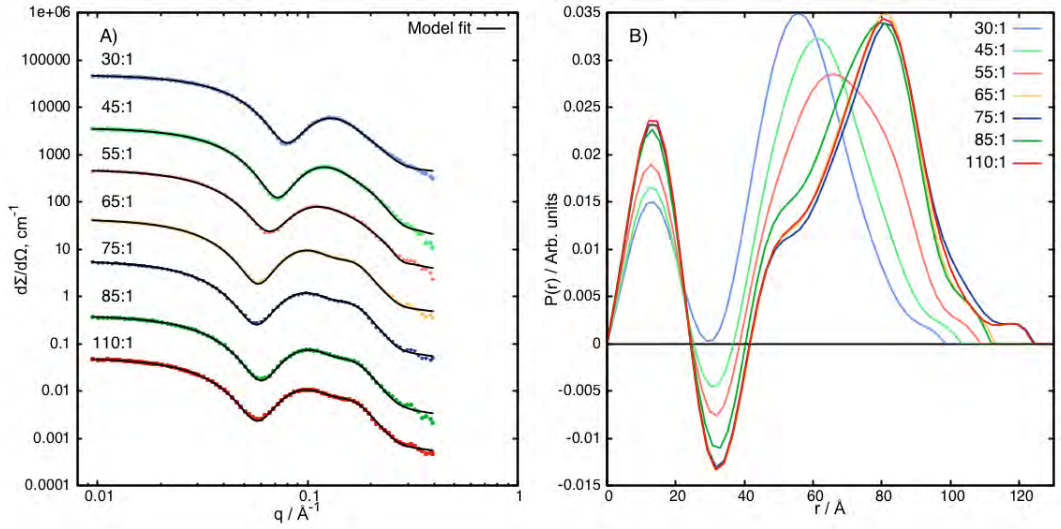


Figure 5: A) SAXS data from nanodiscs prepared with an increasing lipid to msp ratio. B) Indirect Fourier transforms of SAXS data

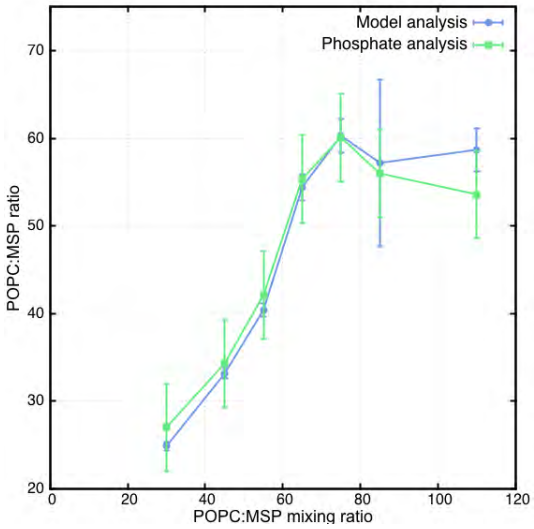


Figure 6: Phospholipid to MSP ratio plotted agains the initial mixing ratio obtained from the model analysis of the scattering data (blue) and by measuring the total phosphorus.

nanodiscs measured as the D_{max} and the lipid-to-MSP ratio measured via the phosphorus analysis are compared ing table 1.The lipid-to-MSP ratio measured via phosphorus analysis is compared to the lipid aggregation number per Nanodisc obtained from the model fits to the scattering data in figure 6. The lipid-to-MSP ratios found by the two measurements are in very good agreement, and both increase with increasing mixing ratio until an initial mixing ratio of 75:1, above which it stabilizes.

Table 1: Size and stoichiometry obtained from the IFT analysis and the phosphate analysis compared to the initial mixing ratio.

Mixing Ratio	D_{max}	POPC:MSP
30:1	96.7 ± 0.9	27 ± 5
45:1	101.4 ± 1.5	34 ± 5
55:1	106.7 ± 0.6	42 ± 5
65:1	110.8 ± 0.7	55 ± 5
75:1	122.1 ± 0.8	60 ± 5
85:1	109.8 ± 0.7	56 ± 5
110:1	121.7 ± 0.6	54 ± 5

Case C: Nanodiscs produced at different detergent removal speeds

The SEC profiles of nanodiscs prepared using NaCholate showed no dependence of the amount of added BioBeads (data not shown). The SAXS data, shown in figure 7, are likewise almost indistinguishable and subsequent model analysis did not indicate any significant dependence on the amount of BioBeads added during the detergent removal step.

Case D: Nanodiscs Produced with different reconstitution detergents

Figure 8 A) show SAXS data from nanodiscs prepared with different detergents. The data show a clear effect on the structure of the purified nanodiscs by the choice of detergent. By using the nanodisc model this effect is quantified in form of a dependence of the lipid aggregation number on the detergent used. The hydrodynamic radius of the mixed detergent lipid micelles, before addition of MSP, was measured using dynamic light scattering. Figure 8 B) show the lipid aggregation

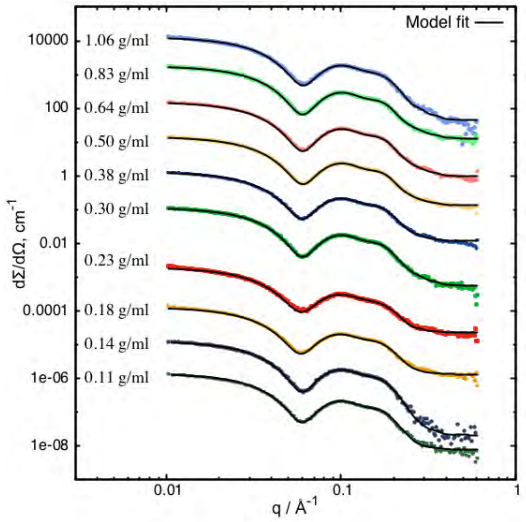


Figure 7: SAXS data of nanodisc prepared using NaCholate and an increasing amount of BioBeads, in order to simulate an increasing detergent removal speed.

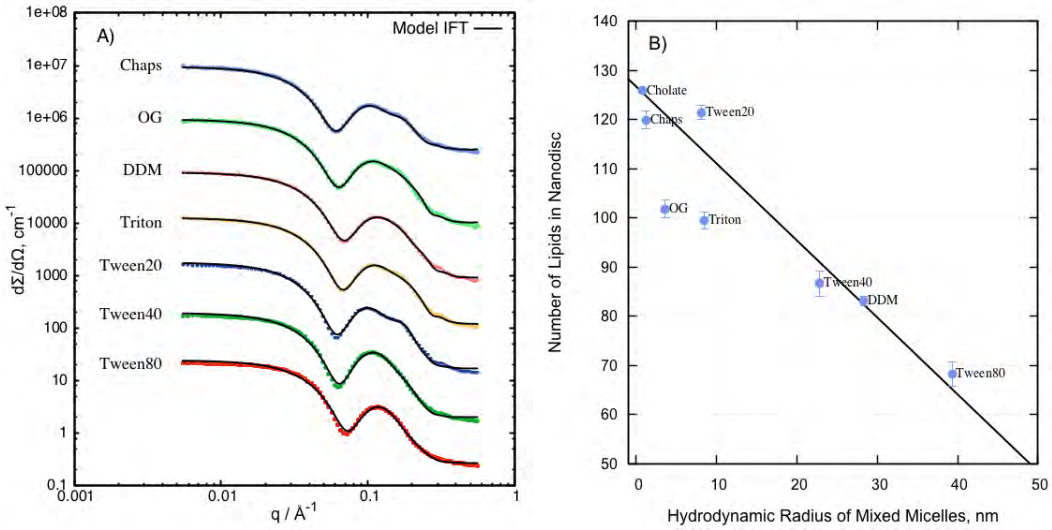


Figure 8: A) SAXS data from Nanodiscs prepared using different detergents and the corresponding model fits. B) Lipid aggregation number found by model fits to the scattering data as a function of the hydrodynamic radius of the POPC/detergent mixed micelles in the reconstitusion mixture.

number per nanodisc, as a function of the hydrodynamic radius of mixed micelles of POPC and the different detergents used. From these data it is evident that there is a clear trend that detergents that form smaller mixed micelles incorporate more lipids into the nanodisc.

General discussion

From the four cases presented above we make several observations:

We observe that within the peak fractions normally collected when purifying nanodiscs using SEC, there exist a distribution of sizes, and that this distribution is not resolved by the standard SEC procedure used for preparing nanodiscs. The model analysis of the SAXS data give a more detailed view, showing that the axis ratio of the nanodisc changes along with the lipid aggregation number, with a larger lipid aggregation number and smaller axis ratios the leading fractions compared to the tailing fractions.

The concentration weighted average values of the aggregation number and axis ratio are calculated to: N_{agg}=116 and ε = 1.45, close to the values for POPC discs (N_{agg}= 124 and ε= 1.44, data not shown) found by SAXS analysis of the whole elution peak of a similar sample.

We observe that it is possible to "under load" the nanodiscs during the self assembly, producing a more elliptic and "lipid poor" nanodisc. However we also observe, that on the other hand, it is not possible to produce a "lipid rich" nanodisc by increasing the lipid-to-MSP ratio. This indicate that the MSP defines a finite size and lipid aggregation number of the nanodisc. Furthermore we find that even "fully loaded" nanodiscs prepared in a large excess of lipids, have an elliptical cross section, ruling out that the finding of an elliptic cross section is simply a consequence sub optimal lipid-to-MSP ratio during the reconstitusion.

The detergent removal experiment, interestingly, shows that the detergent removal speed does not affect the structure of the nanodiscs prepared using the detergent NaCholate, within the timescale accessible in this experiment. This is in contrast to similar experiments on the formation of phospholipid vesicles.

Finally we observe that the choice of detergent has a huge impact on the lipid aggregation number

of the formed disc and that this is correlated to the size of the mixed micelles form of phospholipids and that particular detergent.

There are several parameters governing the self assembly of phospholipid nanodiscs, and we here only probe a subset of these parameters. However from the observations presented above we can try to build up a picture of self assembly process. There is a clear evidence that the size of the mixed lipid/detergent micelle is determining the lipid aggregation number found in the nanodisc. By estimating the number of POPC molecules per lipid we can separate the detergents into two groups in which the self-assembly process follows two distinctly different pathways. The micelles formed with NaCholate, Chaps and OG all have less than 100 lipids per micelle, and will therefore have to fuse with one another in order to form a nanodisc with $\approx$ 130 POPC molecules. The other Detergents form larger micelles and have a much higher number of phospholipid than can be contained in a nanodisc, and will therefor have to be divided into smaller aggregates stabilized by the MSPs during the reconstitusion process. In the above calculations the micelles are naively assumed to be spherical, which is of course not the case for the large micelles. This will also have to be taken into account the get a more complete picture of role of the detergent/lipid micelle size on the self assembly of nanodiscs.
The diffusional dynamics are much faster for the small mixed micelles compared to the large mixed micelles. This dynamic may be fast enough for the small micelles to reach equilibrium while the detergents are removed by the biobeads. This would also explain why the detergent removal speed (simulated by an adding an increased amount of biobeads) did not show and effect on the self assembly, when using NaCholate.

If the picture presented above is correct, a detergent removal experiment using tween 80, in stead of NaCholate, should then show a dependence on the detergent removal speed, if the dynamics are not able to establish equilibrium while the detergent are being removed. Furthermore, a longer incubation time of the MSP with the mixed detergent/lipid micelles then the standard 1 hr

could also yield disc with a higher lipid aggregation number.

Finally we believe that a higher success rate for producing structurally homogenous nanodiscs, can be achieved by carefully tuning lipid-to-detergent ratio as well as the right mixture of detergent the reconstitution, in order to obtain small mixed micelles, with a lipid content comparable to the desired nanodisc. As demonstrated dynamic light scattering can be a useful tool to easily when these conditions has been met.

Conclusion

We have presented a comprehensive study of the factors governing the self assembly of phospholipid nanodiscs and observe that the self assembly of nanodisc, like preparation of phospholipid liposomes, is highly dependent on the preparation procedure. However as opposed to liposomes, the nanodisc self assembly process is not affected by the detergent removal speed but is highly affected by the choice of detergent. An important finding is that the detergents often used for solubilizing membrane proteins while still maintaining function, are not the same detergents that result in an optimal nanodisc preparation. In particular detergents forming small lipid/detergent micelles incorporate more lipids, this information may useful when optimizing incorporation of membrane proteins in tho phospholipid nanodisc.

Acknowledgement

The authors thank Dr. A. Round at the ID14-3 Bio-SAXS beamline at ESRF, the European Synchrotron Radiation Facility. The authors also thank the Lundbeck Foudation and the Danish Government funded UNIK Synthetic Biology program for co-funding of the project.

References

(1) Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. *Nano Letters* **2002**, *2*, 853–856.

(2) Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. *Archives Of Biochemistry And Biophysics* **2006**, *450*, 215–222.

(3) Ritchie, T.; Zolnerciks, J. K.; Sligar, S.; Grinkova, Y. V.; Grinkova, Y. V.; Bayburt, T. H.; Bayburt, T. H.; Denisov, I. G.; Atkins, W. M.; Atkins, W. M. *Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs*; Elsevier, 2009; pp 211–231.

(4) Whorton, M. R.; Yao, X. J.; Sunahara, R. K.; Velez Ruiz, G.; Rasmussen, S. G. F.; DeVree, B. T.; Deupi, X.; Kobilka, B. *Proceedings of the National Academy of Sciences* **2009**, *106*, 9501–9506.

(5) McLean, M. A.; Ye, F.; Hu, G.; Taylor, D.; Ratnikov, B.; Bobkov, A. A.; Sligar, S.; Taylor, K. A.; Ginsberg, M. H. *J Cell Biology* **2010**, *188*, 157–173.

(6) Tajkhorshid, E.; Ohkubo, Y. Z.; Pureza, V. S.; Morrissey, J. H.; Davis-Harrison, R. L.; Sligar, S. *Thrombosis Research* **2008**, *122*, S23–S26.

(7) Bayburt, T. H.; Leitz, A. J.; Xie, G.; Oprian, D. D.; Sligar, S. *Journal Of Biological Chemistry* **2007**, *282*, 14875–14881.

(8) Das, A.; Grinkova, Y. V.; Sligar, S. *Journal Of The American Chemical Society* **2007**, *129*, 13778–+.

(9) 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. *Nature Structural & Molecular Biology* **2011**, *18*, 614–621.

(10) Kynde, S. A. R.; Skar-Gislinge, N.; Pedersen, M. C.; Midtgaard, S. R.; Simonsen, J. B.; Schweins, R.; Mortensen, K.; Arleth, L. *Acta Crystallographica Section D Biological Crystallography* **2014**, *70*, 371–383.

(11) Wagner, G.; Hagn, F.; Raschle, T.; Etzkorn, M. *Journal Of The American Chemical Society* **2013**, 130108043754005.

(12) Lévy, D.; Bluzat, A.; Seigneuret, M.; Rigaud, J. L. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1990**, *1025*, 179–190.

(13) Caponetti, E.; Floriano, M.; Didio, E.; Triolo, R. *Journal Of Applied Crystallography* **1993**, *26*, 612–615.

(14) 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. *Journal of Physics: Conference Series* **2010**, *247*, 012009.

(15) Mylonas, E.; Svergun, D. I. *Journal Of Applied Crystallography* **2007**, *40*, 245–249.

(16) Glatter, O. *Journal Of Applied Crystallography* **1977**, *10*, 415–421.

(17) Hansen, S. L. *Journal Of Applied Crystallography* **2012**, *45*, 566–567.

(18) Rouser, G.; Fleischer, S.; Siakotos, A. N. *Lipids* **1966**, *1*, 85–86.

(19) Skar-Gislinge, N.; Arleth, L. *Phys. Chem. Chem. Phys.* **2010**, *13*, 3161–3170.

(20) Skov Pedersen, J. *Advances In Colloid And Interface Science* **1997**, *70*, 171–210.

(21) Bjørnholm, T.; Møller, B. L.; Feidenhans'l, R.; Simonsen, J.; Skar-Gislinge, N.; Sligar, S.; Mortensen, K.; Arleth, L. *Journal Of The American Chemical Society* **2010**, *123*, 13713–13722.

Chapter 8

Paper 4

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment

Søren A. R. Kynde,[a] Nicholas Skar-Gislinge,[a] Martin Cramer Pedersen,[a] Søren Roi Midtgaard,[a,c] Jens Baek Simonsen,[a] Ralf Schweins,[b] Kell Mortensen[a] and Lise Arleth[a]*

[a]Structural Biophysics, Niels Bohr Institute, Faculty of Science, University of Copenhagen, Denmark, [b]Institute Laue–Langevin, Grenoble, France, and [c]Nano-Bio Science, Department of Chemistry, University of Copenhagen, Denmark

Correspondence e-mail: arleth@nbi.ku.dk

Received 5 June 2013
Accepted 15 October 2013

Monomeric bacteriorhodopsin (bR) reconstituted into POPC/POPG-containing nanodiscs was investigated by combined small-angle neutron and X-ray scattering. A novel hybrid approach to small-angle scattering data analysis was developed. In combination, these provided direct structural insight into membrane-protein localization in the nanodisc and into the protein–lipid interactions. It was found that bR is laterally decentred in the plane of the disc and is slightly tilted in the phospholipid bilayer. The thickness of the bilayer is reduced in response to the incorporation of bR. The observed tilt of bR is in good accordance with previously performed theoretical predictions and computer simulations based on the bR crystal structure. The result is a significant and essential step on the way to developing a general small-angle scattering-based method for determining the low-resolution structures of membrane proteins in physiologically relevant environments.

1. Introduction

Membrane proteins reside in the lipid membranes of the cell. They play a central role in transport and signalling across the cell membrane and are consequently key targets in drug development. The function of membrane proteins is closely coupled to their intrinsic flexibility. Furthermore, membrane proteins require a native membrane-like local environment in order to maintain structural and functional stability. These central features imply that despite the remarkable breakthroughs in the last few years (Morth *et al.*, 2007; Rosenbaum *et al.*, 2009), it is still regarded as extraordinarily challenging to crystallize membrane proteins to obtain high-resolution structural information. Consequently, only 406 structures[1] out of roughly 36 000 unique protein structures[2] published in the Protein Data Bank correspond to membrane-protein structures. This should be seen in light of the fact that more than 26% of the proteins coded for by the human genome are expected to be membrane proteins (Wallin & von Heijne, 1998; Fagerberg *et al.*, 2010).

The function of membrane proteins takes place in close interplay with the surrounding membrane environment. As a result of the disordered nature of lipids and the often rather extreme conditions of membrane-protein crystallization, little information about the structural interplay between membrane proteins and the surrounding fluid lipid environment is

[1] According to a search in the database of membrane-protein structures maintained by Stephen White as of 1 June 2013, counting only entries solved by diffraction methods (http://blanco.biomol.uci.edu).

[2] According to a search of the Protein Data Bank on 1 June 2013, counting only entries solved by diffraction methods and removing similar sequences at 100% identity (http://www.rcsb.org/pdb/search/advSearch.do).

obtained from crystallographic data (Sonntag *et al.*, 2011; Gonen *et al.*, 2005). This means that while numerous theoretical predictions and computer simulations (Lomize *et al.*, 2011; Sonntag *et al.*, 2011) have been made of the interactions between membrane proteins and lipid membranes, very little directly experimentally obtained structural information is available to guide these simulations and theory development.

Nanodiscs are small lipoprotein complexes containing a small sheet of phospholipid bilayer stabilized by two amphipathic membrane scaffolding proteins (MSPs) spanning the rim of the phospholipid sheet. Numerous experimental studies have exploited the fact that membrane proteins can be embedded into the phospholipid interior of the nanodisc (Frauenfeld *et al.*, 2011; Bayburt & Sligar, 2003; Yao *et al.*, 2009; Whorton *et al.*, 2007; El Moustaine *et al.*, 2012) and have used the platform to enable functional studies of membrane proteins in solution-like conditions. While solution studies are also regularly performed on membrane proteins reconstituted into detergent micelles, the nanodisc system has proved to have superior properties with respect to maintaining the function of the membrane proteins (Lyukmanova *et al.*, 2012; Berthaud *et al.*, 2012; Calcutta *et al.*, 2012).

In a recent study, a combination of small-angle X-ray and neutron scattering (SAXS and SANS) allowed us to demonstrate that the unloaded nanodiscs are remarkable structurally homogeneous (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2010), particularly compared with detergent micelles, where the size typically follows a broad size distribution. This structural homogeneity of the nanodisc is most likely to be an effect of the balance between the rather fixed perimeter of the discs as imposed by the amphipathic MSP and the well defined preferred area per head group and hence molecular packing of the incorporated phospholipids. In the analysis of SAXS/SANS data, good structural homogeneity provides the possibility of obtaining high resolution in the structural interpretation. In the study mentioned above this fact allowed unprecedented insight into the nanodisc structure, including the lipid packing inside the discs.

The structural homogeneity of the nanodisc system, its size and its native membrane-like environment make it a promising platform for structural studies of membrane proteins. Interesting results have already been obtained using cryo-EM on a membrane-anchored protein system (Frauenfeld *et al.*, 2011), and the nanodisc platform was recently used in a very interesting NMR study of fully embedded transmembrane structures (Hagn *et al.*, 2013). Small-angle scattering is ideally suited to the solution-based nanodisc system and can even provide information about the phospholipid bilayer and on the interplay between lipids and membrane proteins, information that is neither available from cryo-EM owing to the low electron density of the lipids nor from NMR, where usually the lipids are deuterated to remove the signal. In this light, it is somewhat surprising that to date only a little new insight about membrane proteins inserted into nanodiscs has been obtained with SAXS/SANS.

To date, only a few examples of small-angle scattering studies of a membrane protein incorporated in a nanodisc are available in the literature, all of which are based on SAXS data. The studied proteins include a cytochrome P450-type membrane protein incorporated into a POPC-based nanodisc (Baas *et al.*, 2004), bacteriorhodopsin (bR) incorporated into DMPC-based nanodiscs (Bayburt *et al.*, 2006) and, most recently, the incorporation of a curdlan synthase into two different preparations of nanodiscs based on either POPC or *Escherichia coli* lipids (Periasamy *et al.*, 2013). Owing to the lack of adequate data-analysis methods, however, none of these studies exploit the full information content of the SAXS curves. The data treatment predominately consists of indirect Fourier transform (IFT) analysis, meaning that only the maximum dimensions of the particles can be determined and qualitative comparison with other systems can be made. However, without a model no information about the structure or the localization of the membrane protein into the nanodisc can be obtained. In one case (Baas *et al.*, 2004), an attempt to model the structure of the particle was made, but no fits to the data were shown. This means that the model is in agreement with the overall predicted size, but it is not demonstrated that the model complies with the entire measured scattering function. Including this valuable experimental constraint in the modelling process allows a lot more detail to be extracted, as will be demonstrated here.

In the present article, we report the successful reconstitution of bacteriorhodopsin into nanodiscs containing a 2:1 molar ratio of POPC and POPG. Using a combination of small-angle X-ray and neutron scattering (SAXS and SANS), we show that the vertical position of bR is slightly off-centred, with the C-terminus protruding more than the buried N-terminal region. This has been theoretically predicted (Lomize *et al.*, 2011), but has never been experimentally observed in a real membrane environment. We find that bR is indeed completely surrounded by lipids but is not confined to the centre of the plane of the disc. Furthermore, we observe a small but significant structural modulation of the surrounding phospholipid bilayer, including an increase in the phospholipid area per head group.

The central result of the article is, however, the development of a novel approach for analyzing scattering data from the combined system of nanodisc and membrane protein. We describe a model that is fully compatible with all prior knowledge of the system and with the scattering curves in two different contrast situations. The approach is directly applicable to any system of a membrane protein of well known structure that is embedded in a nanodisc.

2. Materials and methods

2.1. Sample preparation

2.1.1. Bacteriorhodopsin purification. Bacteriorhodopsin (bR) was produced and purified according to previously published reports (Oesterhelt & Stoeckenius, 1974; Dencher, 1982). Briefly, salt medium containing 250 g NaCl (Sigma), 20 g $MgSO_4.7H_2O$ (Sigma), 10 g peptone (Oxoid), 3 g trisodium citrate (Sigma) and 2 g KCl (Sigma) per litre of H_2O

was inoculated with *Halobacterium salinarium*. After 5–6 d of growth at 40°C illuminated with a 500 W lamp, the cells were pelleted and washed in salt buffer containing 250 g NaCl (Sigma), 20 g $MgSO_4.7H_2O$ (Sigma) and 2 g KCl (Sigma) per litre of H_2O. The cells were resuspended in 25 ml Milli-Q H_2O per gram of cells. DNase (Sigma) was added and the solution was left stirring overnight at 4°C. Cell debris was spun down, the supernatant was transferred to centrifuge tubes and the membranes were pelleted at 54 000*g* for 1 h. The supernatant was discarded and the pellet was resuspended in Milli-Q H_2O and spun down a further two times to remove impurities. Isolated membranes were resuspended in a minimum amount of Milli-Q H_2O and carefully layered on top of a sucrose gradient consisting of a step gradient of 30, 40 and 50% sucrose dissolved in STED buffer consisting of 10 m*M* Tris–HCl pH 7.5, 1.5 m*M* EDTA, 0.5 m*M* DTT. The gradient was centrifuged overnight at 200 000*g* to reach equilibrium. The next day, the band containing the purple membrane was extracted and dialyzed against 25 m*M* phosphate buffer pH 7.4 to remove the sucrose. The purple membrane was diluted so that the detergent to bR weight ratio would be ∼20 when octyl glucoside (OG; Applichem) was added to a final concentration of 40 m*M*. The solution was lightly sonicated, left under gentle agitation at room temperature overnight and protected from light. After solubilization, the bR was loaded onto a Superdex 200 column (GE Healthcare) equilibrated in 25 m*M* phosphate buffer pH 7.4, 40 m*M* octyl glucoside. Fractions containing monomeric bR were collected, concentrated, flash-frozen and stored at −80°C until further use.

2.1.2. Biotinylation of bR. bR was biotinylated in order to facilitate the separation of bR-containing nanodiscs from unloaded nanodiscs by affinity purification. bR in 25 m*M* phosphate buffer pH 7.4 containing 40 m*M* octyl glucoside was mixed with a 25-fold excess of succinimidyl-6-(biotinamido)-6-hexanamido hexanoate (NHS-LC-LC-Biotin; Pierce) at 20°C for 1.5 h. The excess unreacted NHS-LC-LC-Biotin was diluted by adding OG–phosphate buffer to the bR solution and concentrating the sample using a 10 kDa cutoff spin filter (Amicon), resulting in a dilution of >5000. A biotinylated bR sample prepared under similar conditions was found to contain 1.6 biotin groups on average using a 2-hydroxyazobenzene-4-carboxylic acid/avidin assay (Green, 1970).

2.1.3. MSP preparation. The amphipathic membrane-scaffolding protein used in the present work was MSP1D1(−). MSP1D1 was expressed and purified according to previously published procedures (Ritchie *et al.*, 2009) and the His tag was cleaved off by adding TEV protease to the sample and utilizing the TEV site incorporated between the His tag and the MSP1D1 sequence. This produced the final MSP1D1(−) protein used in the samples.

2.1.4. Reconstitution of bR into nanodiscs. The preparation of nanodiscs with and without reconstituted bR is described in detail in Bayburt *et al.* (2002) and Bayburt & Sligar (2003), respectively. The molar ratio between MSP1D1(−) and bR was 10:1 in order to minimize the formation of reconstituted nanodiscs containing two or three molecules of bR. The stoichiometry between phospholipids and MSP1D1(−) was 67:1. We used a lipid composition of 2:1 POPC:POPG to form a lipid liquid-phase environment at room temperature and to introduce negatively charged lipids into the lipid bilayer in the nanodisc. These two bilayer features should facilitate a good mimic of the cell membrane compared with DMPC-based bR-nanodiscs. A lipid composition of 2:1 POPC:POPG was used for both the bR-reconstituted nanodisc and the unloaded (non-reconstituted) nanodisc preparations. MSP1D1(−) in 20 m*M* phosphate buffer, bR in OG–phosphate buffer and the lipid–OG–phosphate solutions were mixed. The mixture had a final lipid concentration of 10 m*M* and was incubated at 4°C for 1 h with stirring. SM-2 Bio-Beads (Bio-Rad) were added and the Bio-Bead mixture was stirred overnight at 4°C to remove the detergent. The bR-reconstituted nanodiscs were subsequently separated from the unloaded nanodiscs by affinity chromatography. An immobilized monomeric avidin resin (Pierce) was used according to the manufacturer's instructions. The purified bR-nanodisc sample was stored in a 15% glycerol Tris buffer solution flash-frozen in liquid nitrogen. Finally, the bR-nanodisc solution was thawed and purified using a Superdex 200 size-exclusion column (GE Healthcare) pre-equilibrated in buffer consisting of 20 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl at 4°C. Fractions containing the sample were pooled and concentrated using a 50 kDa cutoff spin filter prior to the SAXS measurements. The buffer was exchanged to a D_2O-based buffer prior to the SANS measurements by diluting the sample ten times in D_2O buffer and then concentrating the sample with a 50 kDa cutoff spin filter. The procedure was repeated four times to ensure almost complete exchange into D_2O buffer.

The theoretical ratio between the absorbance at 280 and 550 nm of the MSP1D1(−)-based nanodisc with only one bR reconstituted into the disc is 2.8 according to UV–Vis studies conducted by Bayburt & Sligar (2003). The final bR-nanodisc used for the small-angle scattering measurements exhibited a ratio of 2.7 according to UV–Vis measurements. Thus, it was confirmed that the solution consists of bR-loaded nanodiscs with only a single bR molecule incorporated in each of the nanodiscs.

2.1.5. Small-angle scattering. For the bR-loaded discs, small-angle X-ray scattering experiments were performed on the BioSAXS beamline ID14-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data were collected at 20°C using the fixed instrument as described in Pernot *et al.* (2010). In Fig. 1(*a*) the data are shown as an absolute scaled intensity as a function of $q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. Radial averaging and q-conversion of data were performed using the standard software at the beamline. Absolute scaling, *i.e.* expressing intensities as scattering cross-section per sample volume in units of cm^{-1}, was performed using BSA as a standard (Mylonas & Svergun, 2007).

The small-angle neutron scattering of the bR-loaded nanodiscs was performed on beamline D11 at the Institut Laue–Langevin, Grenoble, France (Lieutenant *et al.*, 2007). The close proximity to the SAXS facility enabled the measurement of the same sample preparations within a few

hours. In order to cover the desired q range, measurements were performed at two sample-to-detector distances: 1.504 m with 4 m collimation for the q range 0.0060–0.43 Å^{-1} and 9.495 m with 10.5 m collimation for the q range 0.0069–0.072 Å^{-1}. The wavelength spread, $\Delta\lambda/\lambda$, was 10% for all settings. Azimuthal integration of the data as well as absolute scaling with water as a standard were performed using the *LAMP* software package at the beamline. The data are shown in Fig. 1(*a*).

On another occasion, we performed SAXS and SANS measurements of the unloaded nanodiscs. These data are shown in Fig. 1(*c*). Again, the beamtimes were scheduled so that both measurements could be performed on the same sample preparation within 24 h.

SAXS measurements of the unloaded discs were performed on the BioSAXS beamline BM29 at ESRF. This is the same instrument as above but moved from beamline ID14-3 to BM29. The settings and data treatment were the same as described for the loaded nanodiscs above.

SANS measurements of the unloaded discs were performed at KWS 1, Forschungs Neuronenquelle Heinz Maier-Leibnitz, Munich (FRM II). Measurements were performed using

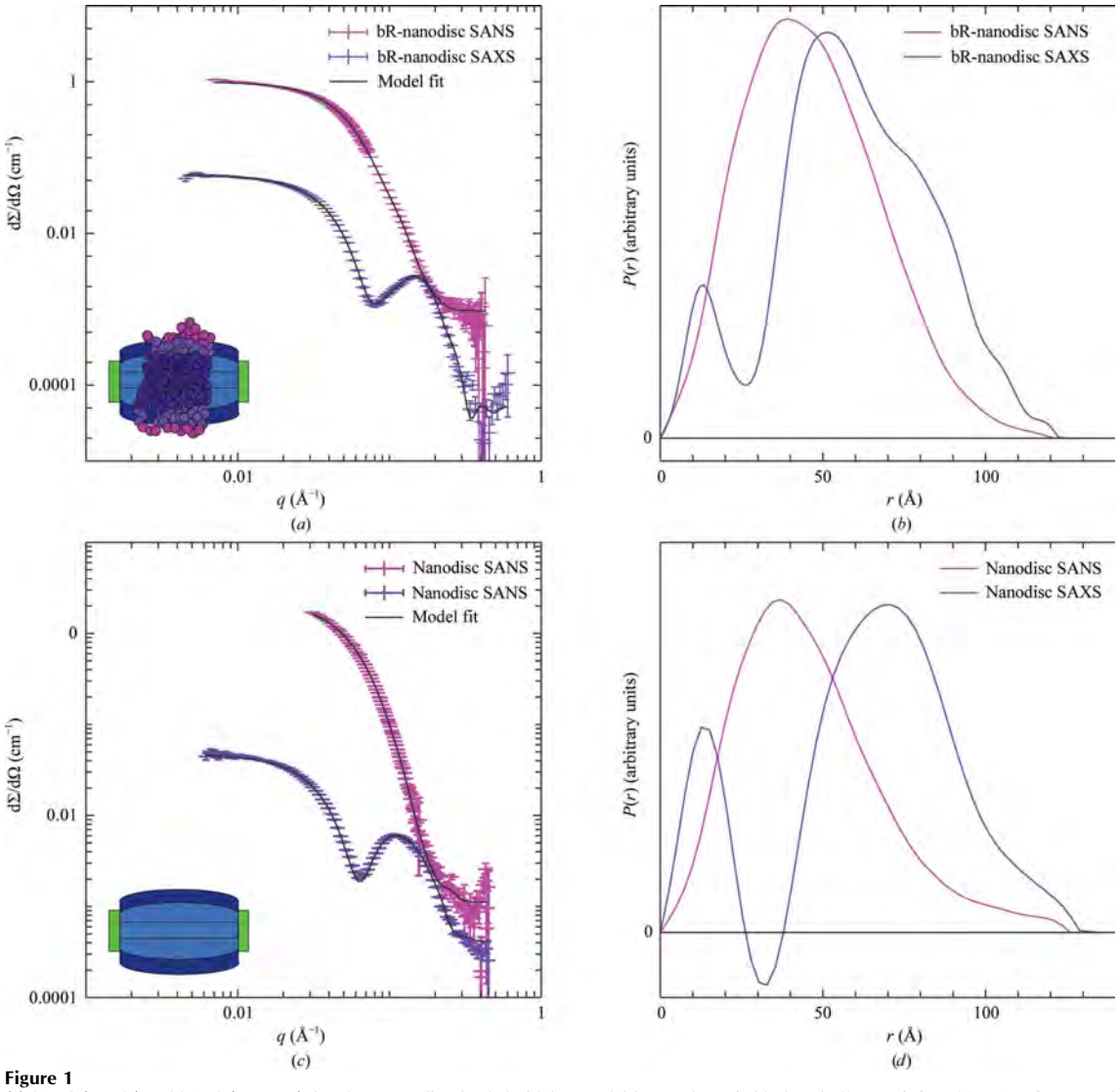


Figure 1
(*a*) SAXS (purple) and SANS (magenta) data from nanodiscs loaded with bR; model fits are shown in black. A buffer consisting of 100% D_2O was used for the SANS measurements. (*b*) Pair-distance distribution functions corresponding to the scattering data in (*a*). (*c*) SAXS and SANS data from unloaded nanodiscs; model fits are shown as black lines. (*d*) Pair-distance distribution functions corresponding to the data shown in (*c*).

neutrons with a wavelength of 4.5 Å at three sample-to-detector distances: 1.27 m with 4 m collimation, 3.77 m with 4 m collimation and 7.77 m with 8 m collimation. These settings covered q-ranges of 0.035–0.45, 0.012–0.16 and 0.0057–0.077 Å^{-1}, respectively. Again, the wavelength spread was 10%.

Immediately before each scattering measurement, the 280 nm absorption of the samples was measured in order to determine the protein concentration using the extinction coefficients for bR and MSP given in Bayburt & Sligar (2003).

2.1.6. Sample handling: SAXS. The samples were thawed and spun at 12 000g for 15 min in order to remove larger aggregates caused by freezing. Afterwards, the glycerol added as a cryoprotectant was removed by diluting the sample by a factor of ten and concentrating using a 50 kDa cutoff centrifuge filter. This was repeated an additional two times, giving a residual glycerol concentration of 0.00015%(v/v); the final flowthrough was kept as an accurate buffer background for the SAXS measurements.

Sample handling was performed using the specialized sample-handling robot available at the BioSAXS beamline. A test run was performed in order to determine the exposure time of the sample before a measurable degree of radiation damage could be detected. Furthermore, the sample was flowed through the beam to make sure that the same volume of sample was not exposed twice. The sample was measured ten times for 500 ms during flow.

In order to obtain a precise buffer background, the flow-through from the spin filter was measured both before and after the sample and the average was used for subtraction.

2.1.7. Sample handling: SANS. To minimize the incoherent scattering from hydrogen, the Tris buffer was exchanged to Tris buffer prepared using D_2O instead of H_2O. This greatly reduced the amount of hydrogen present in the sample and increased the scattering contrast of the protein and lipids in the sample. The buffer exchange was performed using a 50 kDa cutoff centrifuge filter; again, the sample was diluted by a factor of ten with the D_2O buffer and concentrated back to the original concentration. This was repeated an additional two times, giving a residual H_2O concentration of 0.001%(v/v).

2.1.8. Software and implementation. The presented model was implemented in a larger data-analysis framework written in C and controlled *via* an interface written in Python. Using this framework, the model was fitted to the presented data with an adaptation of the Levenberg–Marquardt algorithm (Marquardt, 1963) using the so-called gridsearch implementation (Pedersen, 1997).

The minimization was cross-checked using the Broyden–Fletcher–Goldfarb–Shanno algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) to ensure correctness and convergence. The source code for traditional implementations of these algorithms can be found in the literature (Press *et al.*, 1992).

Local minima are difficult to avoid in complex models. However, the inclusion of molecular constraints and the fitting on an absolute scale together with the condition that the model must fit SAXS and SANS data simultaneously significantly minimizes the number of truly different local minima as experienced in the model-fitting process.

Owing to the heavy nature of the computational tasks, parallelization was implemented using the OpenMP protocol for C, thereby expanding the calculation of $I(q)$ on the available CPUs. With these implementations, the time needed to compute the χ^2 of a given model is a couple of seconds on a modern high-end desktop PC. Thus, execution of the entire minimization procedure takes from a couple of minutes to a couple of hours depending on the convergence criteria and the precision setup of the fitting routine as well as the initial parameter values for the model.

The mathematical tools used in the computation were based on the built-in routines in the mathematical library of C as well as the Gnu Scientific Library (GSL).

The presented confidence intervals were estimated using the profile-likelihood framework. The upper and lower bounds for each parameter were determined such that relaxation of the remaining parameters yields a fit with an increase in χ^2 of no higher than 1.0, corresponding to the 68.3th percentile.

3. Theory and mathematical modelling

3.1. Development of a hybrid approach for structural modelling and analysis of data

For the data analysis, we have developed a novel general approach that combines the advantages of the two prevalent approaches used to analyze small-angle scattering data from particles in solution.

In the first of these approaches, which may be termed *continuous modelling*, particles are represented by simple geometrical objects and the spherical average of the analytical form-factor intensity is fitted to the data. Internal structure can be modelled by combining the form-factor amplitudes of various objects. These could, for instance, be concentric spheres of alternating scattering-length densitiy to account for the core-shell structure of detergent micelles (Cabane *et al.*, 1985). The particle form factors can be combined with analytical structure factors and size distributions, which allows the inclusion of concentration effects and polydispersity in the modelling. Analytic form factors are available for several shapes (Pedersen, 2002), and the approach has successfully been applied to describe a large number of systems, *e.g.* phospholipid vesicles (Kučerka *et al.*, 2010; Andersen *et al.*, 2011), detergent micelles (Lipfert *et al.*, 2007) and micro emulsions (Chen, 1986). Molecular constraints can easily be incorporated into the model, for example information about the molecular components and sample concentration.

The other approach, which may be termed *bead-based modelling*, has proved very successful in systems of monodisperse proteins in solution. Proteins may have complicated shapes but a relatively homogeneous scattering-length density on the relevant length scale. The method utilizes the fact that the scattering from an assembly of spherical beads can readily be calculated (Svergun & Koch, 2003). One bead can, for

instance, represent a single atom or an amino-acid residue. Examples of bead-based methods are *ab initio* and rigid-body modelling (Petoukhov & Svergun, 2005). In *ab initio* modelling, no prior knowledge of the shape of the protein is necessary. By moving one bead at a time, assuming homogeneous packing of the residues, a multitude of shapes can be explored. The overall shapes that fit the data are often strikingly similar and thus are likely to represent the actual shape of the protein (Svergun, 1999). If part of the protein has been crystallized, its residue positions can be determined from the crystal structure. This part is then treated as a rigid body that can be translated and rotated relative to the rest of the structure.

The hybrid approach proposed in the present work combines the two methods to allow a quick computation of the scattering from the nanodisc and membrane protein taking into account all of the *a priori* knowledge of the system. Fig. 2 illustrates how a dummy-residue model based on the crystal structure of a membrane protein is combined with a continuous model of the nanodisc within the hybrid approach. Various model parameters, such as the height of the lipid bilayer and the relative orientation of the membrane protein, can be varied to fit the experimental data.

In a previous study (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2010), it was shown that a nanodisc without a membrane protein could be described by the continuous approach. The disc was modelled as a set of concentric cylinders with elliptical cross-sections. The phospholipid bilayer consisted of three regions of different contrast representing lipid head-groups, lipid alkyl chains and lipid methyl groups, respectively. The MSP was modelled by an elliptical ring.

For the present work, we have used a slight modification of this model in which the head-group and alkyl-chain regions are slightly lens-shaped instead of flat (Kaya, 2004). This accounts for the possible effect of the lipids adjusting to the different hydrophobic heights of the MSP and the membrane protein. As described previously (Skar-Gislinge & Arleth, 2010), molecular constraints are systematically incorporated into the model to minimize the number of free parameters. The bacteriorhodopsin molecule is represented by point-like dummy residues. Each residue corresponds to one amino acid. Their relative positions are based on the coordinates of the crystal structure deposited in the PDB as entry 1py6 (Faham *et al.*, 2004). The position of each residue is determined by the centre of scattering of all of the atoms of that residue. For this calculation, it is taken into account that the position of the scattering centre depends on the scattering-length density of the displaced phase. The lateral position and rotation of the whole membrane protein relative to the nanodisc are the fitting parameters.

3.2. Computation of scattering from the hybrid model

The scattering amplitude from the membrane protein is combined with the description of the unloaded nanodisc in the following way: if the centre of mass of a residue is placed inside the lipid membrane in the nanodisc, the scattering from the displaced lipid (head group, alkyl chain or methyl group as appropriate) has to be subtracted, exactly as one usually subtracts scattering from displaced solvent. If the residue is outside the lipid membrane the scattering from the displaced solvent is instead subtracted.

In practice, the residue j is weighted by its excess scattering length $\Delta b_j = b_j - \rho_j^{\rm bg} v_j$, where b_j is its total scattering length and v_j is its partial specific volume. The background $\rho_j^{\rm bg}$ is the scattering-length density of the displaced solvent if the residue is outside the lipid bilayer, or the excess scattering-length density $\Delta\rho_i$ of the appropriate part of the lipid if it is inside the

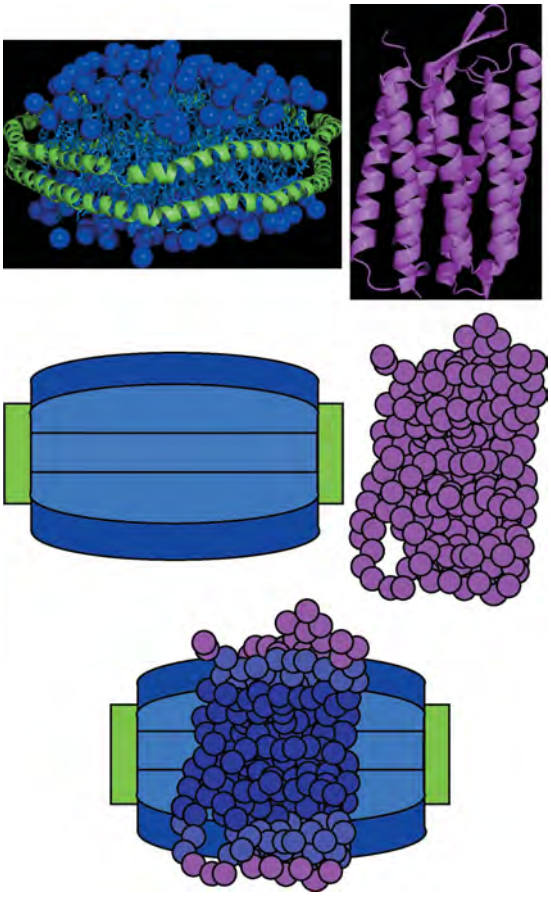


Figure 2
Top row: detailed atomic models from molecular dynamics (Shih *et al.*, 2005) of the unloaded nanodisc (left) and X-ray crystallography of bacteriorhodopsin (right). Middle row: simplified representations of the constituents corresponding to the level of resolution of small-angle scattering. The disc is represented by a few geometrical objects. The membrane protein is represented by a larger number of point-like dummy residues. Bottom: the simplified representations allow quick computation of the simulated scattering from the combined system. The different colours of the dummy residues signify that their weight factors have been adjusted to account for the displaced lipids.

Table 1
Excess scattering-length densities for X-rays ($\Delta\rho_x$) and neutrons in 100% D_2O ($\Delta\rho_n$).

For the lipid head groups the appropriate value for the 2:1 PC:PG mixture is given. The bR molecule is built from beads representing amino-acid residues with varying scattering contrast. The reported value is the weighted mean.

	$\Delta\rho_x$ (cm^{-2})	$\Delta\rho_n$ (cm^{-2})
Lipids		
Head group	5.91×10^{10}	-3.66×10^{10}
Alkyl chains	-1.20×10^{10}	-6.60×10^{10}
Methyl	-4.73×10^{10}	-7.22×10^{10}
Protein		
bR (mean)	3.94×10^{10}	-3.24×10^{10}
MSP	3.79×10^{10}	-2.88×10^{10}

bilayer. If the model membrane protein is translated relative to the membrane, some or all of the residues may move from one phase to another, for instance from the lipid head-group region to the solvent, and their excess scattering length must be adjusted accordingly.

In order to gain computational speed, the amplitudes of both the bead-based model of the membrane protein and the continuous description of the nanodisc are expanded in terms of spherical harmonics.

The spherical harmonic expansion coefficients, A_{lm}, of an assembly of M beads is given by (Svergun *et al.*, 1995)

$$A_{lm}(q) = i^l \left(\frac{2}{\pi}\right)^{1/2} \sum_{j=1}^{M} \Delta b_j J_l(qr_j) Y^*_{lm}(\theta_j, \varphi_j). \quad (1)$$

The position of the jth bead in spherical coordinates is $\mathbf{r}_j = (r_j, \theta_j, \varphi_j)$. Its excess scattering length is given by Δb_j as explained above. J_l are the Bessel functions of the first kind and Y^*_{lm} are the complex conjugates of the spherical harmonics.

On the other hand, the spherical harmonic expansion coefficients, B_{lm}, of the analytical form factor amplitudes, F, describing the continuous model of the unloaded nanodisc can be calculated by

$$B_{lm}(q) = \Delta b_{\rm nd} \left\langle F(\mathbf{q}) \frac{\exp(-im\beta)}{(2\pi)^{1/2}} \tilde{P}_{lm}[\cos(\alpha)] \right\rangle_{\alpha,\beta}, \quad (2)$$

where $\mathbf{q} = (q, \alpha, \beta)$ is the momentum-transfer vector in spherical coordinates, $\tilde{P}_{lm}$ are the Legendre polynomials and $\langle \ldots \rangle_{\alpha,\beta}$ denotes the spherical average. This is a double integral that has to be solved numerically.

The excess scattering length $\Delta b_{\rm nd}$ of the disc is given by $\Delta b_{\rm nd} = \sum_i V_i \Delta\rho_i$, where V_i and $\Delta\rho_i$ are the volumes and excess length densities of the different phases, *e.g.* the lipid head group or MSP.

Even though the expansion (2) is computationally costly, this is compensated by the fact that the combined spherically averaged scattering intensity of the whole system, $I(q)$, can now easily be calculated by

$$I(q) = n \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(q) + B_{lm}(q)|^2. \quad (3)$$

The expression becomes exact for $L \to \infty$, but typically L around 20 is sufficient. The intensity is proportional to the number density of scatterers n and has units of scattering cross-section per unit volume.

Note that if the position of a single bead is changed, only one term in (1) and all of the terms of (3) need to be recalculated to obtain the scattering from the new configuration. This means that the scheme outlined here in principle allows *ab initio* modelling of proteins inside a lipid membrane in the case where a crystal structure is not available.

3.3. Model implementation

To fit the scattering data from bacteriorhodopsin incorporated into the nanodisc, the following parameters were adjusted. (i) The partial specific molecular volumes of the single lipids, the MSP and the membrane protein. These molecular volumes were not allowed to deviate more than a few percent from the pre-estimated values reported in the literature (Kučerka *et al.*, 2008; Skar-Gislinge *et al.*, 2010; Faham *et al.*, 2004; Fraser *et al.*, 1978). (ii) The height of the hydrophobic part of the lipid bilayer (at the rim) and the curvature of the bilayer. Together with the number of lipids in the disc, these parameters determine the cross-sectional area of the disc. (iii) The axis ratio of the elliptic cylinder representing the lipid bilayer was also taken as a fitting parameter, and thus the lengths of the major and minor axes of the elliptical cylinders constituting the lipid bilayer could be deduced. (iv) In addition, the number of hydration water molecules per lipid head group was fitted along with two roughness terms correcting for the fact that the interfaces are not perfectly smooth. These parameters are explained in more detail in a recent publication (Skar-Gislinge & Arleth, 2010). The fitting parameters governing the positioning of the membrane protein were a vertical shift defined as the distance of the centre of mass of the membrane protein to the centre of the bilayer along the bilayer normal, a horizontal shift along the major semi-axis of the ellipsis and finally a tilt around the major semi-axis of the ellipsis of the main axis of the bR molecule relative to the surface normal of the lipid bilayer.

Hence, a total number of 13 free parameters were fitted. The model had to fit the SAXS curve and the SANS curve simultaneously, and the curves were fitted on an absolute scale. This meant that the total amount of excess scattering cross-section per unit volume from the electrons in the model had to add up to the forward scattering of the SAXS curve and at the same time the combined scattering cross-sections per unit volume of the nuclei had to add up to the forward scattering of the SANS curve. Some typical examples of excess scattering-length densities are given in Table 1. The excess scattering length, Δb, of an object of volume V filled with a particular phase of contrast $\Delta\rho$ is $\Delta b = V\Delta\rho$. The forward scattering of a solution of particles that each have a total excess scattering length of ΔB is $n(\Delta B)^2$, where n is the number of particles per unit volume of the solution.

Table 2
Structural parameters from the model fits shown in Fig. 1.

A total number of 13 parameters were fitted to the data from the bR-carrying nanodisc and nine parameters were fitted to the unloaded nanodisc. The parameter marked (nf) was not fitted. The upper and lower bounds indicate the confidence interval corresponding to the 68.3rd percentile.

	bR-loaded nanodisc			Unloaded nanodisc		
	Lower bound	Fitted value	Upper bound	Lower bound	Fitted value	Upper bound
Fit parameters						
bR tilt (°)	23.1	**31.2**	42.6		—	
bR vertical shift (Å)	−7.0	**−4.2**	0		—	
bR horizontal shift (Å)	10.4	**12.1**	13.3		—	
bR molecular volume (Å^3)	32400	**32900**	33400		—	
Hydrophobic thickness at rim (Å)	23.3	**24.6**	25.5	30.3	**30.8**	31.2
Additional thickness at centre owing to curvature of bilayer (Å)	0	**1.8**	6.44		0 (nf)	
No. of lipids in disc	128	**130.6**	133	125.7	**126.66**	127.9
No. of bound H_2O per lipid	5.6	**8.7**	12.2	0	**0**	1.6
Axis ratio of bilayer	1.29	**1.44**	1.59	1.59	**1.66**	1.74
Molecular volume per lipid (Å^3)	1255	**1261**	1268	1280	**1284**	1288
Molecular volume of one MSP (Å^3)	26100	**26500**	26900	24600	**25100**	25400
Interface roughness, X-rays (Å)	5.4	**5.58**	5.9	5.08	**5.36**	5.67
Interface roughness, neutrons (Å)	5.0	**5.32**	5.7	4.41	**4.56**	4.73
Radius of protruding His tags (Å)		—		5.8	**12.5**	20.4
Deduced parameters						
Interface area per lipid (Å^2)		77.3			62.9	
Bilayer thickness at rim (Å)		39.4			40.8	
Bilayer thickness at centre (Å)		40.2			40.8	
Bilayer circumference (Å)		272			234	

4. Results

4.1. Visual inspection of SAXS and SANS data

The recorded two-dimensional scattering data were reduced and scaled as described in §2 and the appropriate backgrounds were subtracted. The background-subtracted SAXS and SANS data from nanodiscs with embedded bacteriorhodopsin are shown in Fig. 1 along with the corresponding indirect Fourier transforms (IFTs; Glatter, 1977). Also shown are the SAXS and SANS data sets of nanodiscs without incorporated bacteriorhodopsin.

The SAXS data of both the unloaded and loaded nanodiscs exhibit a characteristic minimum at around 0.08 Å^{-1}. This oscillatory behaviour is owing to the combination of positive and negative excess scattering-length densities characteristic of these phospholipid-based systems (see Table 1). While the oscillation is apparent both with and without the presence of bR, the specific shapes are clearly different in the two types of systems.

The oscillation is not seen in the SANS data since all of the constituents of the system have similar neutron contrasts on the background of 100% D_2O. This clearly shows that there is complementary information in the two data sets.

4.2. Pair-distance distribution functions

Indirect Fourier transformations were performed using the Bayesian indirect Fourier transformation method (BIFT; Hansen, 2000). Using Bayesian statistics, this method determines the pair-distance distribution function, $p(r)$, without prior user input, *e.g.* the maximum internal distance of the sample or the damping parameter (Glatter, 1977). This ensures that the $p(r)$ is obtained on a statistically sound basis without user bias. From the $p(r)$ functions, we see that the maximum distance within both the loaded and the unloaded nanodiscs is about 120 Å (see Fig. 1). The well defined maximum distances apparent from the $p(r)$ functions strongly confirm a high sample quality free of unintended large aggregates.

A further advantage of the BIFT method is that it provides a more reliable estimate of the experimental information content than the more commonly used number of Shannon channels (Vestergaard & Hansen, 2006). As part of this, the method provides an estimate of the so-called 'number of good parameters' present in the evaluated data set. This number gives a good estimate of the maximum number of independent model parameters that it is meaningful to fit to the experimental data.

The number of good parameters is calculated from the number of spline functions needed to describe a $p(r)$ function that gives a good fit to the data when the inverse Fourier transform is invoked. In this way, the quality of the data set is taken into account since fewer splines are necessary to fit a data set with large error bars.

From the BIFT analysis, the number of good parameters were calculated as 10.6 ± 0.8 in the SAXS data and 8.5 ± 0.2 in the SANS data.

The number of good parameters in the combined SAXS and SANS data set depends on the redundancy of the two measurements. If the contrast situations are similar, the information content will only be slightly improved by the second measurement. If the two situations are complementary the true number of free parameters may be as high as the sum of the numbers from the individual measurements. Unfortunately we have not found a way to formalize this, but it is safe to assume that in the data presented here the number of good parameters is significantly higher than 11 and is definitely smaller than 19.

From previous studies of the unloaded nanodisc, we found that the addition of another SANS contrast at the match point of the protein, *i.e.* 42% D_2O, only weakly constrains the fits.

4.3. Model-fit results

The data from the bR-loaded nanodiscs were fitted using (1), (2) and (3), whereas those from the unloaded nanodiscs were fitted using the previously published model of the phospholipid nanodisc (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2010). The fit results are plotted together with the experimental data for both systems in Fig. 1 and the

corresponding fit parameters are listed in Table 2. Note that each MSP has a protruding histidine tag. These were cleaved from the loaded nanodisc samples before measurement, while they remained on the unloaded nanodisc sample. The model takes this into account as described previously (Skar-Gislinge & Arleth, 2010).

We found that the bR molecule is significantly decentred in the nanodiscs but without touching the rim. The centre of mass of bR is displaced about 12 Å from the nanodisc centre (see Fig. 3), leaving room for a layer of about two phospholipids between the bR molecule and the membrane scaffolding protein. A significant tilt of bR is also observed when it is located in the nanodisc. The best model fit was obtained at a bR tilt angle of 31°. The tilt angle is defined to be zero when the mean directions of the seven transmembrane helices points in the direction of the bilayer normal. The vertical position of the bR molecule is close to centred, and the best model fit was found, when the centre of mass was displaced by −4.2 Å along the bilayer normal with respect to the bilayer centre.

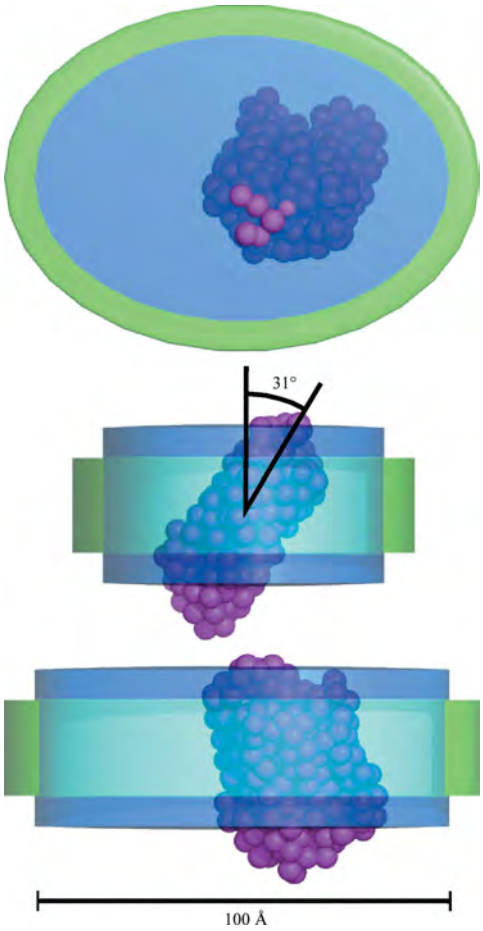


Figure 3
The best model fit to the data shown to scale viewed along the disc normal (top), major axis (middle) and minor axis (bottom). Green, membrane scaffolding protein (MSP). Cyan, hydrophobic part of the lipids. Blue, hydrophilic part of the lipids. Magenta, bacteriorhodopsin.

The number of bound water molecules at the head groups is difficult to determine using this method. This is reflected by the large confidence intervals. The fitted number of 8.7 water molecules per head group for the bR-loaded nanodisc is surprisingly high. On the other hand, zero bound water molecules in the unloaded nanodisc appears to be unrealistically low.

Owing to the high degree of symmetry of bacteriorhodopsin, it was possible to find alternative solutions with tilt angles of approximately 31° and/or a positive lateral shift. However, taking the polarity of the surface residues into account these solutions are unlikely, and we conclude that it is the region containing the C-terminus that protrudes the most, while the region around the N-terminus is partly buried in the hydrophilic head groups of the bilayer. As an example, Fig. 4 shows the most likely solution together with an alternative solution with a positive lateral shift. The amino acids are coloured according to their hydrophobicity on the Wimley–White scale (Wimley & White, 1996). The alternative solution has been deemed to be less plausible owing to a larger number of charged amino-acid residues extending into the solvent and a similar increase in the number of aromatic residues residing inside the hydrophobic core of the lipid bilayer. Furthermore, the high curvature of the bilayer leads to a large hydrophobic mismatch between the height of the MSP and the height of the hydrophobic core. Except for the sign of the vertical shift and the hydrophobic thickness at the rim, the parameters of the alternative solution all fall within the confidence intervals given in Table 2.

The presence of bR perturbs the surrounding nanodisc significantly. This is observed in the axis ratio of the discs, with a decrease from 1.66 to 1.44, and more notably in the hydrophobic interface area per lipid, where a significant change from 62.6 $Å^2$ in the unloaded discs to 78.4 $Å^2$ in the bR-loaded discs is observed. Surprisingly, the introduction of bR does not appear to displace any of the lipids. Instead, the fitted number of lipids increases from 126 in the unloaded disc to 130 in the bR-loaded disc. The combined result of these effects is a quite significant increase in the circumference of the disc from 238 Å in the unloaded disc to 275 Å in the loaded disc, an increase of 16%. While it was surprising to us that the MSP exhibits this flexibility, it is noted that the circumference of the loaded disc corresponds well to the expected maximal length of the helices constituting the MSP.

5. Discussion

Our method gives unique experimental insight into how a membrane protein is organized in a lipid-membrane environment, including how the lipids themselves adapt structurally to accommodate the membrane protein. This gives us the opportunity to test some of the many theoretical predictions

and computer simulations of these interactions (Lomize *et al.*, 2011; Mouritsen & Bloom, 1984; Israelachvili *et al.*, 1980).

From a visual inspection of the crystal structure of bacteriorhodopsin based on the polarity of the surface residues, one would estimate the hydrophobic height to be 35 Å (Lee, 2003).

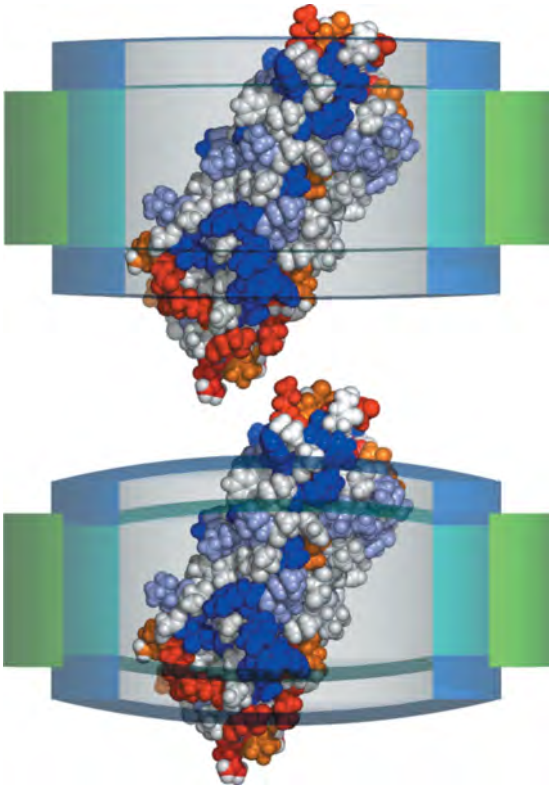


Figure 4
Top, the solution shown in Figs. 2 and 3. Bottom, an alternative solution giving an equally good fit to the data but a less favourable hydrophobic match. Blue, charged hydrophobic residues. Light blue, slightly hydrophobic residues. Orange, slightly hydrophilic residues. Red, aromatic hydrophilic residues.

Table 3
Comparison of theoretical results from the Orientation of Membrane Proteins (OMP) database (Lomize *et al.*, 2011) and experimental results from this work.

	Tilt (°)	Lateral shift (Å)	Hydrophobic thickness (Å)
OPM	24 ± 8	−3.3	29.6 ± 2.2
This work	31	−4.2	26.4 (centre)

On the other hand, the height of the hydrophobic part of an unstretched bilayer of POPC has been measured to be 27.1 Å using X-ray diffraction (Kučerka *et al.*, 2008). Finally, the hydrophobic height of the MSP is estimated to be 24 Å.

To account for this mismatch, the lipids may alter their hydrophobic height by stretching or compressing and a membrane protein may lower its hydrophobic height by tilting such that none of the transmembrane segments penetrate the membrane perpendicularly. Finally, in some situations it may be thermodynamically favourable to settle for a conformation with a slight hydrophobic mismatch. The present work is to our knowledge the first experimental quantification of the interplay between all these effects.

In this work, we clearly see how bR tilts to accommodate the mismatch of the hydrophobic heights. While we had anticipated that the lipids would stretch to cover the hydrophobic parts of bR, we observe the opposite behaviour. This may be the result of an outwards pressure on the MSP owing to a less structured packing of the lipids caused by the presence of bR. Thereby, the MSP loses flexibility and its hydrophobic height comes close to the 24 Å expected from two parallel helical segments. In a competition where the bilayer is compressed at the rim and stretched near the centre, the MSP is likely to have the largest effect because the rim is much longer than the circumference of the membrane protein (see Fig. 5).

In the same context, we observe that the bR molecule is significantly decentred in the nanodisc. Whether this observation is owing to the protein being fixed near the rim owing to more optimal lipid-bilayer packing in this region or owing to a lack of confinement of the protein position, we cannot say from the data. However, it is interesting to observe that the centre position is definitely not preferred by the membrane protein.

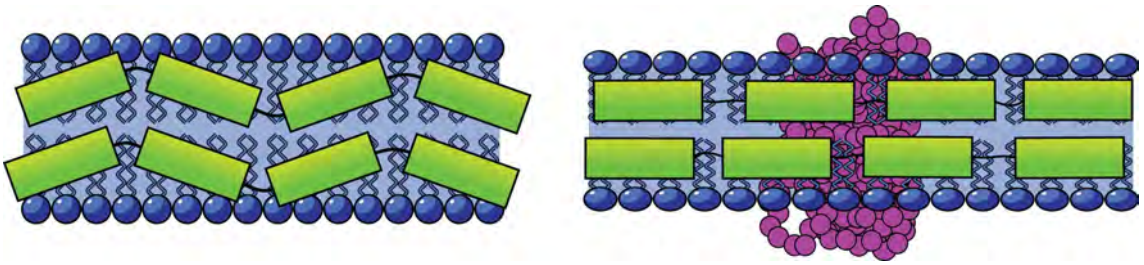


Figure 5
The loaded nanodisc (right) has a larger circumference than the unloaded nanodisc (left). This is a combined effect of the facts that more material has to fit inside the MSP (*i.e.* slightly more lipids plus the membrane protein) and that each lipid has an increased interface area. The sketch illustrates how the latter effect is self-amplifying because the lipids are perturbed by the stretched MSP of the loaded disc.

It is very interesting to compare the above results with the theoretical predictions deposited in the Orientation of Proteins in Membranes (OPM) database (Lomize *et al.*, 2011), which contains calculations of optimum lateral position and tilt angle as well as ideal bilayer thickness for most of the crystallized membrane proteins in the PDB. These calculations seek to minimize the transfer energy of the residues from water to the lipid bilayer, represented by a planar slab of adjustable thickness (Lomize *et al.*, 2006). As seen in Table 3, the tilt and lateral shift are comparable, whereas we observe a slightly lower hydrophobic thickness than the theoretical optimum predicted by the OPM. The differences may arise from the fact that the bilayer of the nanodisc is not a naturally relaxed membrane.

A few experimental studies exist that reveal structural information about proteins in membranes. In particular, bR has been investigated (Altenbach *et al.*, 1990, 1994; Dumas *et al.*, 1999). For example, experiments in which residues in the fourth α-helix (residues 103–129) of bR have been systematically replaced with spin-labelled cysteines have given insight into the depth of the residues in the lipid bilayer (Altenbach *et al.*, 1994). According to this study, residue Ile117 is centred in the bilayer. This method relies on the fact that no native cysteines are present in bR and that the mutant folds in the same way as the native protein. In the present work the lateral position of the residue Ile117 in the best fit to data is 0.6 Å relative to the centre of the bilayer, which is in excellent agreement with the spin-labelling study.

Lipid-bilayer thickness has been estimated by investigating changes in melting temperature owing to the incorporation of membrane proteins in lipid vesicles. Data for bR show that a positive shift in melting temperatures is observed for di(C12:0)PC, whereas a negative shift of comparable size is seen for di(C18:0)PC (Piknová *et al.*, 1993). This indicates that the hydrophobic matching of bR is achieved at chain lengths of around 30 Å. This is midway between the average chain lengths of the two lipid types. The slightly lens-shaped bilayer with hydrophobic height between 24.6 and 26.4 Å obtained in this study is only partly comparable to the result of the phase-transition study and the OPM value. This is because the lipid bilayer of the nanodisc is constrained at the rim by the MSP to a thickness of only 24 Å. This is less than most membrane proteins and has the effect that the bilayer becomes stretched like the head of a drum. The high degree of stress on the phospholipids may explain some of the difficulties in reconstituting certain membrane proteins in the nanodiscs. It would be interesting to investigate whether or not this could be resolved by using a mixture of lipids with different chain lengths.

5.1. Perspectives

At present, the primary bottleneck of the method is the difficulty in obtaining pure and well defined samples of reconstituted membrane proteins in sufficient amounts to allow both SAXS and SANS measurements. Assuming this can be solved, systems of interest can be grouped into two cases depending on whether or not *a priori* knowledge of the protein is present in the form of a crystal structure.

The presented approach is directly applicable in its present form to determine the localization of any membrane protein in the bilayer membrane of the nanodisc provided that the crystal structure is known. In cases where only extracellular parts have been crystallized, this could, with little effort, be combined with a good hypothesis of the membrane part and allow determination of the position and the orientation of the extracellular parts. Furthermore, it is trivial to generalize the approach to other systems in which the membrane protein is incorporated into other carrier systems, for example detergent micelles or lipid-bilayer vesicles.

For cases in which the membrane-protein structure is not known, it is the goal of our efforts to develop a free-form modelling approach comparable to the methods implemented in software such as *DAMMIN* and *GASBOR* (Svergun & Koch, 2003).

In this approach, each amino-acid residue moves freely and thus the model has hundreds of free parameters. Although we cannot hope for a unique solution, the packing constraints of the residues, such as nearest-neighbour distances, inherent to all proteins together with scattering data may be sufficient to determine the overall shape of the membrane protein. In the present case, we find that a little more experimental information is needed for this approach to be successful.

In order to obtain this, it may be advantageous to deuterate the MSP such that the MSP and the membrane protein do not have the same scattering-length density when investigated using neutrons. An even more optimal approach that we are currently investigating is to systematically control the deuteration levels of the MSP and the phospholipids (head groups and alkyl chains) to completely match the surrounding nanodisc carrier in the SANS experiment.

It is also relevant to further investigate the possibility of including other types of information. We already routinely include molecular constraints, *i.e.* information about the sample concentration, chemical composition and scattering length of the constituents, as well as partial specific molecular volumes. In this work, we have shown how this can be combined with structural information from crystallography. However, other structural methods such as cryo-EM might also provide useful constraints. It might also be possible to include local structure information from spectroscopic methods such as NMR, CD and others. Finally, general bioinformatical knowledge, including the packing density of the residues, could be systematically incorporated.

In order to make use of the improving resolution at high q at modern SAXS and SANS facilities, we are working to introduce realistic local structure on a length scale of nanometres in the nanodisc model to allow close fitting at high q. This may be performed by incorporating information about the internal structure of lipids and MSP. Under all circumstances, we envisage a continuous demand for access to combined X-ray and neutron experiments as well as the possibility of increasingly small sample volumes.

6. Conclusion

We have presented a method to determine the position and orientation of a membrane protein relative to a phospholipid nanodisc. For the specific case of bacteriorhodopsin (bR), we find that bR is displaced by 4.2 Å in the direction perpendicular to the lipid bilayer of the nanodisc. We find that bR is not confined to the centre of the disc but is more likely to be found 12 Å from the centre of the disc. The method also allows us to determine the response of the lipid bilayer. In the investigated case we observe a larger disc perimeter and a larger area per head group of the lipid molecules relative to an unloaded nanodisc. The method is applicable to other systems of membrane proteins embedded into a nanodisc.

The authors gratefully acknowledge the access to large-scale facility SANS and SAXS beamtime at the SANS beamline D11, Institute Laue–Langevin (ILL), at the BM29 (formerly ID14-3) Bio-SAXS beamline, European Synchrotron Radiation Facility (ESRF) and at KWS 1, Forschungs Neutronenquelle Heinz Maier-Leibnitz, Munich (FRM II). In this context, we gratefully acknowledge the access to simultaneous SAXS and SANS beamtime at ILL and ESRF. In addition, the authors thank Dr A. Round at ESRF and Dr H. Frielinghaus at FRM II. The authors also thank B. Justesen for preparation of the unloaded nanodisc samples. The UNIK synthetic biology project and McXtrace are gratefully acknowledged for cofunding of this research.

References

Altenbach, C., Greenhalgh, D. Å., Khorana, H. G. & Hubbell, W. L. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 1667–1671.

Altenbach, C., Matri, T., Khorona, H. G. & Hubbell, W. L. (1990). *Science*, **248**, 1088–1092.

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., Grinkova, Y. V. & Sligar, S. G. (2006). *Arch. Biochem. Biophys.* **450**, 215–222.

Bayburt, T. H. & Sligar, S. G. (2003). *Protein Sci.* **12**, 2476–2481.

Berthaud, A., Manzi, J., Pérez, J. & Mangenot, S. (2012). *J. Am. Chem. Soc.* **134**, 10080–10088.

Broyden, C. G. (1970). *IMA J. Appl. Math. Med. Biol.* **6**, 76–90.

Cabane, B., Duplessix, R. & Zemb, T. (1985). *J. Phys.* **46**, 2161–2178.

Calcutta, A., Jessen, C. M., Behrens, M. A., Oliviera, C. L. P., Renart, M. L., González-Ros, J. M., Pedersen, J. S. & Malmedal, A. (2012). *Biochim. Biophys. Acta*, **1818**, 2290–2301.

Chen, S. (1986). *Annu. Rev. Phys. Chem.* **37**, 351–399.

Dencher, N. (1982). *Methods Enzymol.* **88**, 5–10.

Dumas, F., Lebrun, M. C. & Tocanne, J.-F. (1999). *FEBS Lett.* **458**, 271–277.

El Moustaine, D., Granier, S., Doumazane, E., Scholler, P., Rahmeh, R., Bron, P., Mouillac, B., Banères, J.-L., Rondard, P. & Pin, J.-P. (2012). *Proc. Natl. Acad. Sci. USA*, **109**, 16342–16347.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. (2010). *Proteomics*, **10**, 1141–1149.

Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P. & Bowie, J. U. (2004). *J. Mol. Biol.* **335**, 297–305.

Fletcher, R. (1970). *Comput. J.* **13**, 317–322.

Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1978). *J. Appl. Cryst.* **11**, 693–694.

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.

Goldfarb, D. (1970). *Math. Comput.* **24**, 23–26.

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C. & Walz, T. (2005). *Nature (London)*, **438**, 633–638.

Green, N. (1970). *Methods Enzymol.* **18**, 418–424.

Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. (2013). *J. Am. Chem. Soc.* **135**, 1919–1925.

Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.

Israelachvili, J. N., Marcelja, S. & Horn, R. G. (1980). *Q. Rev. Biophys.* **13**, 121–200.

Kaya, H. (2004). *J. Appl. Cryst.* **37**, 223–230.

Kučerka, N., Nagle, J. F., Sachs, J. N., Feller, S. E., Pencer, J., Jackson, A. & Katsaras, J. (2008). *Biophys. J.* **95**, 2356–2367.

Kučerka, N., Nieh, M. & Katsaras, J. (2010). *Advances in Planar Lipid Bilayers and Liposomes*, Vol. 12, edited by A. Iglič, ch. 4. Oxford: Academic Press.

Lee, A. (2003). *Biochim. Biophys. Acta*, **1612**, 1–40.

Lieutenant, K., Lindner, P. & Gahler, R. (2007). *J. Appl. Cryst.* **40**, 1056–1063.

Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. (2007). *J. Phys. Chem. B*, **111**, 12427–12438.

Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2011). *J. Chem. Inf. Model.* **51**, 930–946.

Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006). *Bioinformatics*, **22**, 623–625.

Lyukmanova, E. N., Shenkarev, Z. O., Khabibullina, N. F., Kopeina, G. S., Shulepko, M. A., Paramonov, A. S., Mineev, K. S., Tikhonov, R. V., Shingarova, L. N., Petrovskaya, L. E., Dolgikh, D. A., Arseniev, A. S. & Kirpichnikov, M. P. (2012). *Biochim. Biophys. Acta*, **1818**, 349–358.

Marquardt, D. (1963). *J. Soc. Ind. Appl. Math.* **11**, 431–441.

Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sørensen, T. L.-M., Petersen, J., Andersen, J. P., Vilsen, B. & Nissen, P. (2007). *Nature (London)*, **450**, 1043–1049.

Mouritsen, O. & Bloom, M. (1984). *Biophys. J.* **46**, 141–153.

Mylonas, E. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s245–s249.

Oesterhelt, D. & Stoeckenius, W. (1974). *Methods Enzymol.* **31**, 667–678.

Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.

Pedersen, J. S. (2002). *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, edited by P. Linder & T. Zemb, ch. 16. Amsterdam: Elsevier.

Periasamy, A., Shadiac, N., Amalraj, A., Garajová, S., Nagarajan, Y., Waters, S., Mertens, H. D. T. & Hrmova, M. (2013). *Biochim. Biophys. Acta*, **1828**, 743–757.

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.

Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.

Piknová, B., Páerochon, E. & Tocanne, J.-F. (1993). *Eur. J. Biochem.* **218**, 385–396.

Press, W. H., Teukolsky, S. A., Vettering, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed. Cambridge University Press.

Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M. & Sligar, S. G. (2009). *Methods Enzymol.* **464**, 211–231.

Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. (2009). *Nature (London)*, **459**, 356–363.

Shanno, D. F. (1970). *Math. Comput.* **24**, 647–656.

Shih, A. Y., Denisov, I. G., Phillips, J. C., Sligar, S. G. & Schulten, K. (2005). *Biophys. J.* **88**, 548–556.

Skar-Gislinge, N. & Arleth, L. (2010). *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.
Sonntag, Y., Musgaard, M., Olesen, C., Schiøtt, B., Møller, J. V., Nissen, P. & Thøgersen, L. (2011). *Nature Commun.* **2**, 304.
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. J. (2003). *Rep. Prog. Phys.* **66**, 1735–1782.
Vestergaard, B. & Hansen, S. (2006). *J. Appl. Cryst.* **39**, 797–804.
Wallin, E. & von Heijne, G. (1998). *Protein Sci.* **7**, 1029–1038.
Whorton, M. R., Bokoch, M. P., Rasmussen, S. G. F., Huang, B., Zare, R. N., Kobilkaa, B. & Sunahara, R. K. (2007). *Proc. Natl. Acad. Sci. USA*, **104**, 7682–7687.
Wimley, W. C. & White, S. H. (1996). *Nature Struct. Mol. Biol.* **3**, 842–848.
Yao, X. J., Vélez Ruiz, G., Whorton, M. R., Rasmussen, S. G. F., Devree, B. T., Deupi, X., Sunahara, R. K. & Kobilka, B. (2009). *Proc. Natl. Acad. Sci. USA*, **106**, 9501–9506.

Chapter 9

Paper 5

Shape and orientation of human cytochrome P450 in a lipid environment: A method to characterize membrane proteins

Nicholas Skar-Gislinge[1], Søren Kynde[1], Ilia G. Denisov[2], Xin Ye[2], Ivan Lenov[2], Stephen G. Sligar[2], and Lise Arleth[1]

[1]Structural Biophysics, Niels Bohr Institute, Faculty of Science, University of Copenhagen, Denmark
[2]Department of Biochemistry, School of Molecular and Cellular Biology, University of Illinois, 505 South Goodwin Avenue Urbana, IL 61801, USA

March 7, 2014

1 Introduction

Membrane proteins constitute about 26% of the proteins in the humane proteome [1]. They perform many critical functions, such as cell signaling, transport, photosynthesis, biosynthesis and metabolism. However, due to their poor solubility and other experimental difficulties, the high resolution structural studies of membrane proteins are challenging. Only 1% of all protein structures resolved by X-ray crystallography or NMR and listed in Protein Data Bank represent membrane proteins, and almost all of them are obtained using proteins solubilized in detergents and other micellar colloid systems. Recently several structures of GPCR and other types of integral membrane proteins have been resolved in lipid cubic phase [2], which was a remarkable success honored by the Nobel Prize in 2012[3]. However, the problem of structural characterization of the membrane protein inserted into the native membrane, including detailed information on the geometry and mode of the protein incorporation into the lipid bilayer and functionally important protein-lipid interactions, is still unresolved.

Cytochromes P450 belong to the broad superfamily of heme enzymes (more than 20,000 isozymes identified by 2012 in genomes of organisms from all biological kingdoms) which catalyze multiple reactions essentially important for biosynthesis as well as for catabolism of xenobiotics [4]. In all eukaryotic organisms cytochrome P450s exist and function as monotopic membrane proteins with

the N-terminal 30 to 50 residues inserted into the membrane. 56 cytochromes P450 in humans are involved in biosynthesis and regulation of steroid hormones and vitamins, and also play a major role in drug metabolism. Importantly, catalytic activity of drug metabolizing cytochromes P450 critically depends on their interactions with the redox partner, another monotopic membrane bound flavoprotein NADPH-dependent cytochrome P450 reductase (CPR). As a result, the membrane properties, such as lipid composition, physical state, charge and mobility of bilayer components [5], 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[6]. 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 [7]. 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 P450 unit.

Recently, the mode of incorporation of the human cytochromes P450 into the membrane has become a subject of a number of studies. Most of them used various realizations of molecular dynamics[8] in order to evaluate the position and orientation of several isozymes in the model bilayer. Several experimental attempts were based on the accessibility probes of the surface amino acids of cytochrome P450 [9]. 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 at the broad range of conditions (temperature, pH, pressure, solvent composition) with no restriction on the lipid composition of the membrane and using a relatively small amount of sample (50 uL of 3 - 10 mg/mL solution).

Nanodiscs are ≈10 nanometer sized phospholipid bilayer discs[10], in which membrane proteins can be incorporated. They consist of a central lipid bilayer of 120 to 150 phospholipids, depending on the lipid type, and two amphiphatic membrane scaffolding proteins (MSP) stabilizing the lipids by wrapping round the rim of the bilayer in a belt like manner[11, 12]. The use of the nanodiscs to stabilize membrane proteins in solution in order to perform solution studies has previously been demonstrated in several cases[13, 14, 15]. It has previously been shown that CYP3A4 can be homogeneously incorporated as monomers into phospholipid nanodiscs [16], which is a strict requirement for 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[17]. And while the information content remains low compared to methods such as NMR and X-Ray Crystallography, the use of several datasets and the combination of SAXS and small angle neutron scattering (SANS) can greatly increase the information available for a detailed structural analysis of SAS data[12]. In the present case two SAXS datasets are analyzed in order to increase the amount of available information. As CYP3A4 is only anchored to the membrane via a hydrophobic alpha helix extending into the hydrophobic part of the phospholipid bilayer and we assume that the perturbation of the bilayer lipids by 3A4, on average, is minimal. An "empty" POPC nanodisc prepared under the same conditions as the CYP3A4 POPC Nanodisc, is used as a reference sample, from which the bilayer parameters can be extracted. These are subsequently used in the modeling 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 from water soluble proteins from small angle x-ray scattering (SAXS), can be extended to also cover membrane anchored proteins. This is made possible through the combination with a recently established description for the phospholipid nanodiscs.

We use this newly developed *hybrid* approach[18] 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 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 are also analyzed using a traditional rigid body approach where a molecular dynamics (MD) model for the CYP3A4[9] is fitted to the data in combination with a model for the nanodisc.

The resulting *ab initio* models for the CYP3A4 are in very good agreement with the corresponding MD simulation. Furthermore, there is a very good agreement between the results for the localization of the CYP3A4 in the nanodisc as obtained through the *ab initio* and rigid body approaches.

The *ab initio* approach[19] has proven to be a very powerful method for studying mono disperse soluble proteins using small angle scattering [20, 21] 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 comparable structural techniques such as Cryo-EM[22] and NMR[23], will be able to accelerate studies of membrane proteins, that have previously been inaccessible for structural studies.

2 Material and Methods

2.1 Preparation of CYP3A4 POPC Nanodiscs

Expression and purification of CYP3A4 Cytochrome P450 3A4 was expressed from the NF-14 construct in the PCWori+ vector, provided by Dr. F. Peter Guengerich, Vanderbilt University, Nashville, TN. The CYP3A4 was expressed and purified as previously described[24] and its concentration determined by carbonmonoxy difference spectra[25].

Purification of membrane scaffold protein Expression and purification of MSP1D1(-) has been previously described[10]. The MSP1D1(-) was engineered with a tobacco etch virus (TEV) protease site to cleave the N-terminal hexahistidine tag used for purification, as described earlier[10, 11]. The MSP1D1 was cleaved using a 100:1 MSP1D1:TEV protease weight ratio in 20 mM Tris/HCl, pH 7.4, 0.1 M NaCl, and 0.01% NaN3 (standard disc buffer) at 28 °C overnight. MSP1D1(-) was purified on a Ni-NTA column and concentrated. Concentration was determined by spectrophotometric analysis at 280 nm.

Self-assembly and Puryfication of Nanodiscs containing Cytochrome P450 3A4 Self-assembly of CYP3A4 Nanodiscs is based on previously described procedures [16]. The dried POPC phospholipids (Avanti Polar Lipids, Alabaster, AL) used in these experiments were solubilized using 100 mM cholate solution to make a final lipid concentration of 50 mM. CYP3A4 solubilized in 0.1% (v/v) Emulgen 913 (Karlan Research Products, Santa Rosa, CA) was combined with MSP1D1(-), POPC phospholipids, and cholate in a molecular ratio of 20:1:64:128 (CYP3A4: MSP1D1(-):POPC:cholate). After incubation, the self-assembly process is initiated by removing cholate and Emulgen 913 by a 4 hour incubation on an orbital shaker at 4 °C in the presence of Amberlite XAD-2 Adsorbent (0.5g per every ml of mixture).

Assembled Nanodiscs are seprated into fractions containing empty Nanodiscs and Nanodiscs containing CYP3A4 on a Ni-NTA column using a pentahistidine affinity tag on 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, using a Waters (Milford, MA) HPLC system equipped with a photodiode array detector.

Self-assembly of empty and Purification of empty POPC Nanodiscs The POPC phospholipids used in these experiments were solubilized using 100 mM cholate solution to make a final lipid concentration of 50 mM. Cholate solubilized MSP1D1(-) is added to the solubilized

POPC phospholipids in a ratio of 65:1 (POPC:MSP1D1(-)). The mixture is incubated at 4 °C for 15 minutes. The self-assembly process is initiated by removing cholate by a 4 hour incubation on an orbital shaker at 4 °C in the presence of Amberlite XAD-2 Adsorbent. The mixture is then filtered and fractioned on a Superdex 200 10/300 column, equilibrated with standard disc buffer running at 0.5 ml/min.

2.2 Modeling of Small Angle Scattering Data

2.2.1 Continuous Modeling

The quantity measured by small-angle scattering is the differential scattering cross-section $\frac{d\Sigma}{d\Omega}$ 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

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

where n is the number density of the particles, $A(\mathbf{q})$ is the scattering amplitude of a single particle and $\langle \cdots \rangle_\Omega$ denotes 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.

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

From this equation the scattering amplitudes of many simple scattering length density functions $\Delta\rho(\mathbf{r})$ have been calculated analytically or semi-analytically[26].

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 is represented by a hollow cylinder, with an inner radius corresponding to the radius of the lipid bilayer and a volume based on the composition two MSPs. The model has been described in more detail in a previous publication[27].

2.2.2 Discrete Modeling

Solving the integral of equation (2) becomes become tedious for objects with complex shapes, such as proteins. In this case the complicated shape may 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[28]:

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

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

here J_l is the bessel function of the first kind, Δb_j the scattering length of the j'th point, and r_j, θ_j and ϕ_j the coordinates of the j'th point and Y_{lm}^* the complex conjugate of the spherical harmonic. This point model approach is used to describe the CYP3A4 molecule.

2.2.3 Hybrid Modeling

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:

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

This hybrid approach is utilized in two different models of 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.

Ab Initio Modeling Representing the membrane protein as a number of points, while representing the nanodisc by a geometrical model, opens for performing *ab initio* shape reconstructions along the lines described by *Svergun et al.*[21]. The parameters describing the nanodisc is 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 equation (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 40Å. 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, $C(X)$ 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 to the PDB data bank. In addition the these constraints, if there are less than 14 amino acids in the lipid bilayer $T(X)$ is calculated as $T(X) = (n - 14)^2$ if there are more than 14 $T(X) = 0$. This was done in order to promote configurations that extend in to the bilayer. If a new configuration, X_2, had $F(X_1) > F(X_2)$ it was accepted, on the other hand If $F(X_1) < F(X_2)$ the move to configuration X_2 was accepted only with the probability

$$P = e^{-(F(X_2)-F(X_1))/B}. \tag{7}$$

One pass completed after n_{aa} successful moves, where n_{aa} is the number of movable points, in this case the number of amino acids. Initially B was set to $\chi^2(X_0)/10$ but after each pass B was decreased by 5%.

Rigid Body Modeling To evaluate the results of the *ab initio* modeling a rigid body model of the CYP3A4 molecule is constructed from a previously published molecular dynamics (MD) simulation, of the full length CYP3A4 inserted into a POPC bilayer[9]. This provides a realistic model of the solution structure of the protein inserted into a lipoid 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.

The atomic structure from the MD model is coarse grained by placing a point scatterer at the center 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 equation 4.

In the rigid body approach equation (5) is again fitted to the scattering data from the CYP3A4 nanodisc. This time with a more traditional steepest descend algorithm[26] minimizing χ^2 with respect to the orientation and relative position of the CYP3A4 protein and the nanodisc.

When combining the CYP3A4 and nanodisc models, the scattering length of every point placed in 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.

2.3 Small-Angle Scattering Measurements

The measurements were performed at the BM29 BioSAXS beamline at the ESRF (Grenoble, France) using the fixed settings of the instrument[29]. The recorded detector intensities were automatically radially averaged and absolute scaled using a H_2O standard [30] by means of the software at the beam line. The resulting scattering curves were subsequently re-binned to about 100 experimental data points in order to improve the statistic on the single data points.

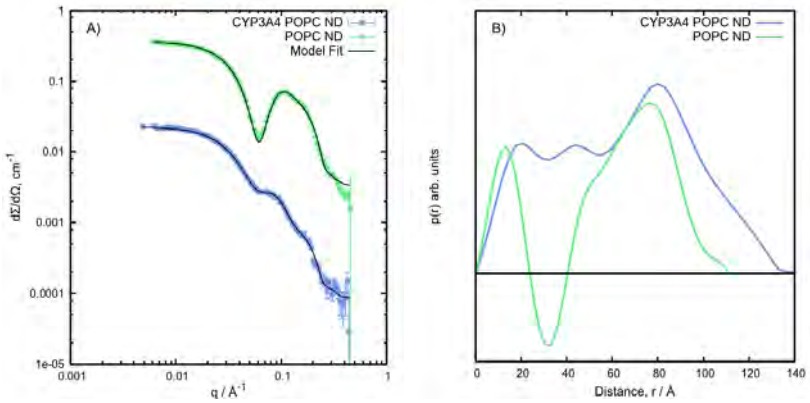


Figure 1: A) Small-angle scattering data of POPC nanodisc (green) and POPC nanodisc with reconstituted CYP3A4 (blue). The full black curves show the corresponding model fits (see text). B) Pair distance distribution functions of the data shown in A, obtained using the Bayesian indirect Fourier transform method.

Indirect Fourier Transform The calibrated scattering curves gives the differential cross-section of the sample as a function of $q = 4\pi\frac{\sin(\theta)}{\lambda}$, where λ is the wavelength of the X-ray source and θ is half the angle between the incident and the scattered beam. The same data can be represented by a real space function by performing a so called indirect Fourier transform[31], yielding the pair distance distribution function (PDDF). All PDDFs presented here were obtained using the Bayesian Indirect Fourier Transform (BIFT) method[32][33]. 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, than the number of Shannon channels, as it takes into account the experimental errors on the data.

3 Results

3.1 Model Free Analysis

The background subtracted SAXS curves from the POPC nanodiscs and the CYP3A4 nanodisc are shown in figure 1 A). Figure 1 B) shows the corresponding Pair 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 [12, 34]. The PDDFs show that the maximum distance present in the samples increase from 110 Å to 135

Å when CYP3A4 is reconstituted in the nanodiscs, however due to the complex nature of the particles a more detailed interpretation is not possible without the use of a model.

Estimation of Information Content In order to avoid over fitting the data when performing model analysis, the information content is estimated, in terms of the number of good parameters, N_g, as provided by the BIFT analysis. For the POPC nanodisc, the number of good parameters are N_g =11.3 ± 0.1, and for the CYP3A4 nanodisc N_g =8.0 ± 0.5.

3.2 Model Analysis

3.2.1 POPC Nanodisc

The scattering data form the POPC nanodiscs are interpreted using a previously published molecular constrained model of the small angle scattering from phospholipid nanodiscs[27], where the nanodisc is modeled 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 equation (2). By systematically incorporating constraints based on the molecular composition of the nanodisc, the number of free parameters in this model is reduced to a total of 7, significantly less than N_g for the data set. The fit to the measured POPC nanodisc data using this model is shown in black in figure 1 A) and the resulting fitting parameters are shown in table 1.

3.2.2 CYP3A4-POPC nanodisc

***Ab Initio* Shape Reconstruction** By describing the membrane protein as an assembly of points, representing each amino acid, the hybrid model allows for making an *ab initio* reconstruction of the shape of a membrane protein anchored in a nanodisc.

Figure 2 shows the results of 10 *ab initio* shape reconstructions. Although the orientation of the 10 *ab initio* models are oriented differently in the nanodisc, they all exhibit an over all similar shape, in particular they all exhibit a stalk like domain inserted into the membrane and a globular domain extending above the membrane. We find no solutions where the CYP3A4 molecule is placed near the center of the nanodisc.

The models were aligned and combined to produce a consensus model using the program DAMAVER [20]. By remapping the models onto a grid, DAMAVER align the models and count the frequency a given point in the grid occurs in the reconstructed models. The combined model is shown in two representations in figure 3 A and B: In figure 3 A each point of the combined model is

Parameter	POPC ND	3A4 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
ν_{Belt}, Å^3	25261.5 ± 71.7	-
ν_{POPC}, Å^3	1292.91	-
ν^a_{PO}, Å^3	961.9	-
ν^a_{PC}, Å^3	331.01	-
ν_{CYP3A4}, Å^3	-	73661 ± 252

Table 1: Fitting parameters for the nanodisc model. All parameters except the interface roughness terms were calculated from the model fit to 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 measure the average interfacial roughness of all interfaces, ν_{Belt} is the partial specific molecular volume of one MSP, ν_{POPC} is the partial specific molecular volume of one POPC molecule and ν_{CYP3A4} is the partial specific molecular volume of CYP3A4. ^{a}: Deduced from the total lipid volume, not an explicit model parameter.[27]

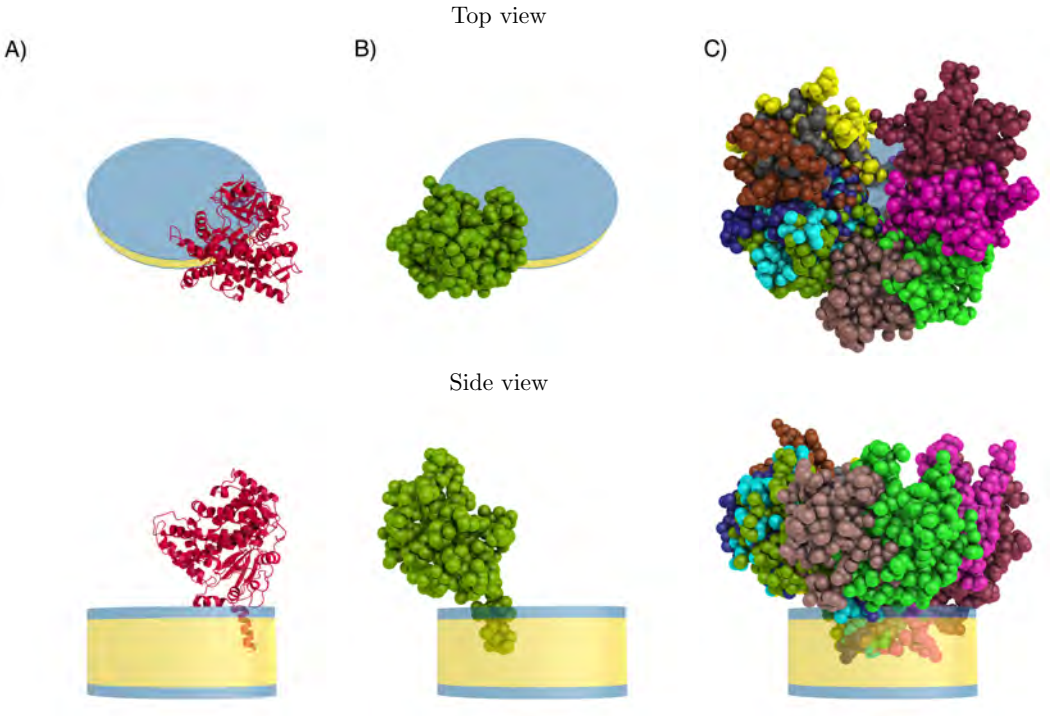


Figure 2: A) Result of rigid body analysis of the scattering data. B) Top and side view of one particular result of an *ab initio* analysis of the scattering data. C) Top and side view of 10 results of subsequent *ab initio* analyses.

represented with a sphere. I3 B each point, of the combined model is represent by a sphere with a radius of $\sqrt{N}$, where N is the number of times that the point occurs in the 10 aligned shape reconstructions. This gives a picture of how conserved each a given point is in the 10 *ab initio* model. Both representations of the combined model is compared to the model of CYP3A4 used in the rigid body analysis, shown in figure 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 the semi-transparent surface in figure 3 B.

Rigid Body Analysis As described in the material and methods section 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[9]. Using the hybrid approach this was combined with a model for the nanodisc using the same model and parameters as the POPC nanodisc. The resulting number of fitting parameters in this model is 6. These are: the three spacial coordinates and the rotation around the bilayer normal of CYP3A4, a correction

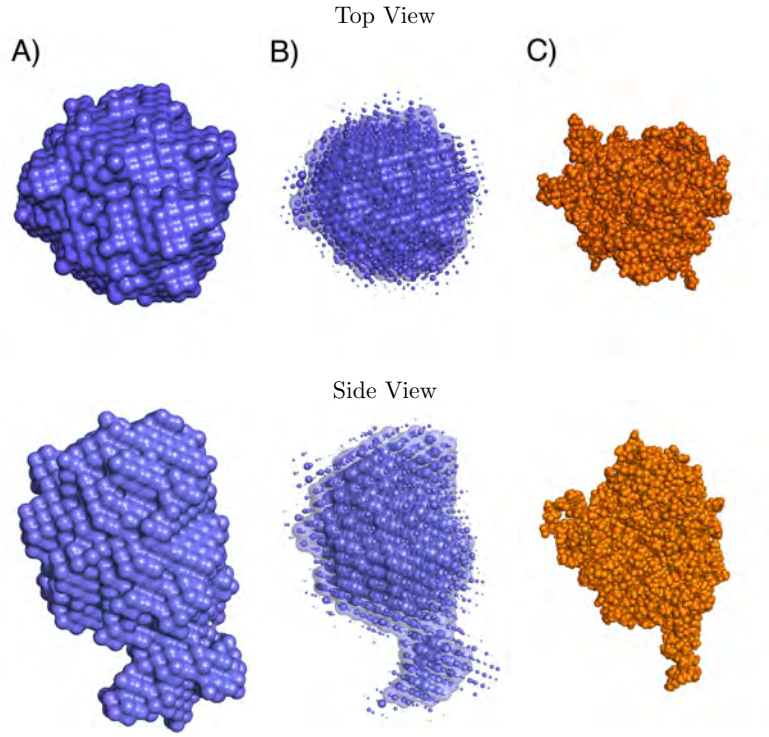


Figure 3: *Ab initio* shape reconstructions of CYP3A4 in nanodiscs. A: Average of all 10 models. B: The average model is shown as spheres where the radius of each sphere is the root of the frequency it appears in all 10 models. The transparent surface represent filtered model calculated via the DAMFILT routine. C: Space filling rendition of the atomic model of CYP3A4 used for the rigid body analysis. All models are drawn to the same scale.

factor of the density of the CYP3A4 molecule and a roughness of the phospholipid bilayer.

The model fit to the scattering data of the CYP3A4 nanodiscs is shown in black in figure 1 A and the resulting model is shown in figure 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 is clear that the solution found using the rigid body analysis is in fact one of several possible solutions.

4 Discussion

The shape reconstruction obtained from the *ab initio* analysis of the SAXS data show an excellent agreement with rigid body analysis by reproducing shape as well as the mode of incorporation. Both methods confirm that the protein is anchored to the membrane via a helix extending into the bilayer. Compared to the one solution found in the rigid body analysis, the *ab initio* results reveal a number of equivalent solutions with the same over all shape distributed along the rim of the nanodisc.

In broader perspective, we find, that is 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 already widely applied for interpreting *ab initio* data from water soluble proteins.

The *ab initio* analyses of the measured data all place the CYP3A4 molecule at the rim of the phospolipid bilayer. This observation can be explained if the CYP3A4 molecule is able to move in the bilayer plane, because the area of the edge region is larger than the area of the center region, increasing the possibility of observing it there. This is in good agreement with the notion that molecule is able to move in the bilayer plane in order to react with its redox partner, CPR, in the natural membrane. The fact that none of the *ab initio* models were found near the center of the nanodisc does not necessarily mean that CYP3A4 is never found there, however, it does show that the center of the discs is not preferred over the edge of the membrane. The apparently even distribution of the CYP3A4 along the edge of the nanodisc, indicate that there is no interaction between the CYP3A4 molecule and any specific part of the two scaffolding proteins. It has previously been shown that the lipids along the edge of the disc are slightly perturbed by the proximity of the MSPs[35, 12], this perturbation may to some degree, be lessened when CYP3A4 is located at the rim replacing replace some of the perturbed edge phospholipids. In this case the a placement along the rim of the lipid bilayer is preferred compare to the center of the disc.

In this context, it is particularly interesting to note, that even though the sample is probably not structurally homogeneous, in respect to the location of CYP3A4 in the nanodisc, the information on the shape of the membrane protein can still be reconstructed through the hybrid approach *ab initio* modelling.

In general SAS from membrane proteins incorporated in nanodiscs will, in addition to information on the membrane protein, also contain direct structural information on the phospholipid bilayer. This information is valuable for the detailed understanding membrane protein function and the understanding of protein-protein interactions in functional complexes of membrane proteins, such as human cytochromes P450 and their redox partner CPR. In the present case we observe that the perturbation of the lipid bilayer by the CYP3A4 membrane anchor is minimal as well as indication of CYP3A4 being able to move in the bilayer plane. This information is not easily

available from other structural techniques such as Cryo-EM, due to the low electron density of the lipids [22], or in NMR schemes where the signal from the nanodisc is suppressed using deuterated phospholipids and MSP [23].

By changing lipid composition of nanodiscs, the effect of bilayer properties on the mode of incorporation, protein mobility and protein-protein interactions may be evaluated. By combining SAXS with small angle neutron scattering it may also be possible investigate the interplay between membrane proteins an the phospholipid bilayer in cases where the perturbation of the lipids by the protein is larger than in the case of CYP3A4.

Acknowledgments

This work was supported by NIH grants GM33775 and GM31756 to S.G.Sligar.

References

[1] von Heijne, G., Fagerberg, L., Jonasson, K., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010).

[2] Cherezov, V. Lipidic cubic phase technologies for membrane protein structural studies. *Current Opinion In Structural Biology* **21**, 559–566 (2011).

[3] Weis, W. I. *et al.* Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).

[4] de Montellano, P. R. O. Cytochrome P450 689 (2005).

[5] Schuler, M. A., Denisov, I. G. & Sligar, S. Nanodiscs as a new tool to examine lipid-protein interactions. *Methods in molecular biology (Clifton, N.J.)* **974**, 415–433 (2013).

[6] McLean, M. A., Grinkova, Y. V., Denisov, I. G. & Sligar, S. Oxidase uncoupling in heme monooxygenases: Human cytochrome P450 CYP3A4 in Nanodiscs. *Biochemical And Biophysical Research Communications* **430**, 1223–1227 (2013).

[7] Yu, X., Cojocaru, V. & Wade, R. C. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnology and applied biochemistry* **60**, 134–145 (2013).

[8] Otyepka, M., Berka, K. & Anzenbacher, P. Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Current drug metabolism* **13**, 130–142 (2012).

[9] Denisov, I. G., Shih, A. Y. & Sligar, S. Structural differences between soluble and membrane bound cytochrome P450s. *Journal of inorganic biochemistry* **108**, 150–158 (2012).

[10] Bayburt, T. H., Grinkova, Y. V. & Sligar, S. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters* **2**, 853–856 (2002).

[11] Denisov, I. G., Grinkova, Y. V., Lazarides, A. & Sligar, S. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *Journal Of The American Chemical Society* **126**, 3477–3487 (2004).

[12] Skar-Gislinge, N. *et al.* Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein. *Journal Of The American Chemical Society* **123**, 13713–13722 (2010).

[13] Bayburt, T. H. & Sligar, S. Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. *Protein Science* **12**, 2476–2481 (2003).

[14] Duan, H., Schuler, M. A., Civjan, N. & Sligar, S. Co-incorporation of heterologously expressed Arabidopsis cytochrome P450 and P450 reductase into soluble nanoscale lipid bilayers. *Archives Of Biochemistry And Biophysics* **424**, 141–153 (2004).

[15] Pureza, V. S., Sligar, S., Morrissey, J. H. & Shaw, A. The Local Phospholipid Environment Modulates the Activation of Blood Clotting. *Journal Of Biological Chemistry* **282**, 6556–6563 (2006).

[16] Baas, B. J., Sligar, S. & Denisov, I. G. ScienceDirect.com - Archives of Biochemistry and Biophysics - Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment. *Archives Of Biochemistry And Biophysics* (2004).

[17] Andersen, H. D., Wang, C., Peters, G. H., Arleth, L. & Westh, P. Reconciliation of opposing views on membrane-sugar interactions. *PNAS* **108**, 1874–1878 (2011).

[18] Kynde, S. A. R. *et al.* Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. *Acta Crystallographica Section D Biological Crystallography* **70**, 371–383 (2014).

[19] Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal* **76**, 2879–2886 (1999).

[20] Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal Of Applied Crystallography* **36**, 860–864 (2003).

[21] Svergun, D. I., Koch, M. H. & Petoukhov, M. V. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* **80**, 2946–2953 (2001).

[22] Frauenfeld, J. *et al.* Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nature Structural & Molecular Biology* **18**, 614–621 (2011).

[23] Wagner, G., Hagn, F., Raschle, T. & Etzkorn, M. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal Of The American Chemical Society* 130108043754005 (2013).

[24] Guengerich, F. P., Hosea, N. A. & Miller, G. P. Elucidation of Distinct Ligand Binding Sites for Cytochrome P450 3A4. *Biochemistry* **39**, 5929–5939 (2000).

[25] Omura, T. & Sato, R. The Carbon Monoxide-Binding Pigment of Liver Microsomes. Ii. Solubilization, Purification, and Properties. *Journal Of Biological Chemistry* **239**, 2379–2385 (1964).

[26] Skov Pedersen, J. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Advances In Colloid And Interface Science* **70**, 171–210 (1997).

[27] Skar-Gislinge, N. & Arleth, L. Small-angle scattering from phospholipid nanodiscs: derivation and refinement of a molecular constrained analytical model form factor. *Phys. Chem. Chem. Phys.* **13**, 3161–3170 (2010).

[28] Svergun, D. I., Barberato, C. & Koch, M. H. CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal Of Applied Crystallography* **28**, 768–773 (1995).

[29] Pernot, P. *et al.* New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Journal of Physics: Conference Series* **247**, 012009 (2010).

[30] Bergmann, A., Glatter, O. & Orthaber, D. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal Of Applied Crystallography* **33**, 218–225 (2000).

[31] Glatter, O. A New Method for the Evaluation of Small-Angle Scattering Data. *Journal Of Applied Crystallography* **10**, 415–421 (1977).

[32] Hansen, S. L. BayesApp: a web site for indirect transformation of small-angle scattering data. *Journal Of Applied Crystallography* **45**, 566–567 (2012).

[33] Vestergaard, B. & Hansen, S. L. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *Journal Of Applied Crystallography* **39**, 797–804 (2006).

[34] Makris, T. M., Denisov, I. G., Sligar, S. & Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chemical Reviews* **105**, 2253–2278 (2005).

[35] McLean, M. A., Denisov, I. G., Shaw, A., Grinkova, Y. V. & Sligar, S. Thermotropic phase transition in soluble nanoscale lipid bilayers. *Journal Of Physical Chemistry B* **109**, 15580–15588 (2005).

Chapter 10

Experimental Work on the Membrane Protein Tissue Factor

The experimental work on Tissue Factor was done as a part of a one month stay at the Sligar lab in October 2011. The main goal was to investigate the influence of the lipid environment on the membrane protein Tissue Factor using SAXS. A secondary goal was to get experience in some of the techniques of molecular biology needed, in order to produce, handle and reconstitute membrane proteins into nanodiscs.

Tissue Factor is a trans membrane enzyme that participate in the blood coagulation cascade by forming a complex with the protein Factor VII, that in turn activates other coagulation factors. The activity of the complex is highly dependent on the local concentration of PS lipids and calcium ions[1][2]. Using nanodiscs it is possible to control the phospholipid mixture, avoiding seggregation[2]. Previous solution SAXS studies of a truncated tissue factor without the trans membrane domain show a structural change when exposed to calcium ions[3].

Reconstitution of Tissue Factor in anionic lipids, in particular phosphatidylserines, and presence of Ca-ions increase the activity of the TF-FactorVII complex[4]. In this study we aimed to investigate the influence of calcium ions and anionic lipids on the Tissue Factor in its trans membrane form, as well as understand the experimental conditions of further studies of the TF-FactorVII complex in nanodiscs.

1 Preparation of Tissue Factor Nanodiscs

Tissue Factor (TF) was expressed according to the protocol developed by Smith *et al*[5] for incorporating TF into liposomes and later modified for reconstituting TF in nanodiscs [6]. The particular Tissue Factor used in this case was a recombinant Tissue Factor without the cytoplasmic domain, known as delTF. The delTF is furthermore optimized for expression in E. coli. and features a histidine tag for purification on a nickel column. The presence of a histidine tag is used both for the purification of the delTF alone, as well as purifying the delTF containing nanodiscs.

1.1 Expression

The e. coli used for expression came from a frozen stock of e. coli containing the delTF plasmid. A small colony from the frozen stock was inoculated in MDG medium and grown over night, at 37° C 300 rpm, to an optical density of 0.8 at 600 nm. 0.5ml of this overnight culture was added to six erlenmeyer flasks each containing 400mlL of BYE media and placed in the incubator at 25° C at 250 rpm. After 2 hrs the expression of delTF was then induced by adding 0.5M IPTG to a final concentration 100μM. The 6 cultures were then grown for 12 hrs at 25° C while shaken at 250 rpm and the cells were pelleted by centrifugation and the pellets were stored at -80° C.

1.1.1 Media for Expression

BYE medium (1L) BYE medium prepared by mixing 5g glycerol, 850mL purified water and 5g Bacto yeast extract. The volume was then brought to 925mL by adding purified water and

then autoclaved. Finally 75mL 10xTB salts and 2mL 50mg/ml kanamycin was added.

10x TB Salts 1L Mix 850mL purified water, 23.12g KH_2PO_4 and 125.41g K_2HPO_4. Add water to a final volume of 1L and autoclave. pH approx. 7.5.

50 mg/ml Kanamycin 500mg kanamycin dissolved in 10ml purified water and sterile filtered.

0.5 M IPTG 1.19g IPTG (isopropyl β-D-1-thiogalactopyranoside) in 10ml purified water. Sterile filtered.

40% (w/v) Glucose 40 g glucose in 74 mL purified water. Autoclaved.

MDG (1 L) The MDG medium was prepared using 915 ml purified water, 2mL 1M $MgSO_4$, 200 μL 1000x trace metals mix (see below), 12.5 mL 40% glucose, 50 mL 5% aspartate, 20 mL 50x M and 2 mL 50 mg/mL kanamycin. The solution was mixed between each addition and finally sterile filtered.

50xM (100 mL) The 50xM buffer consists of 80 mL purified water, 17.75g Na_2HPO_4, 17.0 g KH_2PO_4, 13.4 g NH_4Cl and 3.55 g Na_2SO_4. pH of when diluted 50 fold should be 6.7.

5% Aspartate (100 mL) 5 g aspartic acid dissolved in 97 mL purified water. Set to neutral pH using NaOH (1̃.6 g)

1000x Trace metals mix The 1000x trace metals mix used for the MDG buffer was prepared from 36 mL purified water and the following solutions: 50 mL 0.1 M $FeCl_3$-$6H_2O$, 2 mL 1 M $CaCl_2$, 1 mL 1 M $MnCl_3$-$4H_2O$, 1 mL 1 M $ZnSO_4$-$7H_2O$, 1 mL 0.2 M $CoCl_2$-$2H_2O$, 2 mL 0.1 M $CuCl_2$-$2H_2O$, 2 mL 0.1 M Na_2MoO_4-$2H_2O$, 2 mL 0.1 M Na_2SeO_3-$5H_2O$ and 2 mL 0.1 M H_3BO_3.

1.2 Purification

Each cell pellet was dissolved in a 30 mL lysis buffer and 3 μL Benzonase (Novagen cat. no. 71205-3) and stirred for 15 mins at room temperature. A small amount of Lysozyme was also added, however the precise concentration is not known. When the pellets were dissolved Triton X-100 was added to a final concentration of concentration of 0.5% v/v and the solution was mixed well. Larger aggregates were then pelleted by centrifugation at 20000 g for 15 min. According to the protocol many bacterial proteins could removed by adding Q-Sepharose beads to the supernatant (pH 7.4) and shaking it for 40 min at 4° C. Afterwards the beads were removed by filtration.

Further purification was done by adding 2 M imidiazole and 5 M NaCl, to a final concentration of 20mM and 300mM respectively, and letting the solution precipitate for 2 hrs on ice. The precipitate was removed by centrifugation. The final purification was done by applying the supernatant to the Ni-NTA column, binding the delTF to a nickel column via the histidine tag. The column was then washed with 8 ml of blinding buffer removing unbound proteins. The bound delTF was released by applying 9 mL of the elution buffer to the column and collected in 1 ml fractions. Using SDS gel electrophoresis and A_{280}, the delTF containing fractions were identified and pooled, and subsequently dialysed against the HBS-OG buffer overnight at room temperature, removing imidiazole and residual triton X-100.

1.2.1 Material for purification

HBS-50: 30mM Hepes-NaOH pH 7.4, 50mM NaCl and 0.01% NaN_3

Lysis Buffer: HBS-50 buffer containing 2% (w/v) OG.

HBS-OG: 1% (w/v) OG, 30mM Hepes-NaOH pH 7.4, 100 mM NaCl and 0.01% NaN_3

HPS-Triton: 0.1% (v/v) Triton X-100, 30mM Hepes-NaOH pH 7.4, 100 mM NaCl and 0.01% NaN_3

Q-Sepharose: Q-Sepharose beads (Sigma-Aldrich cat. no. Q1126)equilibrated against HBS-50.

Ni-NTA column 1.5 mL Ni-NTA superflow column (Qiagen cat. no. 30622). Equilibrated against 10 mL NTA binding buffer.

NTA Binding buffer 1% (w/v) OG, 25 mM Sodium phosphate buffer pH 8, 300 mM NaCl, 20 mM imidiazole pH 8.

NTA Elution buffer 1% (w/v) OG, 25 mM Sodium phosphate buffer pH 8, 300 mM NaCl, 500 mM imidiazole pH 8.

1.3 Reconstitution

The delTF was reconstituted in both POPC and 30%POPS/70%POPC nanodiscs. The nanodiscs were prepared following the standard procedure[7], but with OG stabilized delTF was added at

the same time as the MSP[6], at a MSP to delTF ratio of 20:1 in order to avoid nanodiscs with more than one delTF. The assembled discs were purified using size exclusion chromatography and the delTF loaded nanodiscs were separated from the empty discs by binding the histidine tag on the delTF to a Ni-NTA column.

The samples were frozen in 20 mM Tris pH 7.4 with 100 mM NaCl 15% glycerol. Before the SAXS measurements the samples were thawed and the glycerol removed from the buffer by spin filtration.

2 Small-Angle X-ray Scattering

The scattering experiment were carried out at the Chess Synchrotron at Cornell university in New York, at the F2 bioSAXS beam line[8]. Four samples were measured: TF in POPC nanodiscs with and without calcium ions and TF in 30%POPS/70%POPC with and without calcium.

The resulting scattering data are shown on the left in figure 1 along with the corresponding indirect Fourier transforms on the right. First of all, the IFTs show that the samples are all aggregated to some degree, but they also show that adding Ca^{2+} ions has a dramatic effect on the aggregation of TF POPC nanodiscs compared to the POPS/POPC nanodiscs. When comparing the scattering curves there is also a more subtle effect of adding Ca^{2+} to TF in POPS/POPC nanodiscs. This can be seen in figure 2 where the scattering data of TF in POPS/POPC nanodiscs with and without Ca^{2+} are plotted. Here, a clear change in the shape intermediate part of the scattering curve is observed as calcium is added, meaning that there is a structural change as Ca^{2+} is added. However the nature of the change can not be interpreted from the scattering data without a detailed model analysis.

Form the crystal structure of the soluble domain of TF and the size of the nanodisc a maximum length of about 150 Å would be expected, from the IFTs it can be seen that this is not the case. The long tails, extending well beyond 150 Å, present on all the PDDFs in figure 1 indicate a small but significant degree of aggregation present in the sample. This makes a detailed structural analysis of the data unfeasible.

3 Discussion

The wider scientific goal of studying the membrane protein membrane interactions and possible structural effects of the membrane environment on tissue factor, was not reached. However, a partial goal of this experiment was to gain experience in producing membrane proteins, purifying them and reconstituting them into nanodiscs. In this light, at least a partial goal was reached.

The quality of the samples were good enough for performing small-angle scattering experiments, however a low amount of aggregation made a further detailed model analysis unfeasible. All sample used for this study had been frozen in 15% glycerol for transport to the synchrotron. The freezing and subsequent removal of the glycerol may have induced the aggregation. Based on these experiences, a size exclusion step was introduced immediately before the SAXS measurements

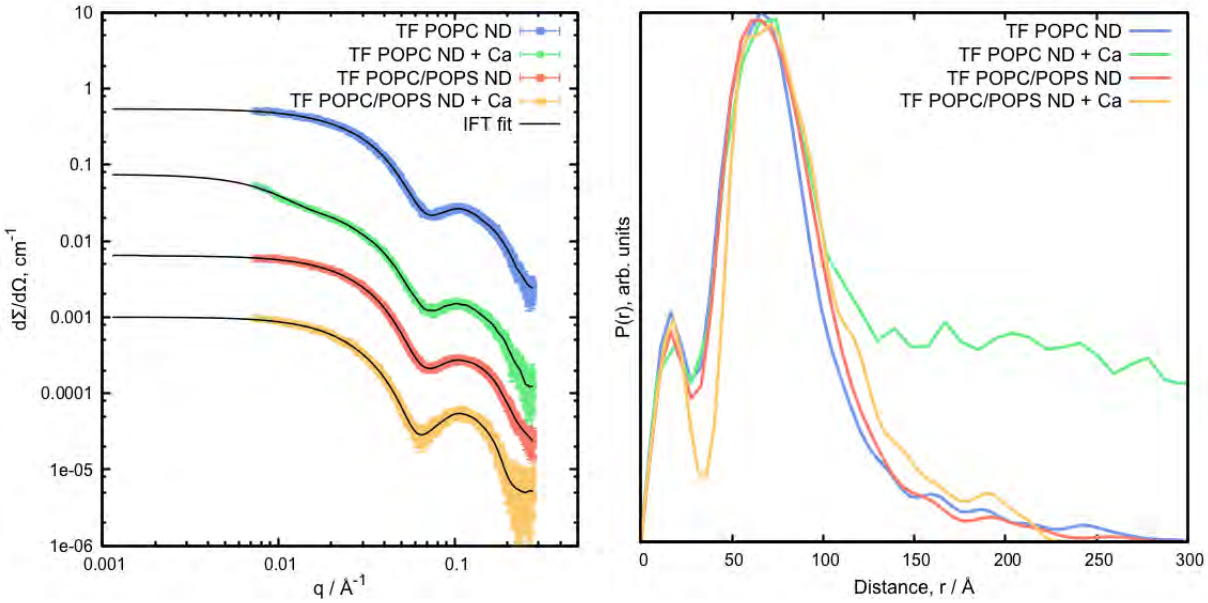


Figure 1: Left: SAXS data of Tissue Factor incorporated in POPC and 30%POPS/70%POPC nanodiscs. Subsequently Ca^{2+} ions were added and the samples measured again. The indirect fourier transforms corresponding the the pair distance distribution functions in the right plot are shown in black. Right: Pair distance distribution function of the data shown left.

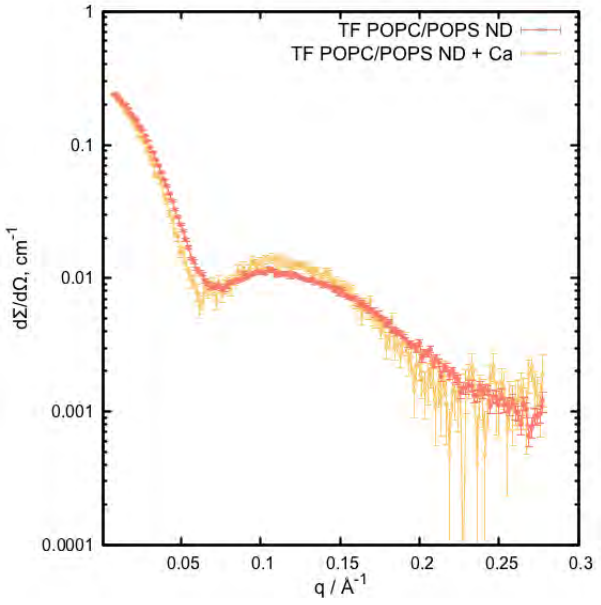


Figure 2: SAXS data of tissue factor incorporated in 30%POPS/70%POPC with and without Ca^{2+}.

in paper 5, to remove any aggregates formed during transport.

The experiment did show a clear effect of adding calcium to the samples, indicating that information on possible interaction between the phospholipids and the reconstituted membrane protein may be extracted from small-angle scattering data. Both PC and PS lipids are known to interact with Ca^{2+}[9][10], however the influence of the change in the lipid environment on TF could, unfortunately not be investigated from these data. A future experiment should try to take this in to account, by comparing the structural changes in nanodiscs with a comparable lipid mixture with and without calcium ions.

References

[1] A. W. Shaw, V. S. Pureza, V. S. Pureza, S. G. Sligar, S. G. Sligar, J. H. Morrissey, and J. H. Morrissey, "The Local Phospholipid Environment Modulates the Activation of Blood Clotting," *Journal Of Biological Chemistry*, vol. 282, pp. 6556–6563, Dec. 2006.

[2] N. Tavoosi, N. Tavoosi, R. L. Davis-Harrison, T. V. Pogorelov, T. V. Pogorelov, Y. Z. Ohkubo, M. J. Arcario, M. J. Arcario, M. C. Clay, M. C. Clay, C. M. Rienstra, E. Tajkhorshid, and J. H. Morrissey, "Molecular Determinants of Phospholipid Synergy in Blood Clotting," *Journal Of Biological Chemistry*, vol. 286, pp. 23247–23253, June 2011.

[3] C. R. Mosbaek, D. Nolan, E. Persson, D. I. Svergun, J. T. Bukrinsky, and B. Vestergaard, "Extensive small-angle X-ray scattering studies of blood coagulation factor VIIa reveal interdomain flexibility.," *Biochemistry*, vol. 49, pp. 9739–9745, Nov. 2010.

[4] J. H. Morrissey, V. Pureza, R. L. Davis-Harrison, S. G. Sligar, C. M. Rienstra, A. Z. Kijac, Y. Z. Ohkubo, E. Tajkhorshid, and E. Tajkhorshid, "Protein-membrane interactions: blood clotting on nanoscale bilayers," *Journal of Thrombosis and Haemostasis*, vol. 7, pp. 169–172, July 2009.

[5] S. A. Smith and J. H. Morrissey, "Rapid and efficient incorporation of tissue factor into liposomes," *Journal of Thrombosis and Haemostasis*, vol. 2, no. 7, pp. 1155–1162, 2004.

[6] J. H. Morrissey, V. Pureza, V. Pureza, R. L. Davis-Harrison, R. L. Davis-Harrison, S. G. Sligar, Y. Z. Ohkubo, Y. Z. Ohkubo, E. Tajkhorshid, and E. Tajkhorshid, "Blood clotting reactions on nanoscale phospholipid bilayers," *Thrombosis Research*, vol. 122, pp. S23–S26, 2008.

[7] T. H. Bayburt, Y. V. Grinkova, and S. G. Sligar, "Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins," *Nano Letters*, vol. 2, pp. 853–856, Aug. 2002.

[8] S. S. Nielsen, M. Møller, and R. E. Gillilan, "High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell," *Journal Of Applied Crystallography*, vol. 45, pp. 213–223, Feb. 2012.

[9] D. Uhríková, D. Uhríková, N. Kucerka, J. Teixeira, J. Teixeira, V. Gordeliy, V. Gordeliy, P. Balgavý, and P. Balgavý, "Structural changes in dipalmitoylphosphatidylcholine bilayer promoted by Ca2+ ions: a small-angle neutron scattering study.," *Chemistry And Physics Of Lipids*, vol. 155, pp. 80–89, Oct. 2008.

[10] J. M. Boettcher, J. M. Boettcher, R. L. Davis-Harrison, M. C. Clay, M. C. Clay, A. J. Nieuwkoop, A. J. Nieuwkoop, Y. Z. Ohkubo, E. Tajkhorshid, J. H. Morrissey, and C. M. Rienstra, "Atomic View of Calcium-Induced Clustering of Phosphatidylserine in Mixed Lipid Bilayers," *Biochemistry*, vol. 50, pp. 2264–2273, Mar. 2011.