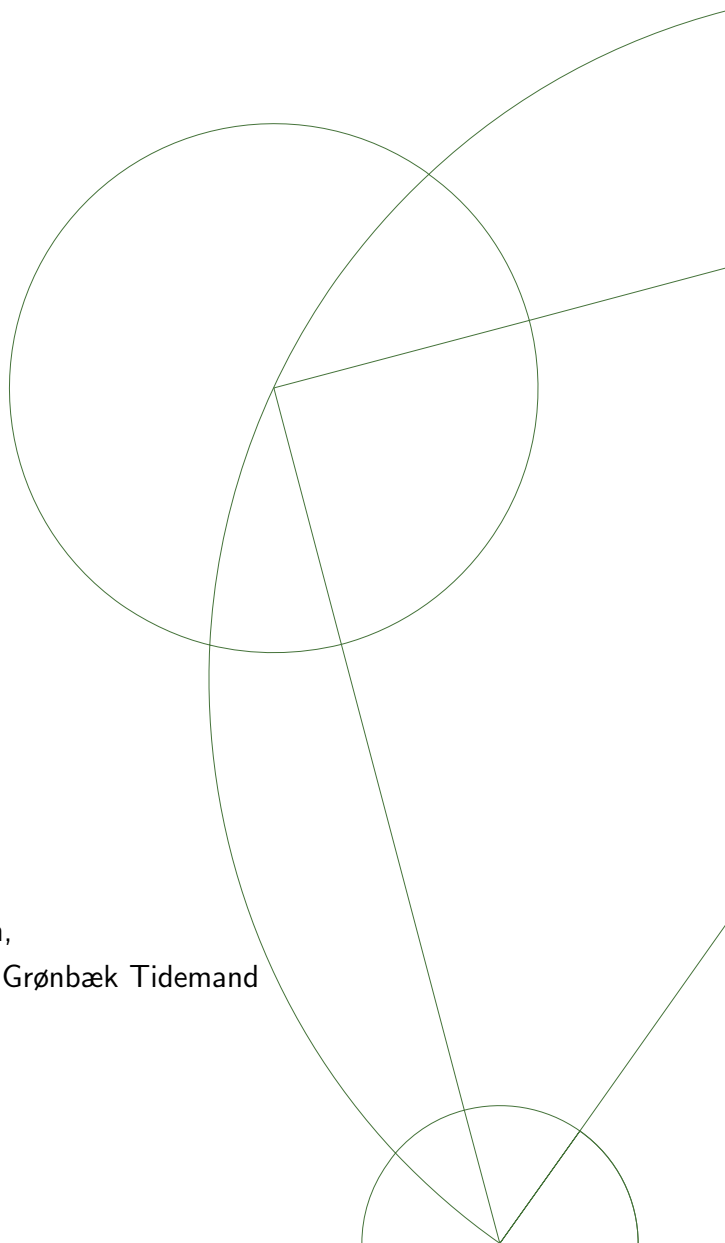
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Investigating Nanodisc Self-assembly

- Effects of the Length of the Membrane Scaffold Protein, Circularization, and Detergent Type and Concentration

Master's thesis

Sara Blemmer



Main supervisor: Lise Arleth and Martin Cramer Pedersen,
Co-supervisors: Nicolai Tidemand Johansen and Frederik Grønbæk Tidemand
Niels Bohr Institute

March 31, 2021

Abstract

Membrane proteins are essential to the body, as they are involved in many vital physiological processes, but they are difficult to study due to their amphiphilic nature. A widely used system for studying membrane proteins in solution are nanodiscs. Nanodiscs consist of a small patch of lipid bilayer enclosed by two membrane scaffold proteins (MSPs), and the reconstitution process of nanodiscs is mediated by detergents. Recent improvements to the nanodisc structure includes the introduction of circularized nanodiscs, where the termini of the MSP are covalently linked. When membrane proteins are to be integrated in nanodiscs, the storage detergent of the membrane protein must also be considered in the reconstitution process. The choice of detergent has been shown to greatly affect the shape and structure of the formed nanodiscs. This project investigated the effect of combining cholate, which is the detergent traditionally used when forming nanodiscs, with DDM, which is often used as storage detergent for membrane proteins. Through size exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS), the structures of the formed nanodiscs were investigated for various lengths of the MSP, different concentrations of DDM present during reconstitution, and the effect of circularization. Geometrical models of the nanodisc structures were successfully fitted to the SAXS data, indicating the formation of nanodiscs in the samples. It was determined that nanodiscs with circularization and longer MSPs are less prone to form larger particles. Furthermore, the use of both cholate and DDM in the reconstitution process also increases the monodispersity of the formed nanodiscs. However, DDM alone yields highly elliptical and lipid-poor nanodiscs. SEC experiments where the membrane protein proteorhodopsin (PR) was integrated in nanodiscs indicated that adding DDM during the reconstitution process increased both monodispersity and the amount of PR integrated in the discs. These results suggest that the addition of DDM to the reconstitution process has the potential for integrating other and more unstable membrane proteins in nanodiscs for structural examinations.

Acknowledgements

- Thank you to my brilliant supervisors for excellent supervision, even during these strange pandemic times.
- Thank you to Nicolai T. Johansen and Frederik G. Tidemand for protein expression and purification and help and guidance in the lab.
- Thank you to the P12 beamline at DESY and the B21 beamline at Diamond Light Source for providing the facilities and granting the beamtimes for performing the small-angle X-ray scattering experiments.
- Thank you to Abigail Barclay and Pernille L. Rasmussen for proofreading the thesis.
- Thank you to my family and friends for encouragement and support when I needed it the most.

List of abbreviations

Symbol	Meaning
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, a phospholipid
MSP	Membrane scaffold protein
ND	Nanodisc
lsND	Linear solubility enhanced nanodisc
csND	circularized solubility enhanced nanodisc
lsMSP1ΔH5	linear solubility enhanced MSP with truncation of helix 5 from the original MSP
csMSP1ΔH5	circularized solubility enhanced MSP with truncation of helix 5 from the original MSP
lsMSP1D1	linear solubility enhanced version of the original MSP
csMSP1D1	circularized solubility enhanced version of the original MSP
lsMSP1D1E3	linear solubility enhanced MSP with repetitions of helix 4, 5, and 6 from the original MSP
csMSP1D1E3	circularized solubility enhanced MSP with repetitions of helix 4, 5, and 6 from the original MSP
CMC	Critical micelle concentration, concentration limit above which a given detergent or lipid will form micelles
EMG	Exponentially modified Gaussian function
IFT	Indirect Fourier transformation
PR	Proteorhodopsin, a membrane protein that absorbs UV-light in the green range (530 nm)
SEC	Size exclusion chromatography
SAXS	Small-Angle X-ray scattering

Contents

1 Introduction

1.1 Motivation

Proteins are large molecules, that are vital to all living organisms. They carry out many different biological functions, including transportation of molecules from one site to another, catalysing metabolic reactions, and communication between cells, to name a few. Some proteins are situated in, or interact with, the cell membrane; these are known as membrane proteins. Almost a third of the human genome encodes membrane proteins [37], and as they are often targeted by pharmaceutical drugs, they are of great scientific interest. Nonetheless, out of the approximately 154,000 known protein structures [51] less than 1,300 [45] are membrane proteins. The main reason is that membrane proteins are difficult to study, as they will generally denature when removed from the membrane. Different approaches have been developed to account for this, including the development of nanodiscs [7]. Nanodiscs consist of a patch of lipid bilayer enclosed by two stabilizing membrane scaffold proteins (MSPs), and they can provide an environment very similar to that of the cell membrane. Recently, the introduction of covalently linked termini of the MSPs, to form so-called circularized nanodiscs, improved the stability and homogeneity compared to the nanodiscs formed with conventional linear MSPs [47]. Afterwards, solubility-enhanced nanodiscs were developed by adding negative charges to the surface of the MSPs to increase the solubility and prevent aggregation of nanodiscs [33]. However, the details of the structure and self-assembly process of these new-generation nanodiscs are widely unexplored.

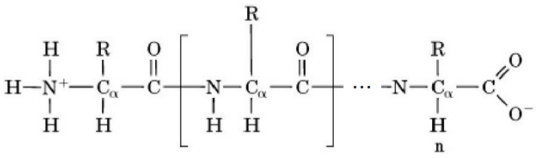
The nanodisc reconstitution process is mediated by detergents, with the ionic detergent cholate being the conventional choice. As the additional charges of the solubility-enhanced nanodiscs might clash with the negative charges of cholate, it could potentially affect the reconstitution ability, so it is relevant to explore the use of nonionic detergents. Furthermore, the integration of membrane proteins in nanodiscs will include a membrane protein storage detergent in the reconstitution process. A commonly used storage detergent is the nonionic detergent DDM. This project explores the effect of combining the two detergents, cholate and DDM, which would mimic a typical experimental situation.

This thesis aims to systematically investigate and compare how the length of the MSP, the concentration of DDM, and circularization affects the structure of the formed nanodiscs, and verify that the nanodiscs are compatible with membrane protein insertion. The self-assembly of the nanodiscs will be studied with biophysical and structural methods and compared to the conventional MSPs. The methods include size exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS), as these are excellent methods studying nanoscale structures, such as nanodiscs.

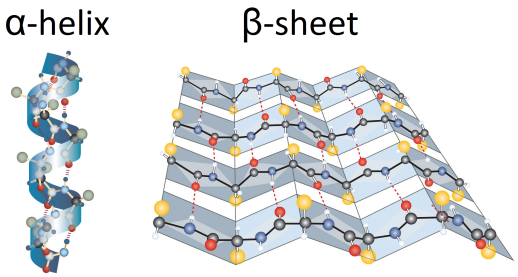
1.2 Proteins and their structure

All proteins consist of the organic compounds amino acids. The structure of amino acids is a carbon atom bound to a hydrogen atom, a carboxylic acid (-COOH), an amine (-NH2), and a unique side chain (-R) that distinguishes one amino acid from another, as seen in figure 1.1a. The side chains can be hydrophobic (non-polar) or hydrophilic (polar or charged) and they greatly affect the structure of the protein. Naturally there exist 20 different amino acids and they form covalent bonds between each other to create long chains known as polypeptides. Polypeptides have a backbone of repeated carbon and nitrogen atoms and when they fold into three-dimensional structures, they are known as proteins.

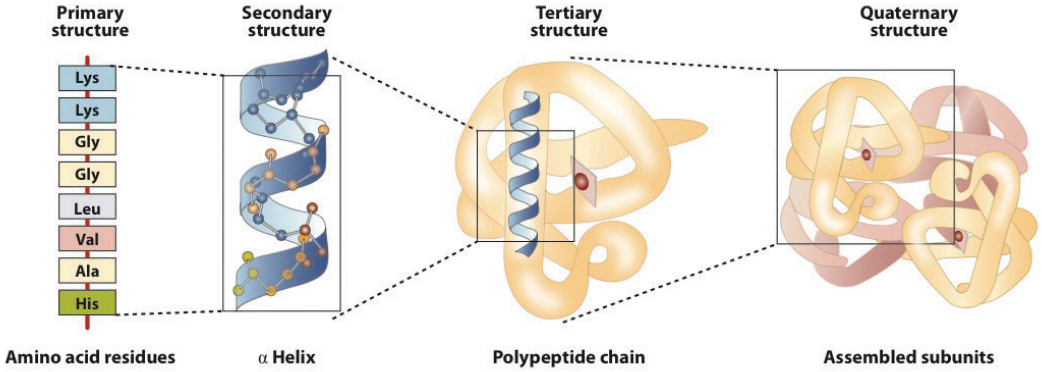
The structure of proteins is described at four different levels. The primary structure is the order of the amino acid sequence. The secondary structure is the local formation of repeated structures stabilized by the creation of hydrogen bonds. The most common secondary structures are α-helices and β-sheets



(a) *The chemical structure of a polypeptide, i.e. a chain of amino acids. The square brackets display the bonds between the different amino acids. Reproduced from [55].*



(b) *Examples of different kinds of secondary protein structure; an α-helix and a β-sheet. Reproduced from [1].*



(c) *The different levels of protein structure. From [1].*

Figure 1.1

as seen in figure 1.1b. The spiral conformation of the α-helix occurs when an N-H group in the backbone of one amino acid binds to the backbone C=O group of another amino acid located four residues before it in the amino acid sequence. β-sheets form when sections of amino acids that could be far away in the sequence are brought into close proximity to construct an accordion pleated-like sheet.

The tertiary structure is the three-dimensional structure of a folded protein. The protein is stabilized through different kinds of chemical bonds, such as disulphide bridges and hydrogen bonds, and molecular interactions like ionic interactions. The fourth and last level of the protein structure description is the quarternary structure. As many proteins consist of multiple sub-units of similar or different proteins, the quarternary structure describes how the sub-unit proteins assemble to form the fully functional protein [4]. The levels of protein structure can be seen in figure 1.1c

Protein folding is the physical process where the polypeptide assembles into the three-dimensional structure of a protein. Many polypeptides will instantly fold into the active spatial conformation known as the native state, as this usually is the most energetically favourable state. An unfolded polypeptide chain has a tremendous degree of rotational freedom so it might seem contradictory that it should be an advantage to fold, as the folding process constrains the mobility of the polypeptide. However, there are different forces driving protein folding, the most important one being the hydrophobic effect. As non-polar substances cannot interact with water, they disrupt the hydrogen bonds between the water molecules. As a consequence, the water molecules surrounding the non-polar substance will lose rotational freedom and the entropy will decrease. To account for this the non-polar molecules tend to aggregate in

aqueous environments to minimize the surface area and thus the number of disrupted hydrogen bonds [31].

As a result of the hydrophobic effect, the non-polar amino acids tend to reside in the interior of proteins, while the polar and charged amino acids can be found on the surface where they can interact with the surrounding water molecules. When the protein is folded it is further stabilized by the formation of chemical bonds, such as hydrogen bonds and disulphide bridges, and Van der Waals forces.

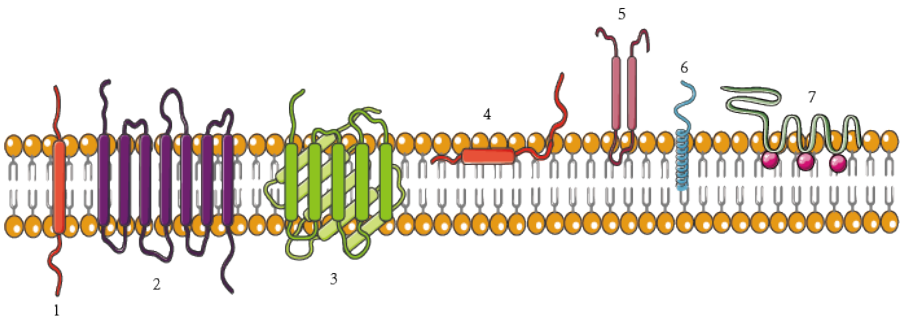
The stability and function of biological proteins is only maintained at the right conditions, depending on external parameters such as temperature, pH, acidity, and salt concentration. If these conditions are not satisfied the proteins might unravel and denature, thus losing their function.

1.3 Membrane proteins

Membrane proteins are a group of proteins on or embedded within the cell membrane as seen in figure 1.2a. The cell membrane is an approximately 5 nm thick biological membrane of fluidic structure enclosing the cell, thus separating the intracellular and extracellular environment [3]. Amongst other molecules, the cell membrane consists different of lipids, the most abundant type being phospholipids. Phospholipids are a group of amphiphilic lipids with a hydrophilic head group and two hydrophobic carbon tails.

Due to the hydrophobic effect, the phospholipids will spontaneously arrange themselves in a bilayer with the polar headgroups facing outwards towards the water and the non-polar tails protected inside in a non-aqueous environment. The cell membrane is a semipermeable barrier, as some molecules can easily cross the membrane by diffusion, while others require active transportation.

Membrane proteins are vital as they carry out various tasks including transportation of ions and molecules across the cell membrane, communicating with the surrounding cells, or reacting to stimuli. Transmembrane proteins are proteins extending through the cell membrane with parts on both sides, while monotopic membrane proteins are anchored to just one side of the membrane. Both types of membrane proteins can be seen in figure 1.2a. Membrane proteins are also amphiphilic molecules with the hydrophobic parts hidden in the midst of the membrane, where they can interact with the lipid tails, and the hydrophilic regions exposed to the surrounding aqueous environments on either side of the membrane [3].



(a) *Different kinds of membrane proteins in a membrane. The phospholipids of the cell membrane are depicted with yellow headgroups and grey tails. 1-3 are transmembrane proteins, while 4-7 are monotopic membrane proteins, i.e. they interact with only one side of the membrane. Figure from [62].*

(b) *The heptahelical structure of proteorhodopsin. Figure from [2].*

Figure 1.2

1.3.1 Proteorhodopsin (PR)

The membrane protein proteorhodopsin (PR), which was studied in this project, is from a family of transmembrane proton transport proteins found in various bacterial organisms. The function of PR is driven by light, so depending on the species, PR absorbs light at specific wavelengths [24]. The PR used in this project was expressed and purified from *E. Coli* and absorbs in the green spectra with an absorption peak at 525 nm[6]. The overall structure of PR is heptahelical, i.e. seven α-helices extending through the membrane as seen in figure 1.2b. In the native cell membrane, PR may exist as a monomer, i.e. a single PR, or form larger oligomeric structures, that is the assembly of multiple PR-proteins, usually five or six, into one ring-shaped unit [36].

1.4 Studying membrane proteins

Because of the amphiphilic nature of membrane proteins, they are notoriously difficult to study when in the native state. Membrane proteins are not water-soluble, as they denature and aggregate uncontrollably when removed from the membrane, in order to minimize the amount of exposed hydrophobic parts. Thus, it is necessary to use a different strategy when examining the function and structure of native membrane proteins. Many of these methods usually involve inserting the membrane protein in a carrier particle of other amphiphilic molecules, thereby making the membrane protein water-soluble.

Amphiphilic molecules such as lipids can self-assemble, i.e. form an ordered structure spontaneously in aqueous environments, as a consequence of the hydrophobic effect. The shape and size of lipid aggregates depend on various conditions of the solution, such as temperature, pH, and concentration, but also how the lipids pack, which depends on the phase, and the shape and geometry of the given lipid.

The lipid packing parameter P, given as $P = {}^{V}/_{A_{head} \cdot L_{lip}}$, where V is the volume of the lipid, A_{head} is the area of the lipid headgroup, and L_{lip} is the length of the lipid. Depending on the value of P, the aggregates will have different shapes [30]. The most common shapes of aggregates can be seen in figure 1.3. For $P < {}^{1}/_{3}$ the lipids will form micelles, that are small spherical aggregates with a hydrophobic core. Liposomes have a packing parameter $P \sim {}^{1}/_{2} - 1$ and a bilayer structure with an aqueous centre. For a packing parameter $P \sim 1$ the lipid shape is almost cylindrical, and these will form bilayers. Micelles and liposomes have both been used to study membrane proteins, but many methods require monodispersity which is difficult to achieve, as the sizes of liposomes and micelles can only be controlled to some degree [41].

In this project, the phosphatidylcholine POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) was used. POPC is a phospholipid naturally occurring in the cell membrane, and its two carbon tails are 18 and 16 carbon atoms long. The longer tail is unsaturated, so it has a double-bond between the ninth and tenth carbon atom giving the tail a slight kink. POPC has a packing parameter $P \sim 1$, and forms bilayers, and has a melting temperature of -2 °C [54], thus it is always in the liquid crystalline phase at physiologically relevant temperatures.

1.4.1 Detergents

Detergents are another type of amphiphilic molecules that is often used for structural examinations of membrane proteins. The structure of detergents may vary and the hydrophilic region can either be charged or uncharged. Detergents are more soluble in water than lipids, as at low concentrations detergents are monomeric, i.e. single molecules. At higher concentrations detergents aggregate to form micelles. The concentration limit where micelles can form is known as the critical micelle concentration (CMC). At concentrations exceeding the CMC, the monomeric concentration is constant while the concentration of

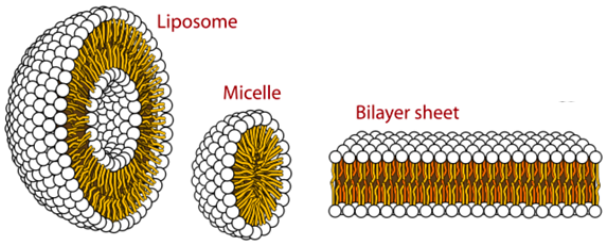


Figure 1.3: *Cross sections of the three most common kinds of three dimensional structures of phospholipids in aqueous environments. Modified from [44].*

micelles increases. The CMC as well as the shape and size of the micelle depends on the type of detergent through the above-mentioned packing parameter, but also external parameters such as the temperature, pH, and salt concentration. Membrane proteins can be protected and stabilized in detergent micelles. As detergent micelles are inexpensive to manufacture and easy to handle in the lab, they are often used when studying membrane proteins [58]. However, as with liposomes and lipid micelles, the sizes of detergent micelles may vary, thus disrupting many structural methods that depends on monodispersity. Moreover, it poses a problem that the environment in a detergent micelle bears little resemblance to the native environment in a cell membrane and stronger detergents can in some cases denature or destabilize membrane proteins, thus changing the structure and functionality of the membrane proteins.

1.5 Nanodiscs

Nanodiscs are water-soluble discoidal structures used for studying membrane proteins. The general structure of nanodiscs is a phospholipid bilayer surrounded and stabilized by two amphiphilic belts, as seen if figure 1.4. Many different kinds of belts have been developed, such as synthetic polymers or peptides, but the original nanodisc was stabilized by so-called membrane scaffolding proteins (MSPs) derived from human ApoA1 [7]. MSPs have an α-helical structure and wrap around the disc with the hydrophobic part protecting the hydrophobic edges of the phospholipid tails in the membrane, and the hydrophilic part facing outwards.

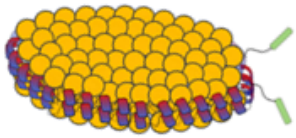


Figure 1.4: *Overall structure of a nanodisc; a bilayer of phospholipids (yellow circles) surrounded by two stacked α-helical MSPs, and two His-tags (green). From [61].*

The first MSP was developed by Bayburt et al. in 2002 [7] and was based on using a truncated, i.e. shortened, version of apolipoprotein A1 (Apo1). The MSP was named MSP1D1 and it has 10 α-helices. Since then, many different modifications of the Apo1 have been made. By truncation or extension of the different helices from the original MSP, the resulting MSPs are of different lengths. As the size of the nanodisc is dependent on the length of the MSPs it is possible to control the size of the nanodisc. The smallest nanodiscs of approximately 8 nm in diameter are small enough to use in solution NMR [10], and recently nanodiscs with diameters up to 50 nm were reported, enabling examination of larger membrane proteins [47].

There are several advantages of using nanodiscs for studying membrane proteins, and the use of

nanodiscs in structural studies have increased. The environment in the nanodiscs resembles that in a cell membrane to a great extent, so it is possible to insert and stabilize membrane proteins in their native state. Furthermore, nanodiscs are reasonably easy to handle as they can self-assemble and it is possible to control the sizes of the nanodiscs to prevent polydispersity.

Recently Nasr et al. managed to upgrade nanodiscs by developing a method for fusing the ends of the MSP, thus circularizing the MSP [47]. The fusing is a covalent link catalyzed by the enzyme sortase A (*Srt*), introduced before the nanodisc reconstitution process. Nasr and coworkers discovered that circularized nanodiscs have a more uniform size distribution and a greater degree of stability at physiologically relevant temperatures than the original linear nanodiscs. After the circularization Johansen et al. [33] introduced mutations of the MSP amino acid sequence to add additional negative charges. The addition of more negative charges increases solubility and the repulsion between proteins thus making aggregation less likely [69].

For this project, six different MSPs were used. They were all solubility enhanced through the above-mentioned introduction of negative charges, while half were linear and the other half circularized versions: ls- and csMSP1D1 are solubility enhanced circularized and linear versions of the original MSP. Ls- and csMSP1ΔH5 are truncated versions where the fifth helix has been deleted. Ls- and csMSP1E3D1 are extended versions, where helix four, five, and six have been repeated. The linearized nanodiscs have slightly longer amino acid sequences as they have a His-tag, which is a short sequence of Histidine amino acids. The His-tag does not contribute to the function of the belt and during the circularization process, it is cleaved off. An illustration of the general structure and size can be seen in figure 1.5.

1.5.1 Nanodisc reconstitution

The reconstitution process of nanodiscs is illustrated in figure 1.6. In a buffer solution MSPs are mixed with phospholipids and detergent to form preaggregates, i.e. aggregates containing a mix of MSPs, lipids and detergent. Afterwards, the detergent is removed by detergent absorbing resin beads [57] and nanodiscs will self-assemble. Afterwards, the sample is purified by size exclusion chromatography, which separates the sample molecules by their sizes.

The concentrations and ratios of lipids, detergent, and MSP are crucial to the self-assembly. It was shown by Skar-Gislinge et al. that the number of lipids in the fully formed nanodiscs is dependent on the choice of detergent which, again affects the number and concentration of lipids in the preaggregate [61].

As the MSP determines the maximum circumference of the formed nanodisc, each nanodisc has a limit to the amount of contained lipids, and cannot be loaded beyond that. However, as there is only one configuration for a fully loaded disc, but many configurations with fewer lipids, the fully loaded disc is entropically unlikely to be favoured. Assuming the number of lipids per nanodiscs follows a normal distribution, the average nanodisc will not be maximally packed with lipids [61]. If the ratio of mixed lipids and MSPs is not right during the reconstitution, there will be an excess of either lipids or MSPs appearing as aggregates of varying size when removing the detergent [8]. This can affect the structural studies, and the MSP:lipid ratio is important to consider.

As the reconstitution process of nanodiscs is driven by detergents, the choice of detergent has been shown to have a major effect on the structure of the formed nanodiscs [61].

Detergents have higher CMCs compared to lipids, and when mixing detergent with lipids they form mixed micelles or mixed vesicles. Whether micelles or vesicles are formed depends on the temperature, salt concentration, type of detergent, and concentration of the respective lipids and detergent. These parameters also determines the size of the micelles/vesicles.

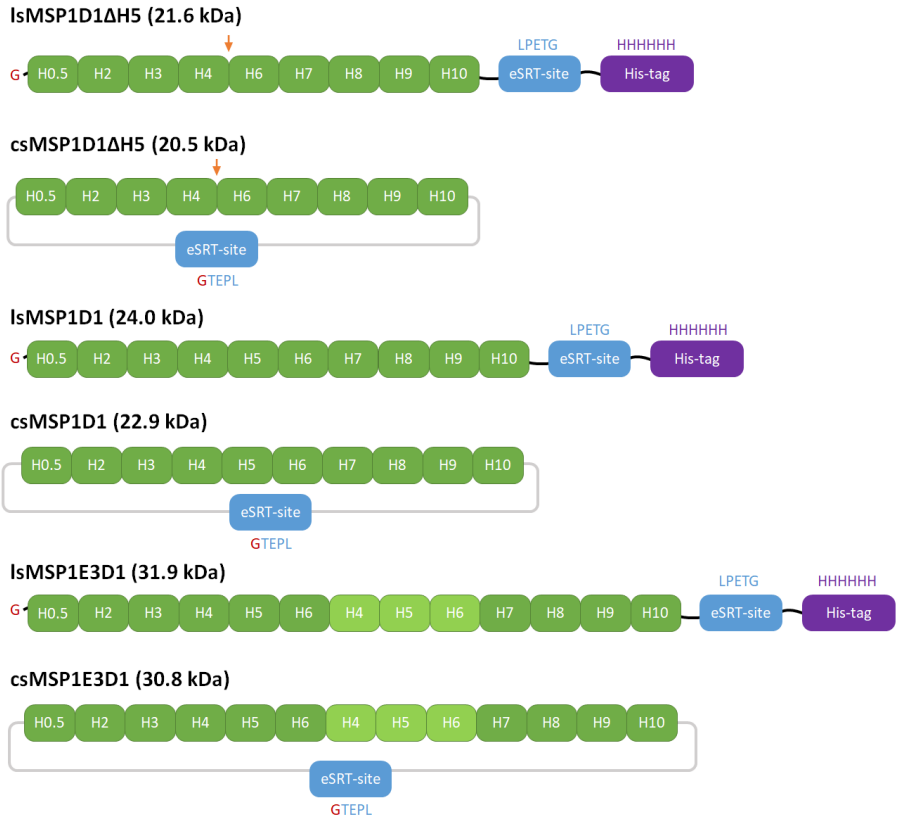


Figure 1.5: *Illustration of the different MSPs used in this project indicating their sizes. The green "boxes" are the helices. The blue eSRT site is the site for binding of the sortase A enzyme for catalyzing the circularization process, in which the purple His-tag is cleaved of. The letters over/under the sites indicate the amino acid sequences. For ls- and csMSP1D1ΔH5 the fifth helix has been deleted as indicated by the orange arrows, while ls- and csMSP1E3D1 have helix 4,5, and 6 repeated.*

The original nanodisc was reconstituted using sodium cholate [7], which is an ionic detergent with a CMC of 10 mM in pure water [20].

Hildebrand et al. examined the solubility of POPC and sodium cholate mixtures and determined the phase transition between vesicles and micelles [28]. Figure 1.7 depicts the behaviour for POPC and sodium cholate with data from table 4 in [28]. At low concentrations of sodium cholate (blue region) POPC and sodium cholate form mixed vesicles as well as detergent monomers. At high sodium cholate concentrations (orange region) POPC and sodium cholate form mixed micelles and detergent monomers. In the intermediate region (green) mixed micelles and mixed vesicles co-exist with detergent monomers.

Nanodiscs mainly form from the mixed micelle phase, thus it is important to use the right concentrations and ratios of lipid and cholate for the reconstitution process. Experiments with reconstitution in the vesicle phase showed that the resulting aggregates were heterogeneous and slightly larger, but contained a smaller number of lipids than the nanodiscs reconstituted from mixed detergent-lipid micelles [7].

A popular detergent for solubilizing membrane proteins is n-dodecyl-β-D-maltopyranoside (DDM). DDM is a mild non-ionic detergent with a CMC of 0.15 mM [46]. DDM has a smaller effect on the stability of the membrane proteins compared to other detergents [19], and it has proven successful in many structural studies of membrane proteins [50]. However, the environment in DDM micelles is still far from the native environment in the cell membrane and the function and structure of some membrane proteins can still be affected.

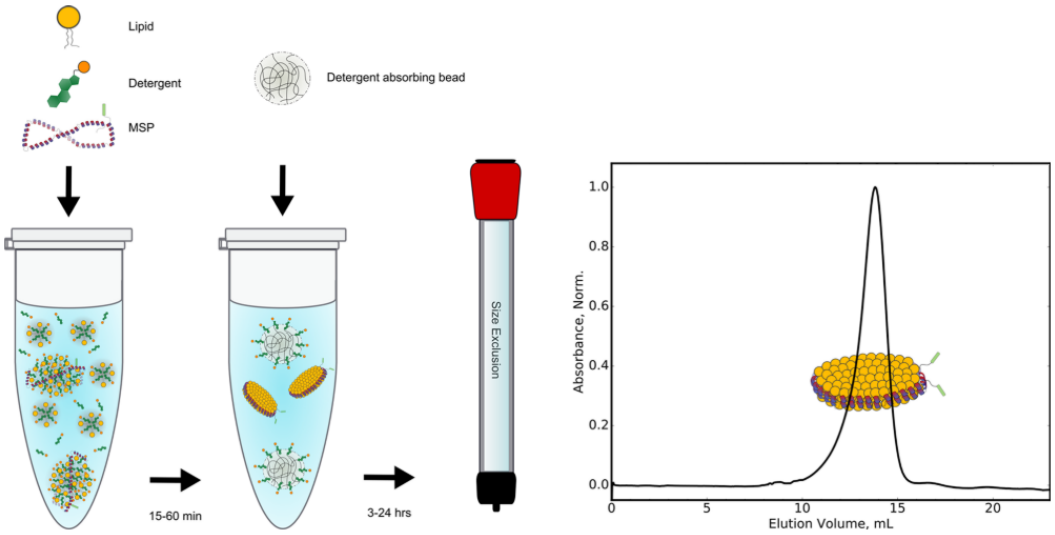


Figure 1.6: *Nanodisc reconstitution process. Lipids, detergent, and MSP are mixed in buffer solution. Detergent removing resin beads are added and the nanodiscs self-assemble. Afterwards, the sample is purified by size exclusion chromatography, which separates the sample molecules by their sizes. Modified from [61].*

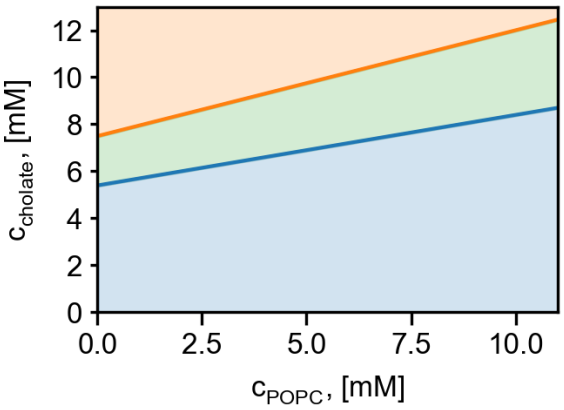


Figure 1.7: *The different phases of mixed POPC and sodium cholate depending on their respective concentrations, c_{POPC} and $c_{cholate}$. At low sodium cholate concentrations (blue region) POPC and sodium cholate form mixed vesicles as well as detergent monomers. At high cholate concentrations (orange region) POPC and sodium cholate form mixed micelles and detergent monomers. In the intermediate region (green) mixed micelles and mixed vesicles co-exist with detergent monomers. Solubilization data from [28].*

2 small-angle X-ray scattering

In small-angle X-ray scattering (SAXS), X-rays are utilized to obtain structural information of macromolecules in solutions at length scales of approximately 1-1000 Å. SAXS is a low-resolution technique as it cannot be used to examine structures at atomic length scales, but rather yields information about the shape, size, polydispersity, and composition of the given macromolecules in a sample. In crystallographic methods, the sample is required to be crystalline, while the advantage of SAXS is the possibility to study solubilized samples. SAXS is used for many different applications, including structural studies of polymers and bio-molecules [65], [13].

The following sections on scattering theory are mainly based on "Elements of Modern X-ray Physics: Second Edition" by J. Als-Nielsen and D. McMorrow [5], "Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules" by D. Svergun et al [66], and "Neutron Scattering: Theory, Instrumentation and Simulation" by K. Lefmann [42].

2.1 Experimental X-rays and X-ray production

X-rays can be produced at in-house laboratories using X-ray tubes, which can be sufficient for some structural scattering experiments. However, many biological samples scatter only weakly, so scattering experiments involving biological samples are often performed at large-scale synchrotron facilities, which can generate X-rays with higher flux and collimation, i.e. larger degree of parallelization.

The specifics of large-scale facilities may vary, so the following description is based on the B21 beamline at Diamond Light Source in Didcot, UK, as this was the beamline used for the main SAXS experiments in this project [29]. Figure 2.1 shows an illustration of the Diamond Light Source synchrotron facility.

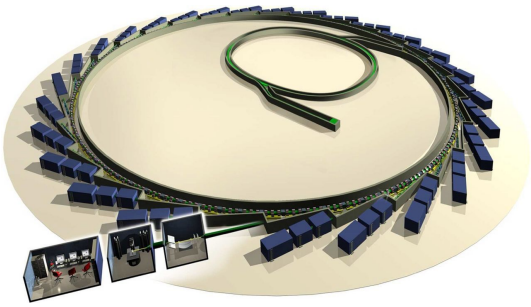


Figure 2.1: *Illustration of the Diamond Light Source synchrotron facility. The electrons are produced and accelerated in the linear accelerator in the centre. The electrons are further accelerated in the booster synchrotron before reaching the storage ring. The beamlines are placed tangentially to the orbit of the electrons in the storage ring. From [16].*

At the synchrotron the electrons are produced by an electron gun, by "evaporating" electrons from a high-voltage cathode using heat in the process thermionic emission, i.e. when the electrons receive enough energy they can escape the cathode [29]. A linear synchrotron and a booster synchrotron accelerate the electron beam to relativistic velocities before the electrons are transferred to a storage ring. In the storage ring, the electrons are kept in circular motion as the beam is bent by the magnetic fields of strong magnets. The electrons emit energy in the form of photons as they change direction in the storage ring. When the electron velocity is large enough, the energy will be emitted as X-rays. The radiation is emitted tangentially to the orbit of motion, so the laboratories for scattering experiments, known as beamlines, are placed along these tangential lines.

SAXS experiments are highly dependent on the X-ray beam being sufficiently monochromatic and

collimated. This is achieved through different optic devices, such as monochromators and slits. The slits limit the divergence of the beam and ensure proper focusing, while monochromators select radiation of a specific wavelength bandwidth. The "price" of focusing and collimating the beam is the loss of photons with every aperture and the resulting X-ray beam will accordingly have a lower intensity.

After the monochromatic X-ray beam has been collimated and focused, it is guided to the sample. Fig. 2.2 illustrates the setup for a SAXS experiment. Most of the photons will travel directly through the sample without interaction with any atoms and hit a beamstop placed in front of the detector. As the detector can be damaged by this direct beam, the beamstop is placed in front to protect the detector. Some of the photons will scatter onto the detector and create a two-dimensional detection pattern. The detector is pixelated and measures the intensity, i.e. the number of photons per pixel per unit time.

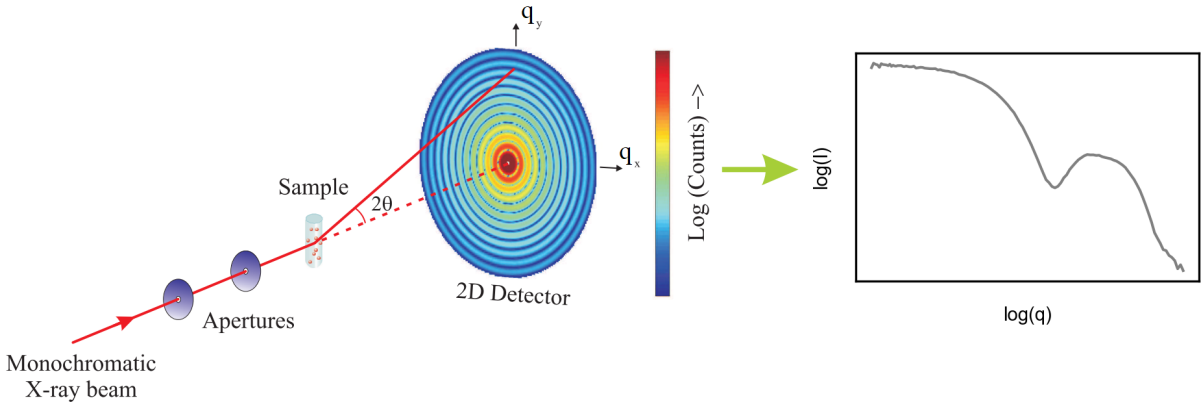


Figure 2.2: *Illustration of a SAXS experiment. The X-ray beam is collimated and focused by different apertures before reaching the sample. Most X-ray travels directly through the sample without interfering, while a small fraction is scattered. The resulting isotropic scattering pattern is azimuthally averaged to obtain a one-dimensional scattering curve, such as the one on the right. From [5].*

Biological samples for SAXS experiments can be kept in aqueous solutions, so the molecules have free mobility and are oriented randomly. In these cases, the resulting scattering pattern will be isotropic, and by azimuthally averaging the pattern, the two-dimensional signal becomes one-dimensional.

The buffer solution in which the sample is dissolved will also scatter and is responsible for a significant part of the measured scattering signal. Furthermore, there will be scattering contributions from the capillary containing the sample. All these accounts for the background scattering. By measuring the signal from buffer and background without sample, it can be subtracted from the signal from the sample, buffer, and background to obtain the signal from the sample only:

$$I_{\text{sample}} = I_{\text{sample+buffer+background}} - I_{\text{buffer+background}} \tag{2.1}$$

where I_{sample} is the scattered intensity from the sample, $I_{\text{sample+buffer+background}}$ is the scattered intensity from the sample, the buffer and the background, and $I_{\text{buffer+background}}$ is the scattered intensity from the buffer and the background. If the scattering signal from the sample is too small compared to the buffer, it can be difficult to distinguish the signal from the buffer background, thus making data processing more difficult.

2.2 Scattering theory

Because of the electromagnetic properties of X-rays, they can interact with electric charges and thus the electrons of atoms. The X-ray/atom interaction can occur in different ways. The most important interactions include absorption of the X-ray and release of an electron or scattering. Scattering events can be coherent or incoherent. The incoherent scattering contributes to the background signal as it spreads uniformly in all directions, and it will be treated as noise. The coherent scattering can be elastic or inelastic; for elastic scattering, the energy is conserved while inelastic scattering includes energy transfer. For SAXS experiments the coherent elastic scattering generates the signal used for structural determinations. SAXS experiments are performed with hard X-rays with wavelengths of order ~ 1 Å (10^{-10} m).

The simplest case of scattering is from a single atom. The initial X-ray is assumed to be a monochromatic plane wave with wave vector $\vec{k}$ of length $|k| = 2\pi/\lambda$, where λ is the wavelength. In the classical description of scattering, a force is applied to the electrons of the atom by the electric field of the incident X-ray. The electrons are thus accelerated and radiate a scattered spherical wave with wave vector $\vec{k'}$, into every direction. When only considering elastic scattering phenomena there will be no loss of energy so the magnitudes of the incident and the scattered wave vectors are identical $|k| = |k'|$.

A central part of scattering experiments is the scattering vector $\vec{q}$, defined as $\vec{q} = \vec{k} - \vec{k'}$. $\vec{q}$ describes the momentum transfer of the scattering event and as seen in figure 2.3 the magnitude of $\vec{q}$ can be calculated from trigonometry:

$$|q| = 2|k|\sin(\theta) = \frac{4\pi \sin(\theta)}{\lambda} \tag{2.2}$$

where 2θ is the scattered angle, i.e. the angle between the incident and the scattered wave vector.

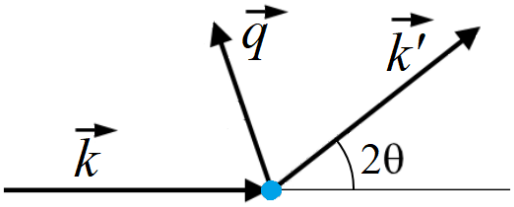


Figure 2.3: *Elastic scattering from an atom. The incoming X-ray has a wave vector $\vec{k}$ while the wave vector of the scattered wave is denoted $\vec{k'}$. The scattering vector is $\vec{q} = \vec{k} - \vec{k'}$ and θ is the scattering angle.*

In SAXS experiments the measured quantity is the intensity I_{meas}, which is the number of scattered photons per unit area per unit time. For a single scatterer the intensity is proportional to the incident X-ray flux Φ_0 and the scattering cross-section $d\sigma$, which is the probability of a photon scattering:

$$I_{meas} = d\sigma \cdot \Phi_0 \tag{2.3}$$

As the measured intensity depends on the incident X-ray flux and the size of the detector, which in turn depends on the specific instrument used, it is convenient to normalize the intensity with the incident flux and the specific solid angle. The normalization ensures that measurements can be compared regardless of the experimental setup. For real-life experiments, the efficiency of the detector should be taken into account, but when assuming all scattered waves are detected the normalized intensity of

scattered photons into a specific solid angle $d\Omega$ is defined as follows:

$$I_{norm} = \frac{I_{meas}}{\Phi_0 d\Omega} = \frac{d\sigma}{d\Omega} \tag{2.4}$$

The normalized intensity $d\sigma/d\Omega$ is also called the differential scattering cross-section and it is an important quantity in scattering experiments. It is assumed that the sample and detector are sufficiently far away from each other, so the incident and scattered X-rays may be considered plane waves. This is known as the far-field limit or the Frauenhofer limit. The differential scattering cross section for each atom is given as:

$$I_{norm} = \frac{d\sigma}{d\Omega} = b_j^2 \tag{2.5}$$

Where b_j is known as the scattering length of the j'th atom. The scattering length describes the amplitude of the scattered wave and for a single free electron, the scattering length will be the classical electron radius (or Thompson scattering length), $r_0 = 2.82 \cdot 10^{-5}$ Å. The scattering length of an atom is proportional to the number of electrons Z in the given atom and is given as $b_j = Z \cdot r_0$.

2.2.1 Scattering from multiple atoms

The scattering amplitude A from one atom can be described in terms of the scattering length of the atom and the phase ϕ_j:

$$A(\phi_j) = b_j \exp(i\phi_j) \tag{2.6}$$

where i is the imaginary unit. For scattering from multiple atoms, there will be interference between the scattered waves, resulting in an interference scattering pattern. As the amplitudes of all the interfering waves are added to get the amplitude of the total scattered wave, the total amplitude depends on the phases of the individually scattered waves. When the scattered waves are in phase the interference is constructive and the amplitude of the total scattered wave is amplified. When the scattered waves are out of phase the interference is destructive and the amplitude of the total scattered wave is diminished or entirely cancelled out.

Consider scattering from two atoms; one placed at the origin with scattering length b_1 and the other at $\vec{r}$ with scattering length b_2. The scattered wave from the first atom with the wave vector $\vec{k_1}$ will then need to travel the extra distance before interfering with the scattered wave from the second atom with wave vector $\vec{k_2}$ as seen in figure 2.4. Hence, there will be a difference in phase $\Delta\phi = (\vec{k_1} - \vec{k_2}) \cdot \vec{r} = \vec{q} \cdot \vec{r}$.

The total amplitude A of the scattered waves is a superposition of the amplitudes from the individual scattered waves, so using eq. 2.6 the amplitude is:

$$A(\vec{q}) = b_1 + b_2 \exp(i\vec{q} \cdot \vec{r}) \tag{2.7}$$

This concept can be extended to an ensemble of n atoms by assuming the amplitude of the total scattered wave is a linear sum, and by using the Born approximation [5]. The Born approximation applies when the scattering from an ensemble of scatterers is so weak that the incident field is only marginally perturbed after a scattering event. Thus it is safe to assume that it is the incident field and not the total field that affects the scatterers in the ensemble. The approximation is usually only valid for thin samples or samples where multiple scattering events happen so rarely that it can be assumed that

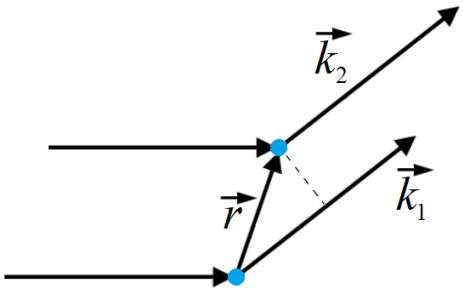


Figure 2.4: *Elastic scattering from two atoms placed a distance $\vec{r}$ apart. The scattering vector $\vec{k_1}$ from the first atom has to travel the extra distance before interfering with the scattering vector $\vec{k_2}$ from the second atom.*

each scatterer will only scatter once. As most biological samples mainly consist of lighter elements with small scattering lengths, these will comply with the first Born approximation. Thus, the total amplitude for scattering from n atoms is:

$$A(\vec{q}) = \sum_{j=1}^{n} b_j \exp(i\vec{q} \cdot \vec{r_j}) \tag{2.8}$$

where $\vec{r_j}$ is the position of the j'th atom with respect to an arbitrary origin. The intensity I of the scattered waves is the product of the total scattering amplitude A with its complex conjugate A*:

$$I(\vec{q}) = A(\vec{q})A^*(\vec{q}) = |A(\vec{q})|^2 = \left| \sum_{j=1}^{n} b_j \exp(i\vec{q} \cdot \vec{r_j}) \right|^2 \tag{2.9}$$

The sum in eq. 2.9 accounts for the scattering from all the atoms in the ensemble. If the ensemble consists of molecules the equation does not distinguish atoms in the same molecule from the atoms in another molecule. Thus, it can be convenient to arrange the atoms in groups and define the position of the j'th atom as $\vec{r_j'} = \vec{R_k} + \vec{r_j}$, where $\vec{R_k}$ is the center of a molecule with respect to an arbitrary origin and $\vec{r_j}$ is the position of the atom with respect to the center of the molecule $\vec{R_k}$ as seen in figure 2.5.

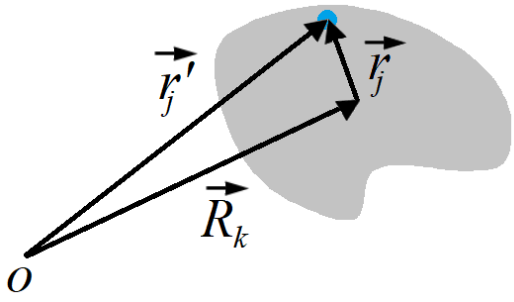


Figure 2.5: *The position of an atom (blue) within a molecule (grey). $\vec{R_k}$ is the center of the grey molecule with respect to an arbitrary origin, while $\vec{r_j}$ is the position of the blue atom with respect to the center of the molecule, so $\vec{r_j'} = \vec{r_j} + \vec{R_k}$.*

Eq. 2.9 can be split into two sums; k is summed over all the N molecules in the sample while j is

summed over the n atoms in the k'th molecule:

$$I(\vec{q}) = \left| \sum_{k=1}^{N} \sum_{j=1}^{n} b_j \exp\left(i\vec{q} \cdot (\vec{R_k} + \vec{r_j})\right) \right|^2 \tag{2.10}$$

$$= \left| \sum_{k=1}^{N} \exp(i\vec{q} \cdot \vec{R_k}) \right|^2 \left| \sum_{j=1}^{n} b_j \exp(i\vec{q} \cdot \vec{r_j}) \right|^2$$

$$= N\, S(\vec{q}) \left| \sum_{j=1}^{n} b_j \exp(i\vec{q} \cdot \vec{r_j}) \right|^2 \tag{2.11}$$

Where $S(\vec{q})$ is known as the structure factor. It describes how the molecules are organized in relation to each other, and thus the interparticle interactions such as repulsion or attraction. For dilute samples the structure factor approaches unity, $S(\vec{q}) \simeq 1$.

2.2.2 Scattering from samples in solution

Usually it is more convenient to use the scattering length density instead of the scattering length. The scattering length density is the sum of the scattering length contributions b_j from the n atoms per unit volume V':

$$\rho = \sum_{j}^{n} \frac{b_j}{V'} \tag{2.12}$$

Using the definition in eq. 2.12, and assuming the sample is dilute, the scattered intensity can be expressed as an integral over the volume of a molecule:

$$I_{molecule}(\vec{q}) = \left| \int_V \rho(\vec{r}) \exp(i\vec{q} \cdot \vec{r})\, d\vec{r} \right|^2 \tag{2.13}$$

where the scattering length density may vary throughout the molecule and thus depends on the position $\rho(\vec{r})$ in the molecule. For scattering experiments with samples in solution, the scattering from the solvent atoms must also be considered. The excess scattering length density $\Delta\rho(\vec{r})$ is the difference in scattering length density of the sample molecules $\rho_s(\vec{r})$ relative to the solvent ρ_0. Assuming the scattering length density of the solvent is constant, the excess scattering length density is given as $\Delta\rho(\vec{r}) = \rho_s(\vec{r}) - \rho_0$. When the sample is monodisperse, i.e. all the molecules are identical, the integral in eq. 2.13 can be extended to cover the entire sample volume of N sample molecules. As the molecules in solution are randomly oriented and distributed, the measured intensity is isotropic and will represent an orientational average. The intensity from the entire sample is then given as follows, where $\langle\rangle_\Omega$ denotes the orientational average:

$$I(q) = N \left\langle \left| \int_V \Delta\rho(\vec{r}) \exp(i\vec{q} \cdot \vec{r})\, d\vec{r} \right|^2 \right\rangle_\Omega \tag{2.14}$$

Eq. 2.14 is the Fourier transform of the excess scattering length density $\Delta\rho(\vec{r})$, so the measured intensity distribution is in reciprocal space. In the reciprocal space, large features are measured at small q-values and hence small-angles, while small features are measured at large q-values.

Usually eq. 2.14 is rewritten:

$$I(q) = c\Delta\rho^2 V^2 P(q) \tag{2.15}$$

where $c = N/V_{sam}$ is the number concentration of the molecules in the sample volume V_{sam} and $P(q)$ is the form factor:

$$P(q) = |F(q)|^2 = \left\langle \left| \frac{1}{V} \int_V \exp(i\vec{q}\cdot\vec{r}) d\vec{r} \right|^2 \right\rangle_\Omega \tag{2.16}$$

$F(q)$ is known as the form factor amplitude. The form factor describes the shape and size of the molecules in the sample and it can be calculated analytically for simple structures. For more complex structures the integral can be evaluated numerically.

2.2.3 Pair-distance distribution

The intensity of the scattered waves from the sample also depends on the correlation between the atoms in the sample.

The scattering from two atoms is considered; one at position $\vec{r_1}$ and the other at position $\vec{r_2}$. From eq. 2.8 and eq. 2.12 the intensity of the pair of scatterers is given as:

$$I(q) = \langle A(\vec{q})A^*(\vec{q})\rangle_\Omega = \left\langle \int_V \int_V \Delta\rho(\vec{r_1})\Delta\rho(\vec{r_2}) \exp(i\vec{q}(\vec{r_1} - \vec{r_2}))\, d\vec{r_1}\, d\vec{r_2} \right\rangle_\Omega \tag{2.17}$$

where $\Delta\rho(\vec{r_1})$ and $\Delta\rho(\vec{r_2})$ are the respective scattering length densities of the atoms. By changing variables to $\vec{r} = \vec{r_1} - \vec{r_2}$, eq. 2.17 can be rewritten:

$$I(q) = \left\langle \int_V \int_V \Delta\rho(\vec{r_1})\Delta\rho(\vec{r_1} - \vec{r}) \exp(i\vec{q}\cdot\vec{r})\, d\vec{r_1}\, d\vec{r} \right\rangle_\Omega \tag{2.18}$$

$$= \left\langle \int_V \gamma(\vec{r}) \exp(i\vec{q}\cdot\vec{r})\, d\vec{r} \right\rangle_\Omega \tag{2.19}$$

Where $\gamma(\vec{r})$ is the autocorrelation function [14]:

$$\gamma(\vec{r}) = \int_V \Delta\rho(\vec{r_1})\Delta\rho(\vec{r_1} - \vec{r})\, d\vec{r_1} \tag{2.20}$$

The orientational average of the complex phase depends on the magnitude of $\vec{q}$ and $\vec{r}$, but not their direction [5]: $\langle \exp(i\vec{q}\cdot\vec{r})\rangle_\Omega = {}^{\sin(q\cdot r)}/_{q\cdot r}$, so the orientationally averaged intensity of the entire sample becomes:

$$I(q) = 4\pi N \int_0^\infty r^2 \langle\gamma(\vec{r})\rangle_\Omega \frac{\sin(q\cdot r)}{q\cdot r}\, dr = 4\pi N \int_0^{D_{max}} p(r) \frac{\sin(q\cdot r)}{q\cdot r}\, dr \tag{2.21}$$

where $p(r) = r^2\langle\gamma(\vec{r})\rangle_\Omega$ is known as the pair-distance distribution. The pair-distance distribution represents a distribution of pair-wise distances within the molecule. D_{max} is the largest possible distance between two scatterers in the molecule, and thus corresponds to the size of the molecule.

3 Methods

3.1 Materials

The membrane scaffolding proteins (MSPs) used in the experiments (csMSP1ΔH5, lsMSP1ΔH5, csMSP1D1, lsMSP1D1, csMSP1E3D1, and lsMSP1E3D1) were expressed and purified by Nicolai T. Johansen and Frederik G. Tidemand using the protocol from [33] and stored at -80 °C. Proteorhodopsin was expressed by Nicolai T. Johansen according to [32]. 1-palmitoyl-2- oleoyl-glycero-3-phosphocholine (POPC) was purchased from Avanti Polar Lipids as powder, while n-dodecyl -D-maltoside (DDM), sodium cholate, Amberlite XAD® 2 resin beads, disodium phosphate (Na_2HPO_4), 70 % perchloric acid ($HClO_4$), ascorbic acid ($C_6H_8O_6$), sulfuric acid (H_2SO_4), and ammonium molybdate tetrahydrate ($(NH_4)6Mo_7O_{24}\times 4\ H_2O$) were purchased from Sigma-Aldrich.

All experiments were performed in buffer containing 20 mM Tris/HCl, 150 mM NaCl at pH 7.5.

3.2 Proteins and UV-absorption

UV-absorption is often used for different biochemical methods. Some molecules absorb radiation at specific wavelengths. Amino acids containing aromatic rings such as tryptophan, tyrosine, and phenylalanine have distinct absorption peaks in the UV-spectra at approximately 280 nm. Hence, most proteins will absorb at this wavelength, and the absorbance depends on the concentration and the number of aromatic amino acids in the given protein. Beer-Lamberts law relates the absorption of radiation at a given wavelength to the characteristics of the sample it travels through [48]. It is given as:

$$Abs = c \cdot \varepsilon \cdot l \tag{3.1}$$

where Abs is the absorption of the sample, c is the molar concentration, ε is the molar extinction coefficient in $M^{-1} \cdot cm^{-1}$ and l is the thickness of the sample given in cm. The molar extinction coefficient is a measure of how much radiation a molecule absorbs at a specific wavelength. For proteins absorption at 280 nm almost exclusively depends on the number of aromatic amino acids, so the molar extinction coefficient can be determined if the amino acid sequence is known [21]. As external factors such as the packing of the amino acids within the protein can affect the degree of absorption, the calculated extinction coefficient will be an estimate. Different programs and online calculators can estimate the extinction coefficient, for instance, ProtParam [56], which was the one used in this project. Beer-Lamberts law has some chemical and instrumental limitations such as assuming the radiation is monochromatic, and that the relation between the concentration and absorbance is always linear. At high sample concentrations and/or if the sample scatters much, the interactions between the molecules in the sample needs to be taken into account as they can cause the linearity to break down [43] and at too low concentrations the signal to noise ratio may be too small to obtain an accurate result.

3.3 Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) is a method for separating molecules in a solution according to their sizes. The sample passes through a separation column, which is packed with porous polymer beads with pores of different sizes. As the smaller molecules can enter more pores, it takes them a longer time flowing through the column while the larger molecules elute faster, as seen in top of figure 3.1. To monitor the elution of proteins with aromatic amino acids, the UV-absorption is measured at 280 nm right as

the sample elutes from the column. Given that all molecules in the sample are loaded at the same time, molecules of the same size will elute at the same time and generate a peak in the absorption signal, as seen in the bottom of figure 3.1. The absorption signal is usually measured as function of the elution volume instead of the elution time as the volume is independent of the flow rate. If the column has been calibrated with a standard of known size, it is possible to transform elution volume to Stokes diameter, which is the diameter of a sphere with the same diffusing rate as the given molecule.

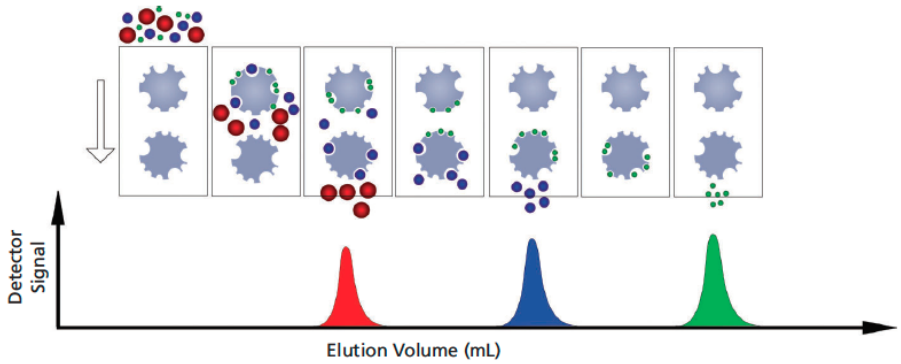


Figure 3.1: *Illustration of the idea behind size exclusion chromatography. The largest (red) molecules flow through the column and elute first, while the medium sized (blue) and the small (green) molecules elute later. Modified from [68].*

As Stokes diameter depends on the structure of the molecule, obtained sizes from SEC can only be estimations. As an example will an unfolded protein tend to have a larger Stokes diameter than the same protein would in the folded state. Nonetheless, it is still a suitable tool for purification and size estimation. The advantages of SEC also include that it does not rely on binding between polymer beads in the column and the molecules of the sample, contrary to many other chromatographic methods.

3.3.1 Experimental SEC

The sample can then be purified by size-exclusion chromatography (SEC) to remove large aggregates and excess molecules. If membrane proteins are to be inserted in the nanodiscs, they should be added in the beginning before adding detergent absorbing beads. If membrane proteins are added, additional purification is needed to remove empty discs.

At first, stock solutions of buffer containing either 100 mM cholate or 75 mM DDM were prepared. For all experiments with cholate, the POPC was dissolved in the cholate stock by sonication in a water bath to reach a concentration of 50 mM POPC. For the experiments with only DDM, POPC was dissolved in the DDM stock to reach a concentration of 25 mM POPC in the same way as with cholate.

In Eppendorf tubes, POPC, MSP and cholate and/or DDM was mixed and diluted with buffer to reach final volumes of 400 μL. For each MSP there were 11 samples containing 10 mM POPC, 20 mM cholate and varying concentrations of DDM (0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, and 50 mM), while the 12th sample had 10 mM POPC and only 30 mM DDM (no cholate). Different stoichiometries were used for the different MSPs as seen in table 3.1; for lsMSP1ΔH5 and csMSP1ΔH5 the ratio MSP:POPC was 1:60, for lsMSP1D1 and csMSP1D1 the ratio MSP:POPC was 1:70, and for lsMSP1E3D1 and csMSP1E3D1 the ratio MSP:POPC was 1:125.

The samples were placed in a thermomixer at 15 °C at 700 rpm for 10 minutes to ensure a thorough mixing before adding the detergent absorbing resin beads Amberlite XAD® 2. Before the addition

	No. of amino acids	Extinction coefficient, ε	MSP:POPC ratio
lsMSP1ΔH5	190	18450	1:60
csMSP1ΔH5	180	18450	1:60
lsMSP1D1	210	18450	1:70
csMSP1D1	200	18450	1:70
lsMSP1E3D1	270	26930	1:125
csMSP1E3D1	260	26930	1:125

Table 3.1: *Table with information on the different MSPs. The number of amino acids and the extinction coefficients are from ProtParam [56].*

to the samples, the Amberlite beads were washed once in methanol (MeOH) and 5 times in Mili-Q water to remove preservatives and resin monomers by decantation. To ensure complete removal of the detergents 0.15 g/mL Amberlite beads were used for all samples and additional $3 \cdot c_{\mathrm{DDM}}$ mg beads were used for samples with DDM, where c_{DDM} is the DDM concentration. The time between adding MSP and Amberlite beads was not important as it has previously been determined that the rate of removing cholate does not have any significant impact on the reconstitution [61].

As the detergent removal process is temperature dependent and may take up to 12 hours [57], the samples were incubated in a thermomixer at 15 °C at 700 rpm for at least 12 hours. After the incubation, the Eppendorf tubes containing the samples were punctured in top and bottom with a needle and centrifuged into 15 mL tubes for 2 min at 200 RCF to separate the beads from the nanodisc samples. The samples with 30 mM DDM and without cholate were more viscous and opaque than the other samples, so they had to be centrifuged again, to remove precipitants.

Afterwards, the samples were fractionated by size exclusion chromatography (SEC) using a Superdex® 200 increase 10/300 GL column on a BioRad NGC™ Quest 10™ Chromatography System. The column was equilibrated with buffer, and the experiments were performed using a 500 μL sample-loop and a flow rate of 1 mL/min. The elution was tracked with UV absorption at 280 nm and samples were collected from the main peaks of the SEC elution profiles and stored at 4 °C. These were the samples used for concentration-determination, SAXS, and phosphate analysis.

3.3.2 SEC of samples with proteorhodopsin (PR)

The nanodisc samples containing the membrane protein proteorhodopsin (PR) were prepared in almost the same way as the samples without. A total of twelve samples were prepared, one for each of the six MSPs, with and without 30 mM DDM.

PR was added in the initial step, where it was mixed with MSP, POPC, cholate and for half of the samples DDM to reach a final volume of 400 μL. All the produced samples contained 20 mM cholate and 5 mM POPC, while the samples with DDM had a concentration of 30 mM DDM. To avoid excess PR, monomer PR was added to the samples in the ratio PR:MSP of 1:6. The produced samples had ratios POPC:MSP of 50:1 for lsMSP1ΔH5 and csMSP1ΔH5, 60:1 for lsMSP1D1 and csMSP1D1, and 115:1 for lsMSP1E3D1 and csMSP1E3D1. Furthermore, 1.5 μL of the reducing dithiothreitol ($C_4H_{10}O_2S_2$,DTT) was added to all samples to prevent the formation of disulfide bridges. The samples were placed in a thermomixer at 700 rpm for 10 minutes before adding Amberlite beads and then the samples were incubated in a thermomixer at 15 °C at 700 rpm for at least 12 hours. The beads were separated from the sample by centrifuging and SEC was performed using a Superdex® 200 increase 10/300 GL column

on an ÄKTApurifier (GE-Healthcare®). A sample-loop of 500 μL and a flow rate of 0.75 mL/min was used and the elution was tracked with UV absorption at 280 nm and 530 nm. As PR does not protrude much from the membrane, the Stokes diameter and thus the elution peak of the nanodiscs reconstituted with PR is not expected to change significantly from that of the empty discs. Accordingly, absorption measurements at 280 nm do not distinguish discs with PR from empty discs. As PR has an absorption peak at 530 nm, where the MSPs will not absorb, measurements of absorption at 530 nm simultaneously with 280 nm enables monitoring of nanodiscs with PR only.

3.4 Protein concentration determination

The protein concentration of the samples were determined through measurements of the UV absorption at 280 nm, using a Thermo Scientific Nanodrop1000 Spectrophotometer. Three different absorption measurements of each 2 μL sample were made for better statistics. The extinction coefficients of the MSP's were calculated from their amino acid sequence using the ProtParam tool [56]. Knowing the extinction coefficients and assuming two MSPs per disc, the concentration of nanodiscs in the sample could be calculated using Beer-Lamberts law, eq. 3.1.

3.5 Phospholipid concentration determination

The concentration of phospholipids was determined through phosphate analysis. The concentration of phosphate is proportional to the concentration of lipids, as each lipid has a phosphate group in the lipid headgroups.

Phosphate analysis was conducted in triplicates. For each sample, a volume containing approximately 50 nmol lipid was transferred to three different long glass tubes with 400 μL of 70 % perchloric acid. The tubes were loosely closed with glass pearls and cooked for two hours at 180 °C to break down the lipids and release the phosphate from the head groups. The use of long test tubes was to ensure a temperature gradient so the perchloric acid would not evaporate and escape but condense on the upper sides of the glass tubes.

After the cooking, the samples were equilibrated to room temperature and 4 mL molybdate reagent were added. The molybdate reagent was made of 2.2 g ammonium molybdate tetrahydrate ($(NH_4)6Mo_7O_{24}\times 4 H_2O$), 14.3 mL concentrated sulfuric acid (H_2SO_4) and Mili-Q water until it reached a volume of 1 L. 500 μL freshly made 10 % w/w ascorbic acid ($C_6H_8O_6$) were added and the samples were heated to 80 °C in a water bath for 10 minutes. The molybdate reagent and the ascorbic acid bind the free phosphate to make a complex, which has an absorption peak in the UV-spectrum at approximately 812 nm, so the absorption was measured at 812 nm on a VWR UV-1600PC spectrophotometer.

As the absorbance and the phosphate concentration, and thus the lipid concentration, are proportional it is possible to determine the phosphate concentration by using a standard curve of known phosphate concentrations. Hence, triplicates of four standards containing 20, 40, 70, and 120 nmol disodium phosphate (Na_2HPO_4) were prepared in the same way as the samples. Furthermore, two blanks without any phosphate were prepared for calibration of the spectrophotometer.

3.6 Small-angle X-ray scattering (SAXS)

Two beamtimes for small-angle X-ray scattering (SAXS) experiments were generously granted: the preliminary experiments were performed at P12 beamline at DESY, Hamburg, while the main SAXS experiments were performed at the B21 beamline at Diamond Light Source, Chilton, Didcot, Oxfordshire [12].

Both beamtimes were mail-in, as the global COVID-19 pandemic did not allow travelling.

Samples from the SEC fractions and buffer for background measurements were shipped, and kept at 4 °C during transport. Measurements were performed with standard settings, wavelength $\lambda = 1$ Å, q-range of 0.0046 $Å^{-1}$ - 0.33 $Å^{-1}$ at 10 °C. All the the experimental settings can be seen in table 3.2.

The samples were measured in groups of eight, with buffer measurements before and after every eighth sample. During exposure, the sample or buffer flowed through the sample cell at a flow rate of 1 $^{\mu L}/_{s}$, to prevent radiation damage. Each measurements consisted of 22 frames with exposure for 2 s.

Wavelength	0.95242 Å (12.4 keV)
Sample cell diameter	1.75 mm
Beam size at sample	1.1 × 0.5 mm
Temperature	10 °C
Sample to detector distance	3705.8347 mm
Beam intensity	4×10^{12} $^{photons}/_{s}$
Exposure time	2 s
Flowrate	1 $^{\mu L}/_{s}$
Number of frames	22
q-range	0.0046-0.33 $Å^{-1}$
Type of detector	Eiger 4M from Dectris

Table 3.2: *Table with parameters of the settings of the B21 beamline [17].*

4 Data analysis

4.1 SAXS data processing

SAXS data has to undergo data processing before modelling can be done. As mentioned in the method section the samples for this project were measured in groups of eight, with buffer measurements before and after every eight samples. For each sample and buffer, the measurements were comprised of 22 frames, and all data frames were visually inspected to check for deviations and outliers before averaging.

At the beamline the data was calibrated with a water standard according to the relation:

$$I_{sample,abs} = (I_{sample,meas} - I_{background,meas})\frac{I_{water,meas}}{I_{water}} \tag{4.1}$$

where $I_{sample,abs}$ is the intensity of the sample on absolute scale $I_{sample,meas}$ is the total measured intensity (sample, buffer, capillary, and noise), $I_{background,meas}$ is the measured intensity from the background (buffer, capillary, and noise), $I_{water,meas}$ is the measured intensity from a water standard, and $I_{water} = 1.632 \cdot 10^{-2}$ cm^{-2} is the known water standard [49].

Each buffer-subtracted and calibrated data set was logarithmically rebinned from 2567 data points to 118 points. The logarithmic rebinning ensures even spacing between the data points on a logarithmic scale while preserving the overall shape of the curve. As a bonus, the rebinning also reduces computation time when fitting the data.

4.2 Modelling of SAXS-data

SAXS data can be interpreted in different ways to extract as much information as possible. They all require a procedure for optimization and evaluation, such as the least-squares method.

4.2.1 Least-squares method

The least-squares method is one of the most widely used approaches when evaluating data and fitted models. The concept is to minimize the deviation between the measured data and the model, as weighted by the uncertainty of the measured data. This is done using the χ^2 (pronounced *chi-squared*) function, which in context of SAXS-experiments, is given as [52]:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{I_i^{exp} - I^{model}(q_i)}{\sigma_i} \right)^2 \tag{4.2}$$

where I_i^{exp} is the experimental intensity as function of the magnitude of scattering vector q with N data points, $I^{model}(q_i)$ is the calculated intensity from the model, and σ_i is the experimental uncertainty. Generally, it is easier to evaluate the reduced χ^2, denoted χ^2_{red} as it also takes the number of fitted parameters and degrees of freedom into account:

$$\chi^2_{red} = \frac{\chi^2}{N - M} \tag{4.3}$$

where N is the number of data points, M is the number of fitted parameters, and N-M is the number of degrees of freedom. χ^2_{red} is extensively used as a measure for goodness of fit. Traditionally an ideal fit

should have $\chi^2_{red} \simeq 1$.

4.2.2 Pair-distance distribution

The pair-distance distribution and the intensity from a diluted and monodisperse sample is related according to eq. 2.21 [64]:

$$I(q) = 4\pi N \int_0^{D_{max}} p(r) \frac{\sin(q \cdot r)}{q \cdot r} \, dr \tag{4.4}$$

where I(q) is the intensity as a function of the scattering vector q, N is the number of sample molecules, D_{max} is the size of the molecules, and *p(r)* is the pair-distance distribution function.

p(r) is the distribution of distances of pairs of scatterers within the molecules as weighted by their excess scattering lengths, and can thus be used to estimate the size of the molecules. The pair-distance distribution can be found as the inverse Fourier transform of the intensity given in eq. 4.4. However, since the intensity not known in the entire range $0 < I(q) < \infty$, and consequently the equation cannot be transformed directly. Glatter developed the Indirect Fourier Transform (IFT) in 1977 [22] to solve this issue. The IFT approximates the *p(r)* distribution by a linear combination of a finite number of weighted basis functions [26], and transforms the scattering data from reciprocal space into real space. The method requires a smoothness parameter, a number of basis functions, and a maximum distance in real space. As the maximum distance typically is an unknown but desired parameter in regard to structural studies, a Bayesian method was developed for estimating the maximum distance as part of the IFT-routine in the Bayesian Indirect Fourier Transform or BIFT [26].

If the scattering length density changes throughout a molecule, the value of *p(r)* may be negative for certain ranges of r. *p(r)* also depends on the difference in scattering length density between the different parts of the molecule and the background solution. If the scattering length density of one region of the molecule is lower than that of the background, while another part of the molecule has a scattering density larger than that of the background, the difference in contrast signs causes a negative values of *p(r)*.

Although IFT is an excellent tool, it strictly depends on a monodisperse system, as larger impurities such as aggregates can cause an overestimation of the molecule size.

4.2.3 BayesApp

The online tool BayesApp [9, 27] has been developed for calculating *p(r)*-functions from SAXS data using a BIFT algorithm. Within seconds *p(r)* can be determined and it only requires a file containing the length of the scattering vector; q, the intensity as function of q; $I(q)$, and the error on I; σ. All *p(r)* functions in this project were obtained using BayesApp.

Another feature of BayesApp is the ability to evaluate whether the experimental errors on the data are under- or overestimated [39, 18]. The χ^2_{red} determined from BIFT, $\chi_{red,BIFT}$, can be used to recale the errors on a dataset according to the relation:

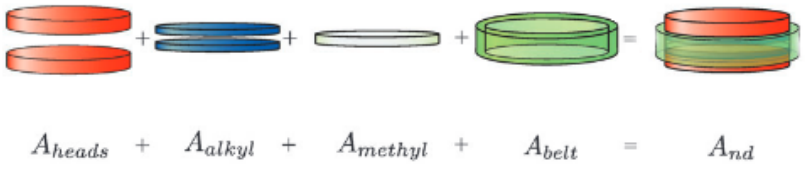
$$\sigma_{rescaled} = \sigma_{meas} \sqrt{\chi^2_{red,BIFT}} \tag{4.5}$$

where $\sigma_{rescaled}$ is the rescaled errors and σ_{meas} is the measured errors. All the errors on the SAXS data for this project were overestimated and was thus rescaled according to eq. 4.5 with a factor $\simeq 0.3$.

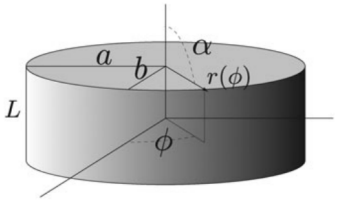
4.2.4 The nanodisc model

Scattering data contain information on the three-dimensional structure of the sample molecules. However, in the orientational averaging of the detector image, the data are reduced to one-dimensional data. A lot of information is bound to get lost, as it is impossible to calculate the three-dimensional structure directly from the scattering data, and multiple structures can yield the same one-dimensional scattering curve. By turning the problem upside down, one can start by developing a model based on the assumptions and limitations of the given system, and then evaluate it in relation to the experimental data. One such model is the nanodisc model.

Nanodiscs are an example of molecules where the scattering length density is not constant throughout the molecule, but changes between the different components. By assuming the scattering length density is constant within the different parts making up the nanodisc, it is possible to build a geometrical model for describing nanodiscs. It was developed and published by Skar-Gislinge et al. in 2010 [59] and is based on expressing different parts of the nanodisc as cylindrical structures with elliptical cross-sections, as seen in figure 4.1a. The top and bottom cylinders represent the hydrophilic head groups while the hydrophobic core of the lipid tails are split into the regions of the methyl-groups (tips of the tails) and alkyl-groups (the rest of the tails). The MSP is represented as a hollow cylinder enclosing the edges of the other layers.



(a) *Graphic of the principle for determining the form factor of the nanodisc. From [59].*



(b) *Illustration of an elliptical cylinder visualizing the different parameters of the nanodisc model. From [59].*

Figure 4.1

From eq. 2.12, eq. 2.15, and eq. 2.16 it can be seen that the spatially averaged intensity from a monodisperse sample of diluted molecules is given as:

$$I(q) = c\left\langle |A(\vec{q})|^2 \right\rangle_{\alpha,\phi} = c\left\langle |\Delta b^2 F(\vec{q})|^2 \right\rangle_{\alpha,\phi} \tag{4.6}$$

where c is the number density $c = N/V$, $A(\vec{q})$ is the scattering amplitude, $F(\vec{q})$ is the form factor amplitude, $\Delta b = V \cdot \Delta\rho$ is the excess scattering length, $\Delta\rho$ is the excess scattering length density and V is the volume of a molecule.

The total scattering amplitude of the disc is a linear sum of the contributions from the individual parts:

$$\begin{aligned} A_{ND} &= A_{head} + A_{alkyl} + A_{methyl} + A_{belt} \\ &= \Delta b_{head} F(\vec{q})_{head} + \Delta b_{alkyl} F(\vec{q})_{alkyl} + \Delta b_{methyl} F(\vec{q})_{methyl} + \Delta b_{belt} F(\vec{q})_{belt} \end{aligned} \tag{4.7}$$

The form factor amplitude for a cylinder can be calculated from eq. 2.16 and is given as:

$$F_{cyl}(q, \alpha, \phi, r, L) = \frac{2J_1(qr(\phi)\sin(\alpha))}{qr(\phi)\sin(\alpha)} \frac{\sin(qL\cos(\alpha)/2)}{ql\cos(\alpha)/2} \tag{4.8}$$

where J_1 is the first order Bessel function, α is the azimuthal angle, L is the height of the cylinder, and r(ϕ) is the variable radius of the cylinder which is a function of the horizontal angle ϕ:

$$r(\phi) = \sqrt{a^2 \sin^2(\phi) + b^2 \cos^2(\phi)} \tag{4.9}$$

Where a and b are the long and short half axes respectively, allowing an elliptical shape of the cylinder. Fig. 4.1b depicts the parameters of the cylindrical form factor amplitude.

For the headgroups, alkyls, and the methyl, the form factor amplitude is obtained using eq. 4.8 with different heights of the layers but the same cross-section. The belt protein is constructed as a hollow cylinder; a large outer elliptical cylinder minus a smaller inner elliptical cylinder.

All the different components can be assembled to calculate the form factor amplitude of the entire nanodisc $A_{ND} = \Delta b_{ND} F(\vec{q})_{ND}$, where Δb_{ND} is the nanodisc excess scattering length and $F(\vec{q})_{ND}$ is the nanodisc form factor. Orientational averaging is required, as the nanodiscs can be randomly oriented, so the intensity can be calculated as an integral over the two angles:

$$I_{ND}(q) = \frac{2c}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} |A_{ND}|^2 \sin(\alpha)\, d\phi\, d\alpha = \frac{2c}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} |\Delta b_{ND} F(q, \phi, \alpha)_{ND}|^2 \sin(\alpha)\, d\phi\, d\alpha \tag{4.10}$$

4.2.5 Parameters in the nanodisc model

The nanodisc model requires some initial inputs to determine the modelled intensity curve, such as the scattering lengths and molecular volumes of the different components of the model (head group, methyl, alkyl, belt protein, and solvent). In addition to the inputs, the model also depends on other fitting parameters:

- **Axis ratio, $\mathbf{A_r}$**: the ratio between the major and the minor axis of the nanodisc, i.e. how elliptical the bilayer is. A completely circular disc will have an axis ratio of 1, while a larger axis ratio indicates a more elliptical disc. Previous studies have shown the elliptical shape of nanodiscs depends on the temperature [60].
- **Area per headgroup, $\mathbf{A_{head}}$**: the area per lipid headgroup. Determines the area of the entire lipid bilayer, $A_{bilayer} = A_{head} \cdot N_{lip}$, where $A_{bilayer}$ is the area of the bilayer and N_{lip} is the number of lipids (parameter described below). As the volume of all lipids is preserved, the area per headgroup also determines the height of the bilayer, $H_{bilayer}$, through the relation $H_{bilayer} = V_{bilayer}/A_{bilayer}$, where $V_{bilayer}$ is the volume of the bilayer. Previous studies have shown that the area per headgroup depends on the temperature [38].
- **Number of lipids, $\mathbf{N_{lip}}$**: the number of lipids contained in one nanodisc. It determines the area of the disc by the relation $A_{ND} = A_{head} \cdot N_{lip}$, where A_{ND} is the entire area of the disc, A_{head} is the area per headgroup.
- **Height of belt, $\mathbf{H_{belt}}$**: the vertical height of the MSP belts. In the same manner as with the lipids, the volume of the belt is also preserved, so the height of the belt regulates the width of the belt, W_{belt}. The height was kept fixed at 25.8 Å [33].

- **Roughness, R**: correction factor to account for roughness at the interfaces. As the model assumes infinitely sharp interfaces it causes some noticeable non-physical features in the scattering curve. Real-life interfaces are not completely smooth but have some degree of random roughness, so a Gaussian function is applied to account for this; $P_{final}(q) = P(q)\exp(-(q \cdot R)^2/2)$, where $P(q) = |F(q)|^2$ is the form factor and R is the interface roughness parameter. To reduce the number of fitted parameters the roughness is assumed to be an average roughness of all interfaces, and it is not specified for individual interfaces.

- **Correction for volume of the belts, $\mathbf{CV_{belts}}$**: correction for the initially estimated volume of the belt, $V_{belt,final} = CV_{belt} \cdot V_{belt,initial}$, where $V_{belt,initial}$ and $V_{belt,final}$ are the initial and final volumes respectively . CV_{belt} affects the scattering length density of the belt, $\Delta\rho_{belt}$, and thus the general scattering pattern of the model through the relation, $\Delta\rho_{belt} = b_{belt}/V_{belt,final}$, where b_{belt} is the scattering length of the belt. Preferably $CV_{belt} \simeq 1$ as this indicates a good initial "guess" of the volume.

- **Correction for the volume of the lipids, $\mathbf{CV_{lip}}$**: as with the belt, it is the correction for the initially estimated volume of the lipids, $V_{lip,final} = CV_{lip} \cdot V_{lip,initial}$, where $V_{lip,initial}$ and $V_{lip,final}$ are the initial and final volumes respectively . Preferably $CV_{lip} \simeq 1$.

- **Background**: constant background value subtracted from the data.

4.2.6 WillItFit

WillItFit is a software developed by Pedersen et al. [70] to fit semi-analytical models to SAXS data [53]. By using the geometrical form factor amplitudes to determine the intensity of a given model, it can fit the model to SAXS data. The software has multiple fitting algorithms incorporated to select for fitting. As input, the software depends on a user-defined model, one or more data sets along with the scattering lengths and partial specific volumes of the components in the given model, including the solvent. In this project, the Levenberg-Marquardt algorithm and the previously described nanodisc model were used. The Levenberg-Marquardt algorithm is a numerical algorithm used to solve non-linear problems, and it relies on the least-squares method as described in section 4.2.1.

The volumes and scattering lengths used can be seen in table 4.1.

4.3 Phosphate analysis

In the phosphate analysis, the phosphate from the lipid head groups is released and bound to molybdate in a compound that has an absorption peak in the blue spectra at approximately 812 nm. As there is one phosphate group in the head group of each lipid, the degree of absorption depends on the amount of lipid in the sample. By using a standard curve, with absorption of standards with known phosphate content, the amount of lipid in the nanodisc samples can be determined.

From Beer-Lambert's law eq. 3.1 it follows that the absorption is proportional to the amount of substance n, as the concentration c is equal to $c = n/V$, where V is the volume of the sample. From the phosphate analysis of the standards with known phosphate contents, the absorption at 812 nm and the amount of substance can be fitted to a straight line using the curve fitting tool from the scipy package in Python and the equation for a straight line:

$$\mathrm{Abs}_{812nm} = a \cdot n_{lip} + b \tag{4.11}$$

	Chemical formula	Partial specific molecular volume, V, [Å^3]	Scattering length, b, [cm]
lsMSP1ΔH5D1	$C_{941}H_{1474}N_{266}O_{310}S_3$	24,329.3	$3.24{\cdot}10^{-9}$
csMSP1ΔH5D1	$C_{895}H_{1416}N_{244}O_{299}S_3$	23,048.8	$3.08{\cdot}10^{-9}$
lsMSP1D1	$C_{1046}H_{1648}N_{296}O_{343}S_3$	26,890.3	$3.60{\cdot}10^{-9}$
csMSP1D1	$C_{1000}H_{1590}N_{274}O_{332}S_3$	25,609.8	$3.44{\cdot}10^{-9}$
lsMSP1E3D1	$C_{1393}H_{2195}N_{395}O_{453}S_5$	35,341.5	$4.80{\cdot}10^{-9}$
csMSP1E3D1	$C_{1347}H_{2137}N_{373}O_{442}S_5$	34,061.0	$4.64{\cdot}10^{-9}$
POPC headgroup	$C_{10}H_{18}NO_8P$	319.0	$4.62{\cdot}10^{-11}$
POPC alkyl	$C_{30}H_{58}$	818.8	$6.71{\cdot}10^{-11}$
POPC methyl	C_2H_6	108.6	$5.08{\cdot}10^{-12}$
Water	H_2O	30	$2.82{\cdot}10^{-12}$

Table 4.1: *Table with scattering parameters of the different MSPs, POPC, the solvent (water) used for fitting in WillItFit [70]. The scattering lengths were calculated from the number of electrons in the molecules from their chemical formulas. The partial specific molecular volumes of POPC were determined in [60], POPC headgroup in [63], the volumes of the methyl groups in the phospholipids were calculated using Tanford's formula [67], and the volumes of the MSP were from ProtParam [56, 21].*

where Abs_{812nm} is the absorption at 812 nm, n_{lip} is the amount of lipids in nmol, and a and b are the fitted parameters. The concentration of lipids in the samples as determined by phosphate analysis is thus:

$$c_{lip} = \frac{n_{lip}}{V} = \frac{\text{Abs}_{812nm} - b}{V \cdot a} \tag{4.12}$$

By measuring the absorption at 280 nm the concentration of nanodiscs in the sample can be determined. Assuming 2 MSPs per nanodisc, the concentration of nanodiscs in the sample can be estimated from absorption measurements using Beer-Lambert's law eq. 3.1:

$$c_{ND} = \frac{\text{Abs}_{280nm}}{2 \cdot l \cdot \varepsilon_{MSP}} \tag{4.13}$$

where c_{ND} is the concentration of nanodiscs in μM, Abs_{280nm} is the absorption at 280 nm, l is the path length in cm, and ε_{MSP} is the extinction coefficient of the MSP in $M^{-1}cm^{-1}$.

The number of lipids per disc as determined by the phosphate analysis, $N_{lip,pi}$ is then given as the ratio between the lipid concentration and the nanodisc concentration:

$$N_{lip,pi} = \frac{c_{lip}}{c_{ND}} = \frac{2(\text{Abs}_{Pi,812nm} - b)l \cdot \varepsilon_{MSP} \cdot 10^{-3}}{a \cdot V \cdot \text{Abs}_{280nm}} \tag{4.14}$$

Where the factor 10^{-3} is too ensure the right units; l has units of cm, ε_{MSP} has units of $M^{-1}cm^{-1} = L \cdot mol^{-1} \cdot cm^{-1}$, a has units of $^1/_{nmol}$, V has unit μL, while the two absorptions and b are dimensionless. By assuming Gaussian errors, the law of propagation of errors can be applied to eq. 4.14 to calculate the uncertainty on the number of lipids per nanodisc as determined by the phosphate analysis.

4.4 SEC curves and Gaussians

For a sample of monodisperse molecules, the elution peak of a SEC curve is expected to resemble a Gaussian distribution [35]. However, the Gaussian distribution does not take the observed skewness of chromatographic peaks into account. Diffusion of molecules inside the column, the length of the column, flowrate, dead volumes in the system, and other effects can influence the peak shape and contribute to band broadening. It seems the function describing elution peaks best is the exponentially modified Gaussian function (EMG) [11, 40]. The EMG is a convolution of a Gaussian function and an exponentially decaying function. It can be described by the following equation, where the $*$ marks the convolution of the two functions, defined as $g(x) * f(x) = \int_{-\infty}^{\infty} f(x - x')\, g(x')\, dx'$ [34, 23]:

$$\begin{aligned} EMG(V, h, \mu, \sigma, \tau) &= \text{gauss}(V, h, \mu, \sigma) * \exp(V, \tau) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{V-\mu}{\sigma^2}\right)^2\right) * \exp\left(-\frac{V}{\tau}\right) \\ &= \frac{h\sigma}{\tau}\sqrt{\frac{\pi}{2}} \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{V-\mu}{\tau}\right) \operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu - V}{\sigma} + \frac{\sigma}{\tau}\right)\right) \end{aligned} \tag{4.15}$$

Where V is the volume, h is the height of the peak, σ and μ are the variance and mean of the Gaussian part respectively, while τ is the relaxation parameter of the exponential part, and erfc(z) is the complementary error function:

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} \exp(-V^2) dV \tag{4.16}$$

Besides the variance, mean, and relaxation parameter, a characteristic of an EMG is the skewness γ_1. As skewness affects the tailings of the EMG, skewness can be used as a measure of the degree of tailing. The skewness increases with increasing relaxation parameter and it is given as [25]:

$$\gamma_1 = \frac{2\tau^3}{(\sigma^2 + \tau^2)^{3/2}} \tag{4.17}$$

5 Results and discussion

In order to investigate the relationship of DDM and the different MSPs, the following experiments were conducted. The nanodiscs were assembled with different MSPs and varying concentrations of DDM and studied using size exclusion chromatography (SEC), small-angle X-ray scattering (SAXS), and phosphate analysis.

First, the samples of each of the following MSP types lsMSP1ΔH5, csMSP1ΔH5, lsMSP1D1, csMSP1D1, lsMSP1E3D1, and csMSP1E3D1 were prepared. For each MSP, all samples but one were prepared with a POPC concentration of 10 mM and a cholate concentration of 20 mM, while the last sample contained 10 mM POPC and 30 mM DDM. In addition, increasing concentrations of DDM were included to finally yield the following series:

- lsMSP1ΔH5 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 60:1.
- csMSP1ΔH5 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 70:1.
- lsMSP1 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 70:1.
- csMSP1 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 60:1.
- lsMSP1E3 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 125:1.
- csMSP1E3 NDs: 11 samples with 10 mM POPC, 20 mM cholate and 0, 0.5, 1, 2, 3, 4, 5, 10, 20, 30, or 50 mM DDM and 1 sample with 10 mM POPC and 30 mM DDM. All samples were prepared with a POPC:MSP ratio of 125:1.

The choice of reconstitution stoichiometries of the POPC:MSP ratios were based on previous studies [32, 8] and bachelor projects conducted in our research group.

As mentioned in the Methods section the detergents are removed by detergent absorbing resin beads, and there are no detergents left when the nanodiscs have formed. However, in the following sections, the samples will be referred to by their reconstitution detergent concentrations, as a means to distinguish between them.

In the following sections, the results are presented and discussed according to the different MSPs. Afterwards, a more general discussion will follow.

5.1 Preliminary experiments

Nanodisc samples were prepared, and the subsequent SAXS measurements were collected at the P12 beamline at DESY, Hamburg, as well as analysed for phosphate content. However, the SAXS curves looked odd and it was discovered that I had made a critical mistake in the lab. The sample volume injected in the sample-loop for the SEC-experiments had not been adjusted, and only about a third of the injected sample had been used. Thus, the concentrations of nanodiscs in the sample fractions collected from SEC were so low that the signal from the SAXS measurements was almost indistinguishable from the buffer background, resulting in cases of negative intensities after averaging and background subtractions. An example of the data from the first experiments can be seen in figure 5.1. The data are from the optimization of the DDM concentration for csMSP1D1ΔH5.

As seen from panel C the samples reconstituted with 2.5 mM DDM, 20 mM DDM, 30 mM DDM, and only DDM have negative signals at the dip at q $\sim$ 0.07 Å and the larger q-values, and thus will not show on a logarithmic scale. It was not possible to fit the nanodisc model to any of the scattering curves with negative signals. Having spent many hours in the lab it was a very unfortunate mistake and a setback. Afterwards, it was decided to repeat the experiments.

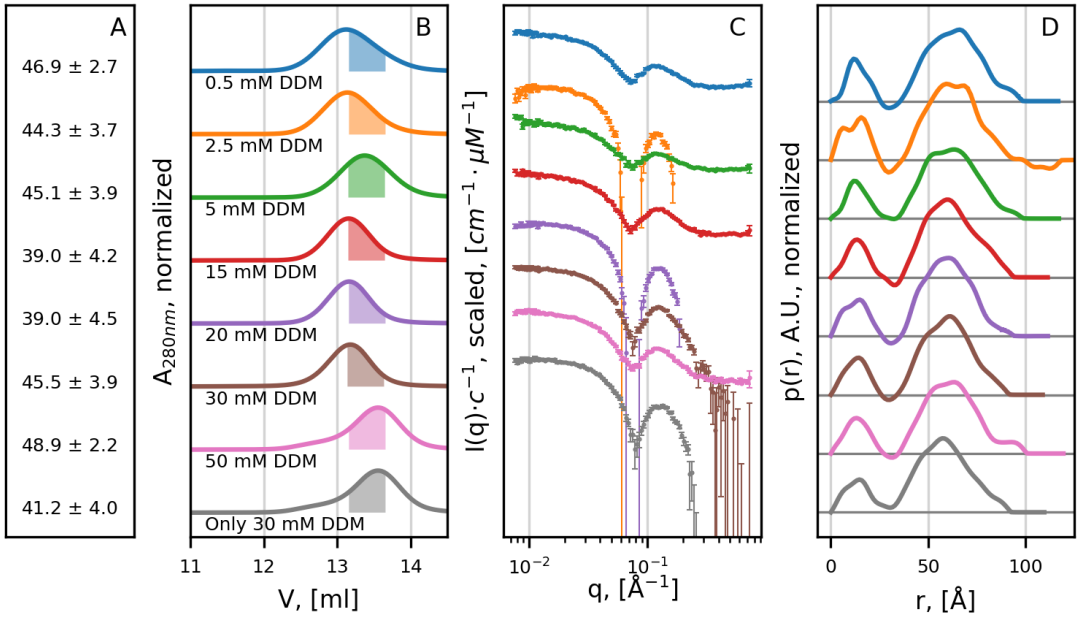


Figure 5.1: *Data from the initial experiments with the DDM concentration for csMSP1D1ΔH5 .* ***Panel A****: POPC:MSP ratio as determined from the phosphate analysis.* ***Panel B****: SEC curves with marked volume fractions used for SAXS and phosphate analysis. The curves are normalized by their height.* ***Panel C****: SAXS curves scaled by* 10^{-i+1}*, where* $i = 1, ..., 8$ *is the number of the curve.* ***Panel D****: Pair-distance distribution functions normalized by their area and scaled by factor* $i+0.2$*, where* i*=0,...,7 is the number of the curve.*

5.2 lsMSP1D1ΔH5

Figure 5.2 displays the results of lsMSP1D1ΔH5 NDs. The SEC chromatograms in panel A suggests the samples without DDM in the reconstitution process and with low concentrations of DDM display polydispersity, as the SEC chromatograms have multiple elution peaks. The nanodiscs are expected to elute at approximately 13 mL while larger particles will elute before. As marked in the figure, samples were collected from the expected nanodisc peaks and used for further analysis.

It seems that including DDM in the reconstitution process ensures a larger degree of monodispersity. With increasing amounts of DDM, the peaks from the larger particles decrease and the nanodisc peak grows more dominant. At 20 mM DDM, there is only a small peak of larger particles and for the sample with only DDM the sample appear completely monodisperse.

Most samples exhibit the approximate same elution volume for the nanodisc peak. However, there is a shift to a smaller volume, indicating a larger size, for the sample with 50 mM DDM in the reconstitution process. The sample with only DDM has a peak shifted to a larger elution volume indicating smaller discs.

The SAXS-curves (figure 5.2, panel B) have the overall characteristic nanodisc features; a flat region at low q-values followed by a decline at q ≈ 0.07 Å^{-1} and then an asymmetric peak. The flat region at low q-values indicates no intermolecular interactions and thus dilute samples. There is a striking difference in curvature for the samples with 1 mM DDM and 3 mM DDM (curve number 3 and 5 from the top) as their minima are less prominent than the other samples'. All the model fits seem to reproduce the features of the SAXS curves.

The difference plot in figure 5.2 panel C displays the difference between the SAXS curves from panel B and a reference, which is the average of the SAXS curves with cholate. In the plot, the differences between the SAXS curves are more clear while common features are diminished. The curves differ the most at low q-values, but there does not seem to be any systematic changes, as the amount of DDM is varied in the reconstitution process. The sample with only DDM deviates from the other curves at approximately 0.1 Å^{-1}. This is also the location of the peak in the SAXS curve, and when comparing the curves in panel B it is evident that the peak of the sample that was reconstituted with no cholate but only DDM, exhibits a single peak at q$\approx$0.1 Å^{-1}, while the other samples exhibits double peaks.

The refined parameters can be seen in figure 5.2 panel D and in table A.2 in appendix A.2. The low χ^2_{red} values (figure 5.2 panel D, bottom) indicate good model fits and suggest nanodiscs are present in the sample. For all samples except the one with only DDM, the number of lipids per disc, N_{lip}, is larger than the 120 number of lipids per disc estimated in the reconstitution mixture. This should not be possible, and it is probably due to experimental uncertainties.

A closer inspection of the parameters shows that the samples 1 mM DDM and 3 mM DDM deviate from the rest. These two samples, which also have visibly deviating SAXS-curves (panel B), exhibit exceptionally large axis ratios A_rs of 2.4 and 2.2, compared to the other samples with axis ratios fluctuating around 1.6. The two deviating samples also display larger areas per headgroup A_{head}s and number of lipids N_{lip}, while the estimated belt volumes V_{belt}s are smaller. The sample without DDM also deviates with the smallest area per headgroup A_{head}, belt volume V_{belt}, and roughness. Furthermore, it also displays the largest errors on all parameters.

As the nanodisc model assumes monodispersity, the variation in the parameters at low DDM concentrations could maybe be explained by the polydispersity observed in the SEC curves. As the peaks in the SEC curves overlap for samples with small DDM concentrations, the collected fractions from the nanodisc peak might be contaminated with particles from the adjacent peaks.

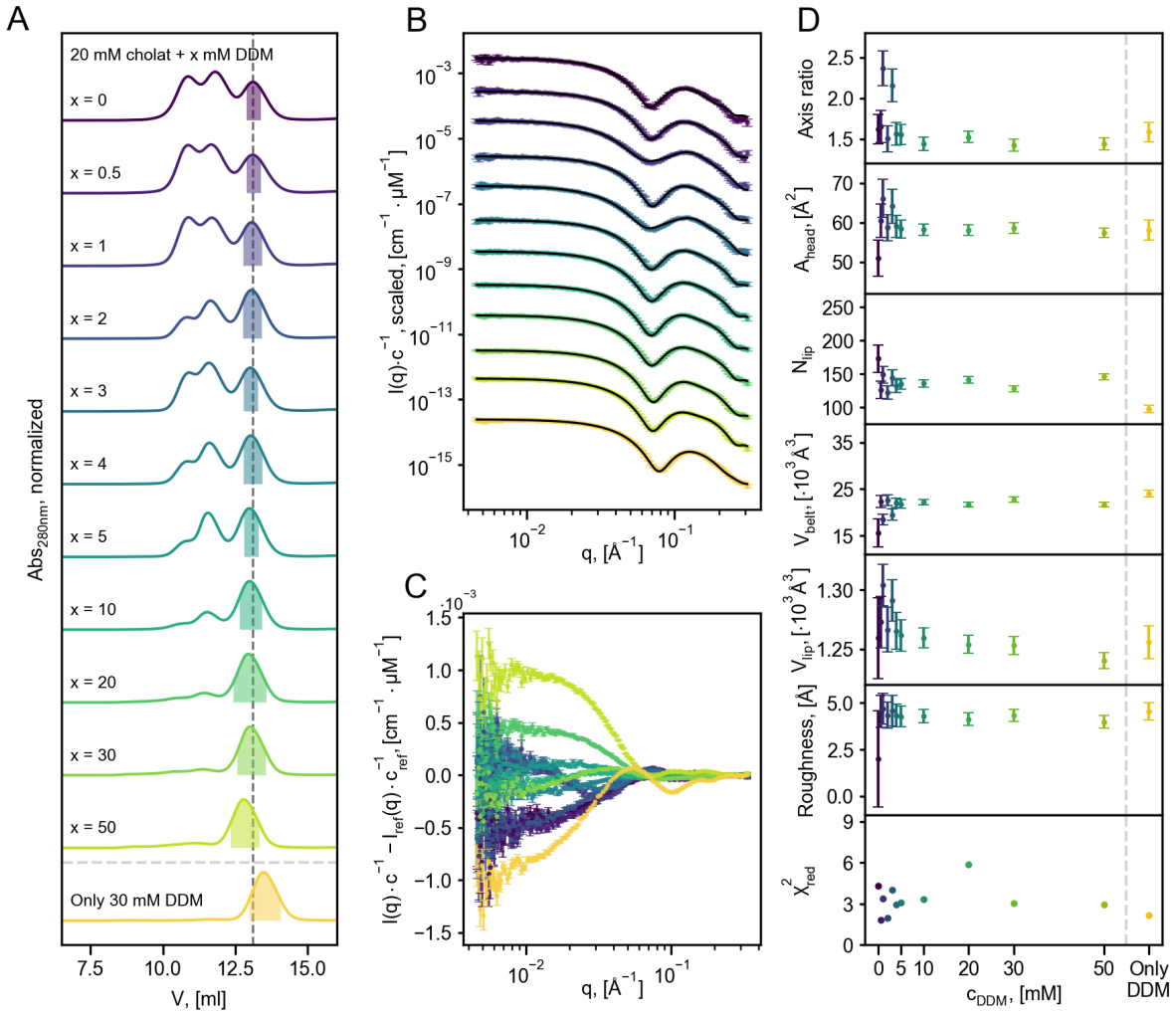


Figure 5.2: *Results from SEC and SAXS of lsMSP1D1ΔH5 NDs.* ***Panel A****: The SEC chromatograms of the samples with 20 mM cholate and different concentrations of DDM. The bottom SEC-curve is for the sample without cholate and only DDM, as separated from the others by the horizontal dotted grey line. Each SEC-chromatogram has been normalized with respect to the highest peak in the individual chromatograms, and thus displays the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The vertical dotted line passes through the maximum of the nanodisc peak for the sample without DDM (the top) and serves as a guide to the eye, for comparison of the nanodisc elution peaks. The fully coloured parts of the peaks are the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70], of the samples collected from the SEC-curves in panel A. The colours correspond to those in panel A, and the fits from WillItFit [70] have been marked in black. For each sample, the measured intensity I has been normalized with the concentration c of the respective sample, so the intensity has the unit $cm^{-1} \cdot \mu M^{-1}$. The intensity is plotted as function of the length of scattering vector q in $Å^{-1}$ on a double-logarithmic scale. The concentration was determined by measuring the absorption at 280 nm and using Beer-Lambert's law (eq. 3.1), as described in previous sections. Furthermore, the curves have been scaled with a factor 10^{-i+1}, where $i = 1, 2, .., 12$ is the number of the curve, so the top curve is scaled with 10^0, the second one from the top is scaled with 10^{-1}, and so forth. The curves have errorbars with rescaled errors, as estimated from Bayes App [9], albeit they are small and difficult to see.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and an average of all the scattering profiles except the one with only DDM.* ***Panel D****: Fit parameters and χ^2_{red} obtained from WillItFit as function of the DDM concentration, c_{DDM}, with the same colour scheme as with previous panels. To the far right is the sample with only DDM, as separated from the others by a grey dotted line. The parameters are described in section 4.2.5, and can be seen in table A.2 in appendix A.2.*

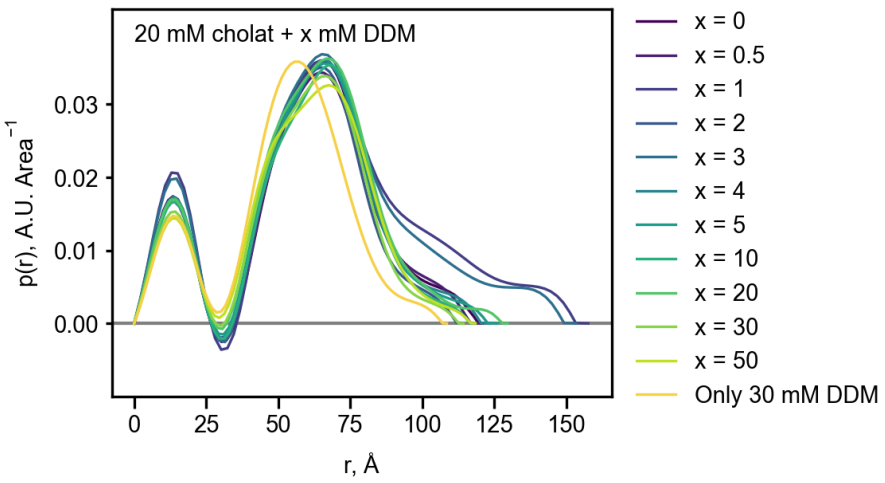


Figure 5.3: *The pair-distance distributions of lsMSP1D1ΔH5 NDs with cholate and varying concentrations of DDM. The p(r) functions have been normalized by their respective areas and the colour scheme is the same as in figure 5.2*

The pair-distance distributions *p(r)* of the lsMSP1D1ΔH5 ND samples as function of the distance r are shown in figure 5.3. All *p(r)* functions display the double-peak features of nanodiscs [61]. The sizes of the nanodiscs are given as the maximum distance r, and as seen from figure 5.3 the sizes vary considerably between the samples. Again, the samples with 1 mM DDM and 3 mM DDM deviate from the rest, as they display significantly larger discs. This could be due to the presence of larger particles.

The shape and position of the large peak at r ≈ 75 Å changes considerably for the sample with only DDM. Furthermore, this sample also displays a smaller size than the others.

5.3 csMSP1D1ΔH5

Figure 5.4 displays the results of csMSP1D1ΔH5 NDs in the same manner as with figure 5.2. All SEC-chromatograms in panel A display one distinct and narrow peak, indicating a monodisperse sample All samples with both cholate and DDM roughly have the same elution volumes indicating that the nanodiscs are of the same size. However, the peaks of these samples seem to become systematically narrower and more skewed with increasing concentration of DDM. The sample with only DDM has a peak shifted to a slightly larger elution volume, indicating smaller nanodiscs. This sample has a fairly symmetric albeit wider peak, indicating a wider size distribution of discs.

As with lsMSP1D1ΔH5 NDs, the SAXS curves for csMSP1D1ΔH5 NDs have the overall nanodisc features. The model fits reproduce most features but have slight troubles at the larger q-values. This is especially evident for the sample without DDM (the curve at the top). From the difference plot in panel C, there seems to be a systematic change in the SAXS-curves for the samples with cholate at lower q-values.

The determined parameters are quite similar for all samples, and χ^2_{red}s of 1.3-1.9 indicate excellent fits. The lipid volume V_{lip} seem to decrease slightly with increasing amounts of DDM for the samples with cholate, while the number of lipids per disc N_{lip} slightly increases. The number of lipids per nanodisc in the range 90-110 is smaller than the 120 estimated in the reconstitution mixture.

In figure 5.5 the pair-distance distributions *p(r)* of the samples can be seen as function of the distance r. The *p(r)* distributions are quite similar and exhibit almost the same nanodisc size D_{max} ≈ 100 Å.

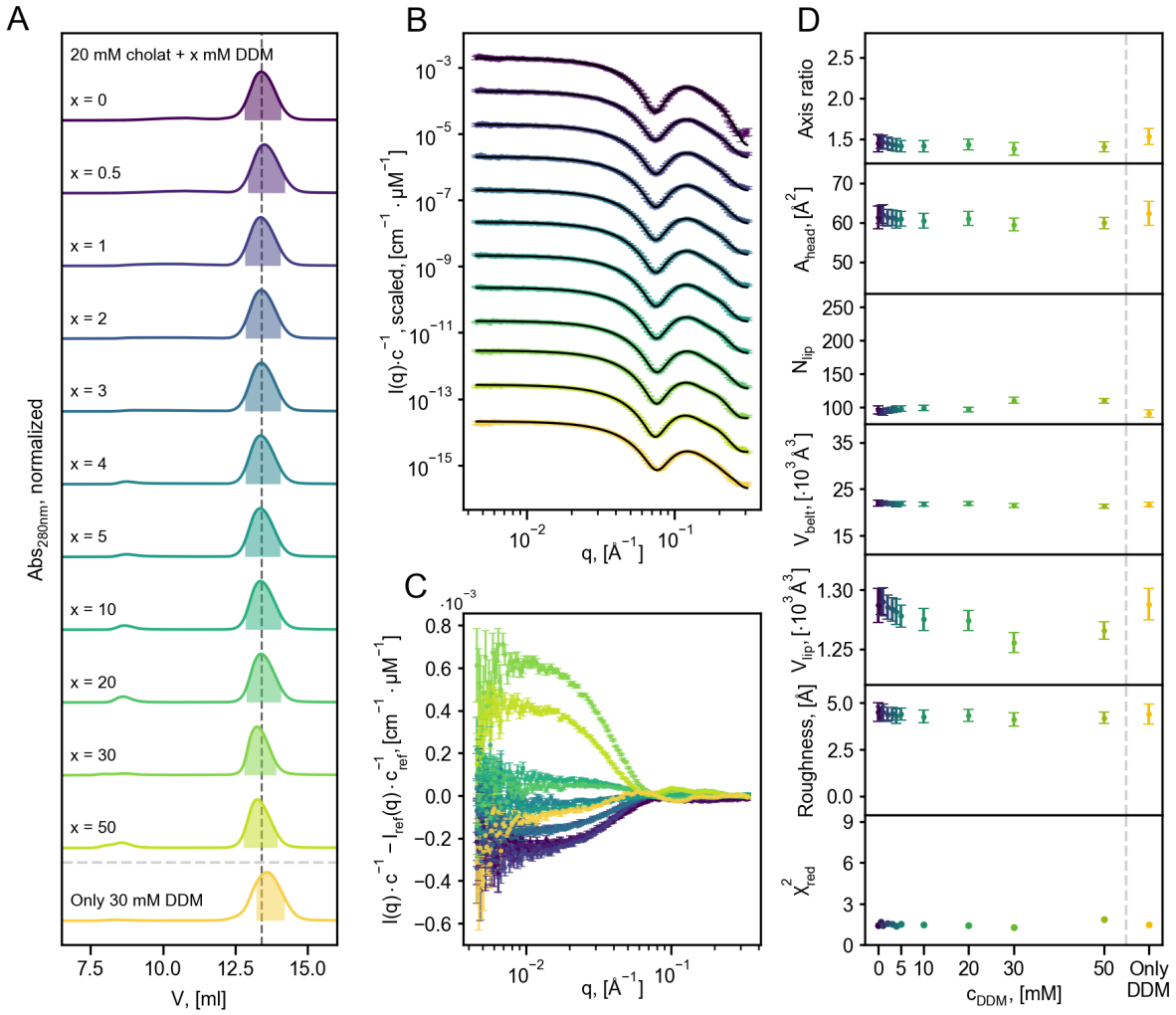


Figure 5.4: *Results from SEC and SAXS of csMSP1D1ΔH5 NDs.* ***Panel A****: The SEC chromatograms with the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The fully coloured parts of each peak display the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70] with the fits marked in black. The intensity I has been normalized with the concentration c and is plotted as function of the length of scattering vector q. The curves have been scaled with a factor 10^{-i+1}, where i=1,2,..,12 is the number of the curve.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and the average of all the scattering profiles except the one with only DDM.* ***Panel D****: parameters and χ^2_{red} with errorbars obtained from WillItFit [70] as function of the DDM concentration. The parameters can be seen in table A.2 in section A.2.*

For the samples with cholate, the shape of the big peak at $r \approx 65$ Å seems to change systematically from somewhat Gaussian-like to an asymmetric peak with a shoulder, while the sample with only DDM returns to the more symmetric and Gaussian-like peak.

5.4 lsMSP1D1

Fig. 5.6 displays the results from SEC and SAXS of lsMSP1D1 NDs. As with lsMSP1D1ΔH5 NDs, many of the SEC-chromatograms for lsMSP1D1 NDs display clear signs of polydispersity, as they have multiple elution peaks. For the samples in grey, with 0.5 mM DDM and 1 mM DDM, the nanodisc peaks were so

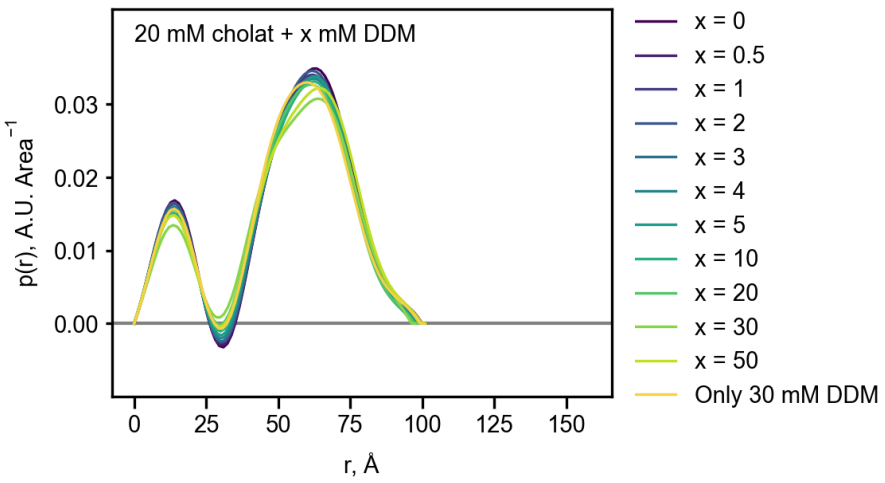


Figure 5.5: *The pair-distance distributions of csMSP1D1ΔH5 NDs with cholate and varying concentrations of DDM. Each distribution has been normalized with its area.*

small and contaminated with larger particles from the adjacent peaks, that it was not possible to collect fractions for further analyses.

From the SEC-chromatograms, it is evident that the nanodisc peak is diminished when introducing small amounts of DDM, compared to the sample 0 mM DDM. However, the nanodisc peak increases again when adding more DDM. This behaviour is contradicting to the behaviour of lsMSP1D1ΔH5 NDs.

The nanodisc peaks seem to shift to a slightly smaller elution volume, and thus a larger size, with increasing DDM concentrations. The sample with only DDM, however, seem to have almost the same elution volume as the sample without DDM. The sample with only DDM is the only sample to appear completely monodisperse, as all the other samples display some degree of polydispersity.

The SAXS curves in panel B have the characteristic nanodisc features, with fits that reproduce the features. Some of the SAXS, especially at lower concentrations of DDM exhibits slight curvature at low q-values, which could indicate imperfect buffer subtraction. The difference plot in panel D shows no systematic change in the SAXS curves.

The χ^2_{red} values indicate good fits. The refined parameters fluctuate slightly, especially at the lower DDM concentrations. This might be due to the polydispersity of the samples.

The determined number of lipids per disc fluctuate from 138-181. The number of lipids estimated in the reconstitution mix was 140 and only the sample with only DDM is smaller. This is the same trend as with the lsMSP1D1ΔH5 NDs.

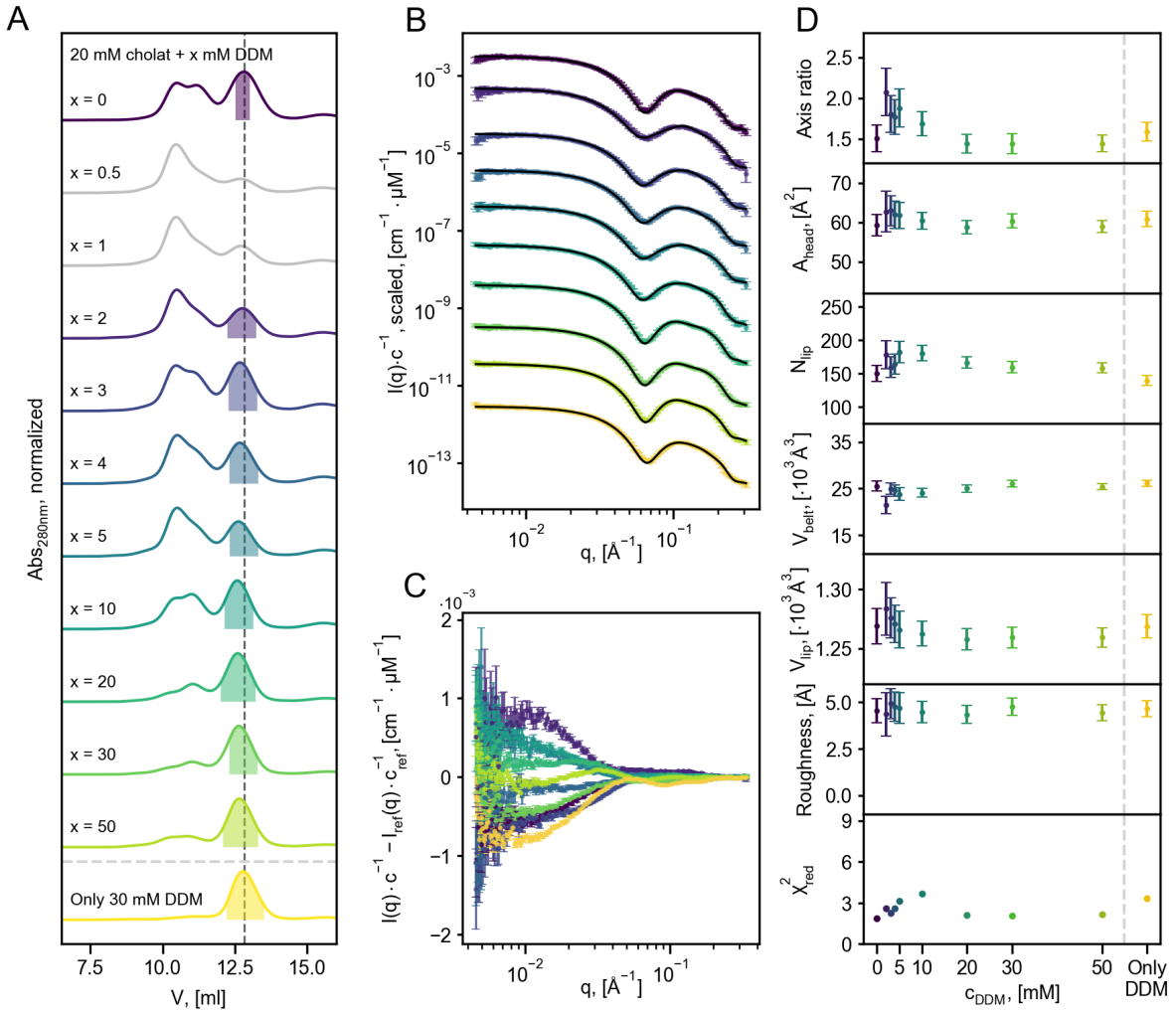


Figure 5.6: *Results from SEC and SAXS of lsMSP1D1 NDs.* ***Panel A****: The SEC chromatograms with the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The fully coloured parts of each peak display the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis. For the two samples containing 0.5 mM DDM and 1 mM DDM, marked in grey, the main nanodisc peaks were too small to extract any useful information, so they were omitted for further analyses.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70] with the fits marked in black. The intensity I has been normalized with the concentration c and is plotted as function of the length of scattering vector q. The curves have been scaled with a factor 10^{-i+1}, where i=1,2,..,10 is the number of the curve.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and the average of all the scattering profiles except the one with only DDM.* ***Panel D****: parameters and χ^2_{red} with errorbars obtained from WillItFit [70] as function of the DDM concentration. The parameters can be seen in table A.2 in section A.2.*

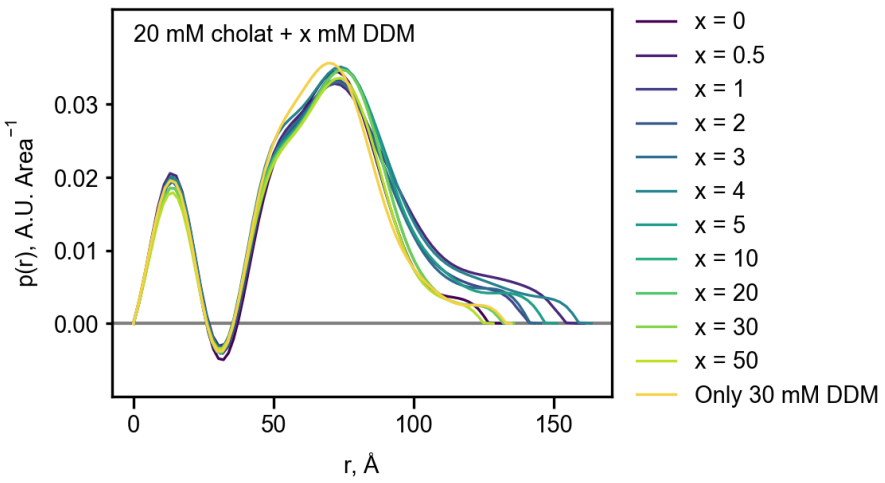


Figure 5.7: *The pair-distance distributions of lsMSP1D1 NDs with cholate and varying concentrations of DDM. Each distribution has been normalized with its area.*

The pair-distance distributions, figure 5.7, of the samples exhibits varying nanodisc sizes. There seem to be some degree of systematic trend, as the samples with smaller amounts of DDM exhibits larger sizes. This supports the perception that there could be some degree of polydispersity at lower DDM concentrations.

5.5 csMSP1D1

The SEC chromatograms of csMSP1D1 NDs are displayed in panel A in figure 5.8. The chromatograms all display distinct monodisperse peaks indicating formation of nanodiscs. For the samples with cholate, the curves become narrower and markedly more skewed with increasing DDM concentration. Furthermore, the elution peaks shift to a smaller volume with increasing DDM concentration. The sample with only DDM has a symmetric shape and almost the same elution volume as the sample without DDM.

The SAXS curves in panel B display nanodisc features and a flat region at small q-values indicating that there is no aggregation. The fitted models readily reproduces the trait of the curve, however, for the samples with cholate and large DDM concentrations the model slightly overshoots at large q-values.

The difference plot in panel C reveals differences at low q-values larger than for the other nanodiscs and a systematic change.

Even though the determined χ^2_{red} indicate good fits, the parameters for csMSP1D1 NDs show odd behaviours. The lipid volume V_{lip} systematically varies, as it decreases with increasing concentration of DDM until 30 mM DDM, and then it increases again. For many of the other types of nanodiscs, the lipid volume was more or less constant. The number of lipids display the opposite behaviour, for the samples with 20, 30, and 50 mM DDM, the number of lipids is too large; it simply is not physically possible to fit that many lipids in such a small disc.

One suggestion to the reason for the behaviour could be whether larger particles were present. However, neither SEC chromatograms nor the SAXS data show any signs of aggregates. Aggregation can appear as upturns in curvature at low q-regions of SAXS data, but as this was not the case for any of the samples.

As nothing appeared to be wrong with the samples beside the great differences in SAXS curves at low q-values, various changes in the fitting routine was tried as a means to "solve" the odd behaviour.

At first, the lipid volume was fixed to 1275 $Å^3$ [33], to "force" the number of lipids down, but as the fitting parameters are correlated, it would just induce fluctuations and differences in other parameters. Then fitting was tried with a scaling parameter, which scales the fitted curve and allows for adjustment of the concentration, in case the concentration determination had been off. But it did not reduce the number of lipids enough or account for the behaviour. The roughness and the belt volume were kept fixed on separate occasions, and the belt volume together with the lipid volume at another, but alas, nothing seemed to do the trick. Every time, the new fit would be worse than the initial fit, and the model would deviate from the scattering curve. This indicates that something is going on at higher DDM concentrations that the model cannot account for.

As far as it is known, this systematic scattering pattern has not been detected before, and it was not possible to find a conclusive reason for the odd behaviour.

The *p(r)* distributions in figure 5.9 show that the sizes systematically follow the same pattern as the particles parameters where it gets larger and then smaller again with increasing DDM concentration. The shape of the large peak at r $\approx$ 75 Å also changes from somewhat Gaussian to more asymmetric. The sample with only DDM is the smallest of size and has an almost Gaussian peak as well.

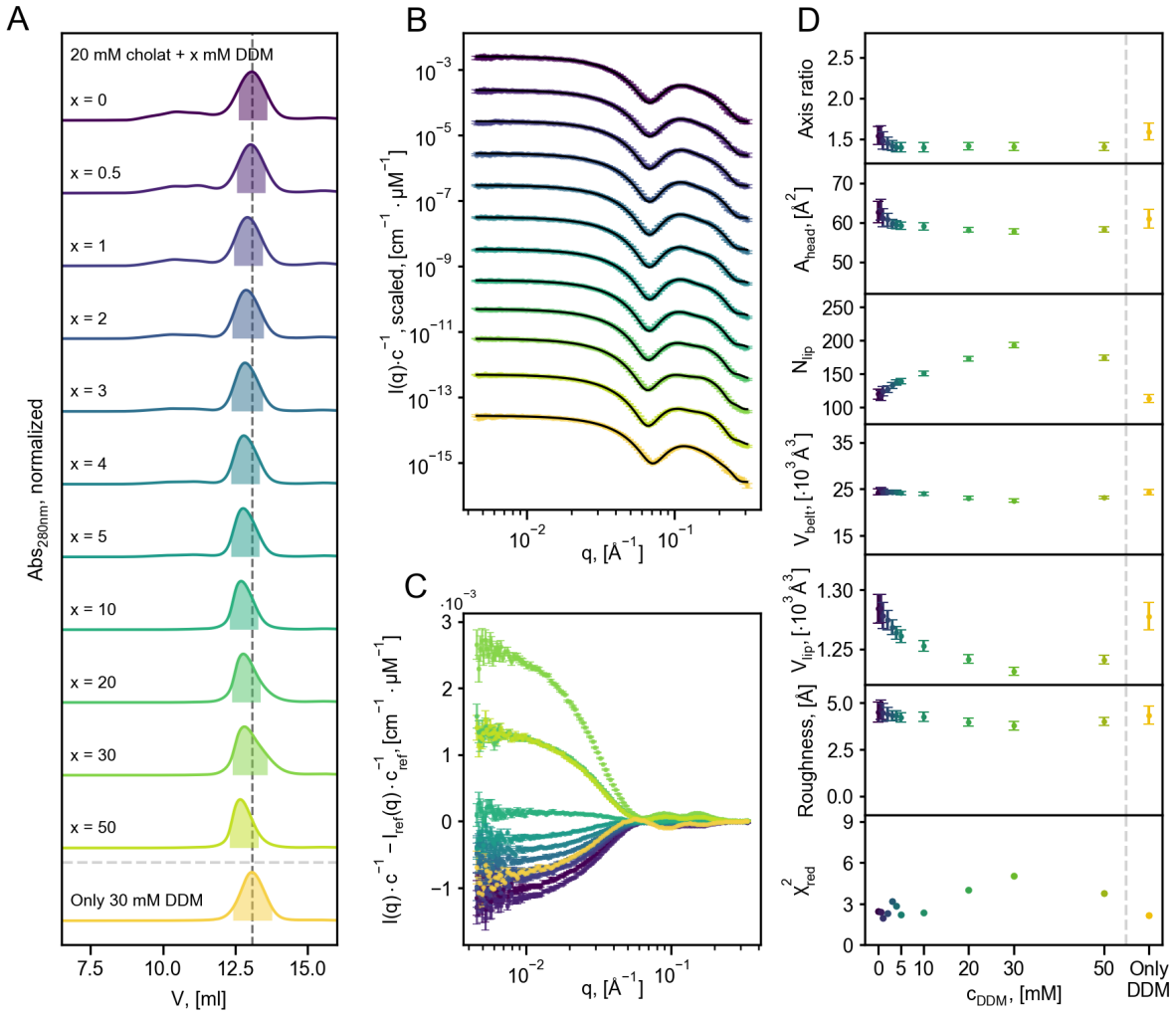


Figure 5.8: *Results from SEC and SAXS of csMSP1D1 NDs.* ***Panel A****: The SEC chromatograms with the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The fully coloured parts of each peak display the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70] with the fits marked in black. The intensity I has been normalized with the concentration c and is plotted as function of the length of scattering vector q. The curves have been scaled with a factor 10^{-i+1}, where i=1,2,..,12 is the number of the curve.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and the average of all the scattering profiles except the one with only DDM.* ***Panel D****: parameters and χ^2_{red} with errorbars obtained from WillItFit [70] as function of the DDM concentration. The parameters can be seen in table A.2 in section A.2.*

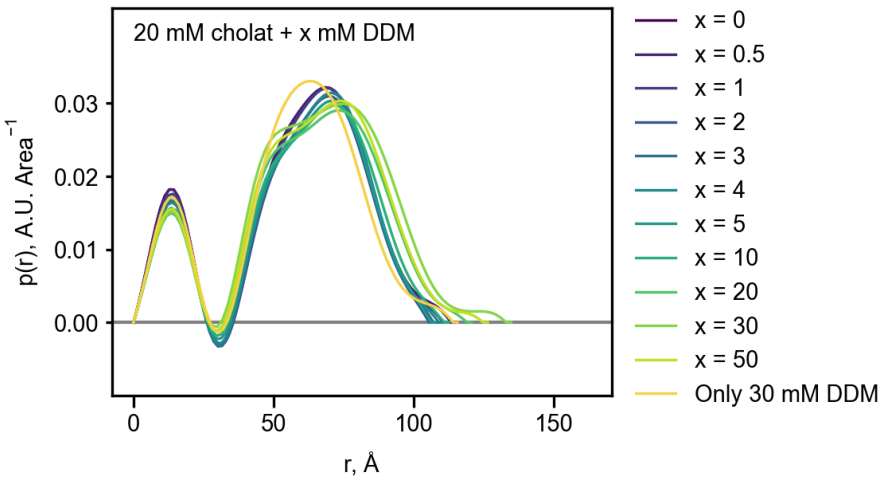


Figure 5.9: *The pair-distance distributions of csMSP1D1 NDs with cholate and varying concentrations of DDM. Each distribution has been normalized with its area.*

5.6 lsMSP1E3D1

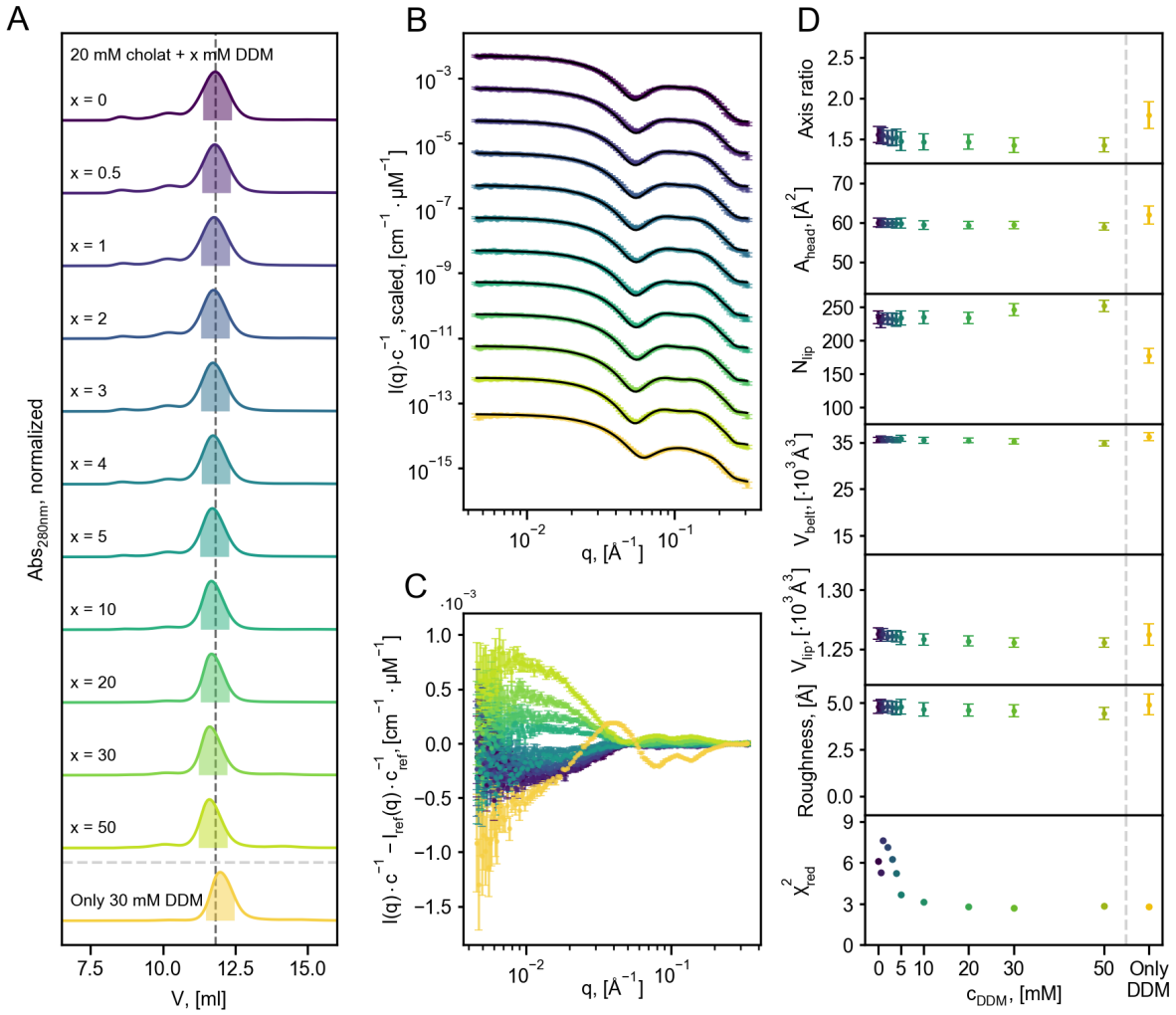


Figure 5.10: *Results from SEC and SAXS of lsMSP1D1E3 NDs.* ***Panel A****: The SEC chromatograms with the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The fully coloured parts of each peak display the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70] with the fits marked in black. The intensity I has been normalized with the concentration c and is plotted as function of the length of scattering vector q. The curves have been scaled with a factor 10^{-i+1}, where i=1,2,..,12 is the number of the curve.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and the average of all the scattering profiles except the one with only DDM.* ***Panel D****: parameters and χ^2_{red} with errorbars obtained from WillItFit [70] as function of the DDM concentration. The parameters can be seen in table A.2 in section A.2.*

The SEC-chromatograms of lsMSP1D1E3 NDs in panel A in figure 5.10 all display a single distinct peak, indicating monodispersity. At low DDM concentrations, there are slight peaks of larger particles. The peaks seem to become visibly more narrower and more skewed with increasing DDM concentrations. The SAXS curves in panel B exhibit nanodisc features and the SAXS curves vary systematically from the reference as seen in panel C. The sample with only DDM is noticeably different from the others and display larger differences at higher q-values than the others. χ^2_{red} is large at small DDM concentrations indicating worse fits, but it decreases with higher DDM concentrations.

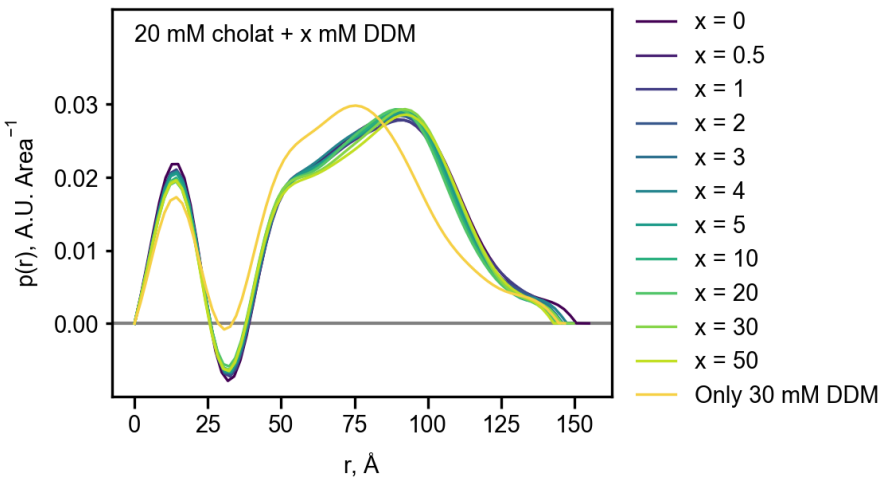


Figure 5.11: *The pair-distance distributions of lsMSP1D1E3 NDs with cholate and varying concentrations of DDM. Each distribution has been normalized with its area.*

The *p(r)* distributions in figure 5.11 display the same size of approximately 150 Å, and the samples with cholate follows the same curvature. However, the sample with only DDM has a rather different shape than the others, even though it has the same size.

5.7 csMSP1E3D1

The SEC-chromatograms in panel A in figure 5.12 are all very monodisperse. The nanodisc peaks develop a larger degree of skewness with an increasing amount of DDM. The peak moves to slightly larger volumes indicating larger nanodiscs. The SAXS curves in panel B exhibit nanodisc features. The sample with only DDM does not exhibit the double bump to the same degree as the other samples. This is also evident in the difference plot in panel C where the sample with only DDM is noticeably different than the other samples. The rest of the samples express systematic changes in differences.

The fit parameters vary slightly, especially at lower DDM concentrations. These samples with lower concentrations of DDM also exhibits larger χ^2_{red}s indicating worse fits.

The *p(r)* functions for csMSP1E3D1 NDs have the same size but the sample with only DDM is very different in shape.

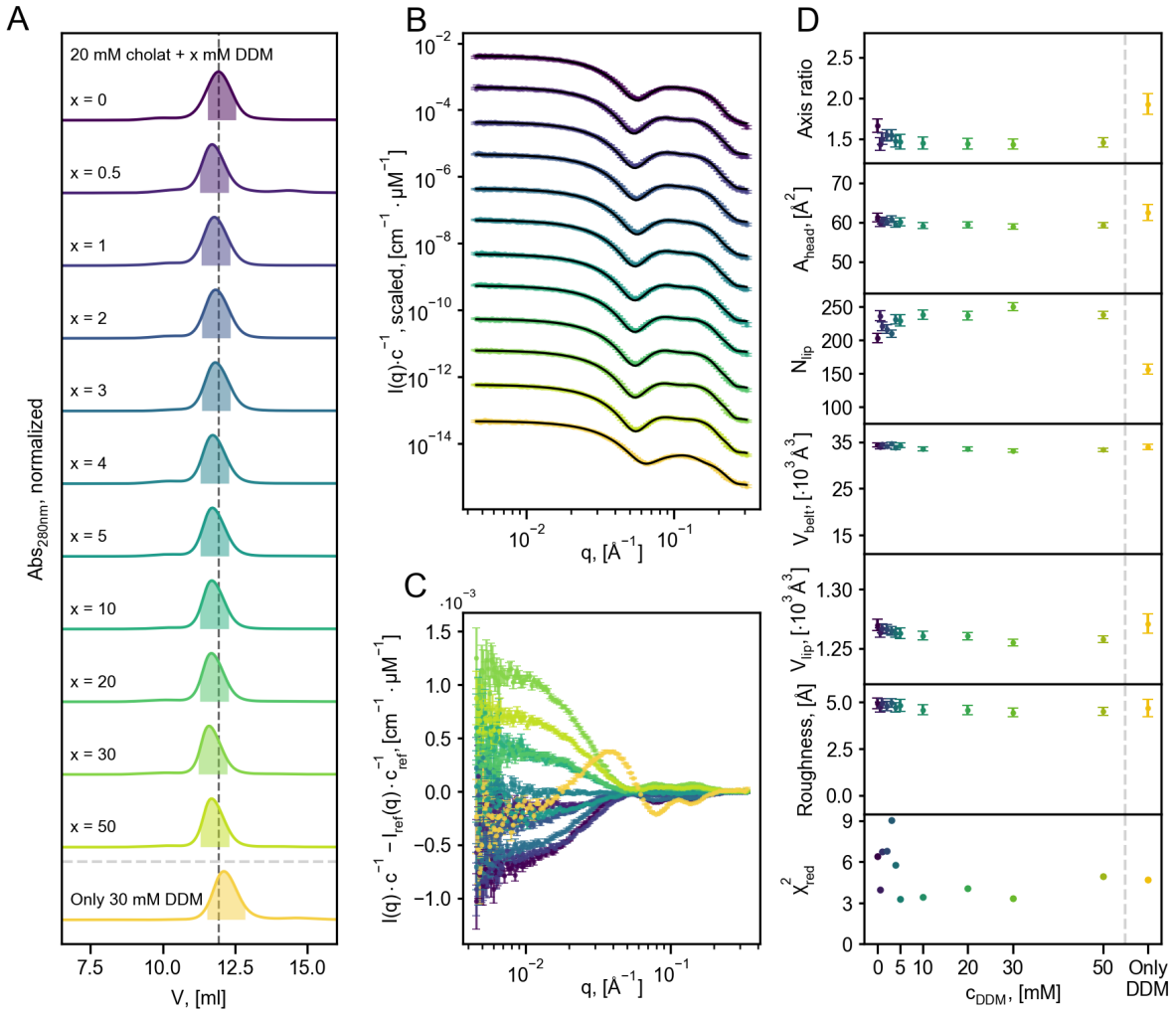


Figure 5.12: *Results from SEC and SAXS of csMSP1D1E3.* ***Panel A****: The SEC chromatograms with the normalized absorption at 280 nm, Abs_{280nm}, as function of the elution volume V in mL. The fully coloured parts of each peak display the collected volumes used for further analysis by SAXS, concentration determination and phosphate analysis.* ***Panel B****: SAXS curves and fits obtained using WillItFit [70] with the fits marked in black. The intensity I has been normalized with the concentration c and is plotted as function of the length of scattering vector q. The curves have been scaled with a factor 10^{-i+1}, where i=1,2,..,12 is the number of the curve.* ***Panel C****: Difference plots; the difference between the scattering profiles in panel B and the average of all the scattering profiles except the one with only DDM.* ***Panel D****: parameters and χ^2_{red} with errorbars obtained from WillItFit [70] as function of the DDM concentration. The parameters can be seen in table A.2 in section A.2.*

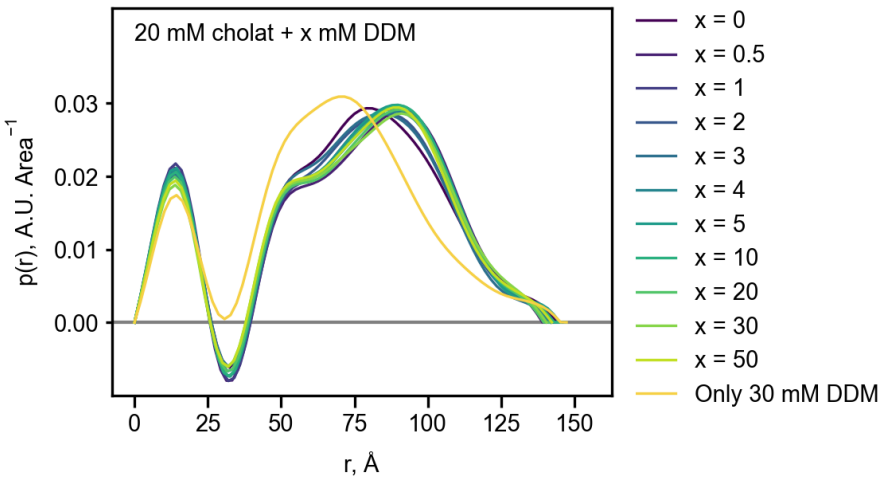


Figure 5.13: *The pair-distance distributions of csMSP1D1E3 with cholate and varying concentrations of DDM. Each distribution has been normalized with its area.*

5.7.1 Phosphate analysis

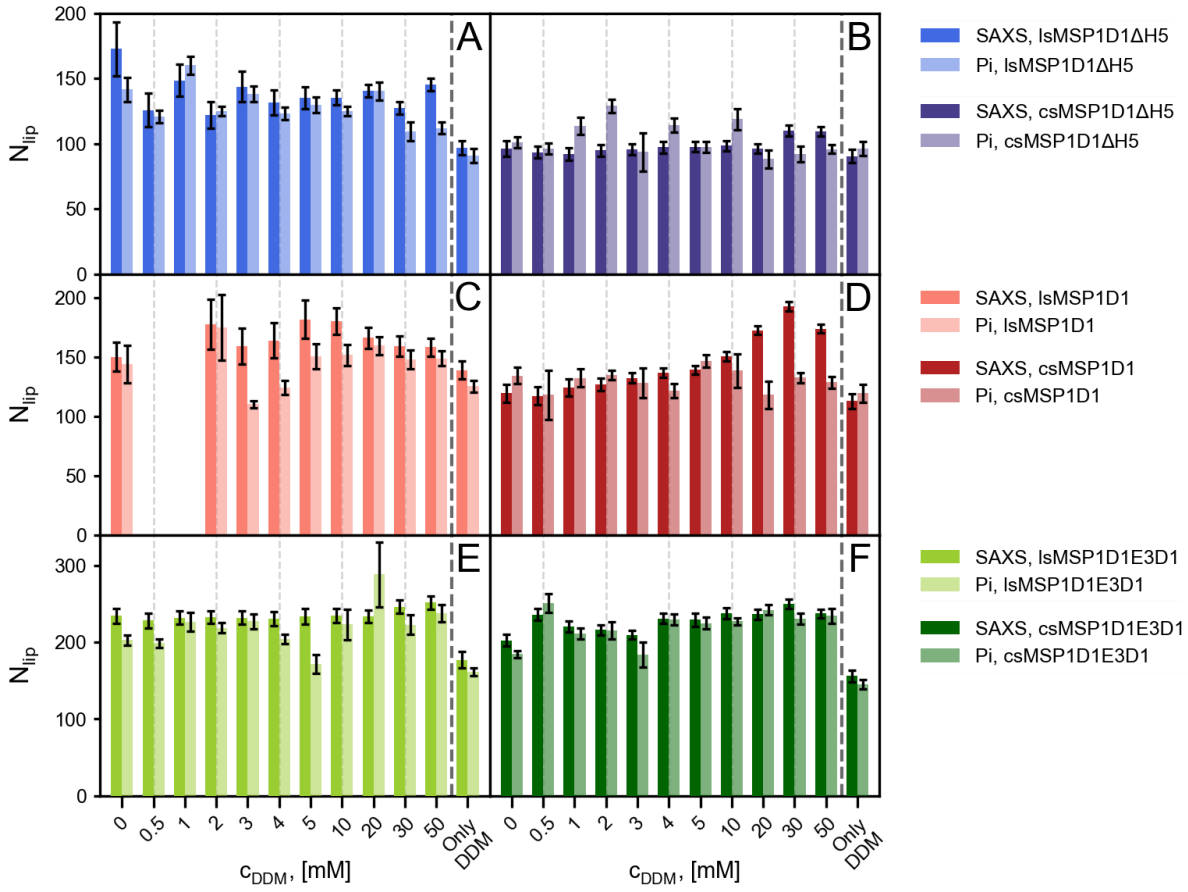


Figure 5.14: *Bar chart of the number of lipids per nanodisc, N_{lip}, determined from SAXS (WillItFit) and phosphate analysis (Pi) as function of the concentration of DDM, c_{DDM}. **A**: lsMSP1D1ΔH5 ND **B**: csMSP1D1ΔH5 ND **C**: lsMSP1D1 ND **D**: csMSP1D1 ND **E**: lsMSP1E3D1 ND **F**: csMSP1E3D1 ND*

The phosphate analysis was conducted as an independent measure of the number of lipids per disc. It was determined as previously described in the Methods section and data analysis section, the results are presented in figure 5.14. Panel A through F shows the results from each MSP; the number of lipids N_{lip} from the phosphate analysis (denoted Pi in the figure) along with the number of lipids determined from the SAXS experiments, as a function of the concentration of DDM c_{DDM}. As mentioned before, there is no data for lsMSP1D1 ND with 0.5 mM and 1 mM DDM, as the nanodisc peaks from SEC were too small and contaminated to extract any useful information from.

The results show that some, but not all, samples agree on the number of lipids determined from the two methods. There is no systematic pattern to the disagreement, as it occurs for all nanodiscs at different DDM concentrations, and it is not the same method yielding the largest number every time. The errors on the smaller linear discs (lsMSP1D1ΔH5 ND and lsMSP1D1 ND) generally seem to be larger, but it could be a coincidence.

Three independent measurements of phosphate were made for each sample, to obtain better statistics. For some samples, e.g. lsMSP1D1E3 ND with 20 mM DDM, one value was strikingly different from the other two and would shift the average. These data points could be outliers, but with only three data points in each series, it was not deemed acceptable to exclude any data points on that basis.

The phosphate analysis was conducted over multiple days, and one of the days, the determined number of lipids for all samples were markedly larger than the number determined by SAXS, thus indicating a systematic error. It was decided to exclude the data from that day, and repeat the phosphate analysis. The initial results can be seen in figure A.1 in appendix section A.1.

The phosphate analysis is sensitive to the amount of phosphate, and previous use of phosphate analysis have shown varying degrees of success with the method. There are multiple sources of error; unclean glass tubes with phosphate remnants from previous phosphate analyzes could lead to overestimation of the number of lipids, while bad pipetting or if the sample from SEC was not thoroughly mixed before extracting sample for phosphate analysis could lead to over- or underestimation.

It is noticeable that the phosphate analysis systematically estimates a markedly lower number of lipids for the three samples of csMSP1D1 NDs with 20, 30, and 50 mM DDM in panel D. These are the same samples that displayed odd behaviours in the SAXS analysis. The individual phosphate measurements for these samples are in agreement, so the discrepancy is striking and questions the number of lipids determined by SAXS.

5.8 General discussion

5.8.1 Effect of length of MSP

As expected, the results show that the longer the MSP, the larger the maximum distance D_{max} and the larger the number of lipids, N_{lip}, per nanodisc. This is consistent with literature [15] . Furthermore, the length of the MSP affects the monodispersity of the formed nanodiscs. From the SEC-chromatograms it is evident that of all the linearized discs, lsMDP1D1E3 ND displayed the largest degree of monodispersity, even without DDM. This is probably due to the larger degree of flexibility in the larger discs.

5.8.2 Effect of circularization

From the SEC-chromatograms it is evident that circularization of the MSPs ensures a larger degree of monodispersity of the formed discs, which is in agreement with previous findings [47]. Especially for the smaller nanodiscs (ls- and cs MSPΔH5 NDs and ls- and csMSP1D1 NDs), there seem to be some improvement. The SEC chromatograms of the nanodisc reconstituted with linear MSPs display polydispersity, while the circularized versions are monodisperse. The circularized nanodiscs contain fewer lipids and are slightly smaller than their linear counterparts. This is in line with previous findings of circularized and solubility enhanced nanodiscs examined by SAXS [33].

5.8.3 Effect of DDM

The overall effect of adding DDM during the reconstitution is increasing monodispersity, as seen in the SEC chromatograms. This is particularly evident for lsMSP1D1ΔH5 NDs and lsMSP1D1 NDs, as these nanodiscs display many peaks corresponding to different particles in the SEC-chromatograms at low DDM concentrations, but the monodispersity increases with addition DDM. However, while the monodispersity for lsMSP1D1ΔH5 NDs only increases when adding DDM, it initially decreases before being restored for lsMSP1D1.

For all nanodiscs, the sample with only DDM displays a smaller number of lipids per disc and a larger axis ratio. This is consistent with previous findings suggesting that nanodiscs assembled with only DDM contain fewer lipids and are more elliptical [61]. The effect is most evident in the two largest discs, ls- and csMSP1D1E3 ND. For all nanodiscs, the *p(r)* distributions of the samples with only DDM also

demonstrate rather different structures, as the shape of the distributions deviates from the other samples with cholate. Furthermore, with the exception of lsMSP1D1 ND, the minima in the *p(r)* distributions at r ≈ 30 Å is less negative for the samples with only DDM, supporting that these samples contain fewer lipids. The contrast between the scattering length densities of the lipid tails and the lipid heads, and the lipid tails and the MSPs, cause negative values in the *p(r)*. Therefore, the growing value of the minima indicates a smaller contribution from the lipids.

This is also in line with the samples with only DDM containing precipitants as mentioned in the Methods section. These precipitants were probably different lipid aggregates, that had not been incorporated in the discs, resulting in lipid-poor nanodiscs.

As previously determined [61], the structure and composition of the preaggregates have a great influence on the structure of the formed discs. The solubility enhanced MSPs have more negative charges than conventional MSPs and it is plausible that the negative charges of the MSPs clash with the negative charges of cholate, thus affecting the ability to reconstitute. This could be the reason for the observed polydispersity in some samples when reconstituted the conventional way with only cholate. The large linear discs lsMSPE3D1 NDs are more flexible and to a great extent can overcome the clash of charges. Meanwhile, the two other linear discs are smaller and more constrained, and maybe they are more likely to fuse to form larger particles with more than two MSPs. The circularized nanodiscs have the termini of the MSPs covalently linked and cannot fuse to form larger particles in the same way the linear nanodiscs can.

For samples of mixed cholate and DDM during reconstitution, the stoichiometry of the preaggregates depends on both the concentration of cholate and the concentration of DDM. At low DDM concentrations, cholate dominates the reconstitution process [61], but by increasing the amount of DDM, the stoichiometry of the preaggregates slowly changes. As DDM is not charged, an explanation for the increase in monodispersity when adding DDM in the reconstitution process might be that DDM can "dilute" charges in the preaggregate and to some degree prevents charge repulsion.

5.8.4 SEC and Gaussians

The nanodisc peaks from the SEC chromatograms were fitted to exponentially modified Gaussians (EMGs), with the goal of determining the area ratio between the area of the main nanodisc peak A_{main} and the area of the entire SEC-curve A_{tot} to use as a measure for the degree of monodispersity. The fitting was done using the curve fitting tool from the scipy package in Python. The samples with polydispersity were fitted to a sum of two Gaussians and one EMG; the nanodisc peaks were assumed to follow the EMG, while the peaks of the larger particles were assumed to be Gaussians. The choice of using Gaussians to describe the peaks of the larger particles instead of EMGs was to limit the number of fitting parameters, and it was estimated that it would not have any significant impact on the determined area. The SEC chromatograms with the area of the nanodisc peaks from the EMG fits for is displayed in figure 5.15 panel A for lsMSP1D1ΔH5 NDs. All fits can be seen in the appendix in A.3.

As evident from the previous figures the size of the main nanodisc peak increases with increasing concentration of DDM for most nanodiscs. Thus, the area ratio also increases as a function of DDM until the area ratio reaches a maximum value. All the other samples seem to exhibit asymptotic behaviour. For the samples with cholate, the area ratios were fitted to an asymptotic function, to evaluate and compare the behaviours of the different discs with increasing amounts of DDM. The fitted equation is similar to that of Michaelis-Menten kinetics, which is based on saturation in enzymatic reactions [4], or in this case; when enough DDM is added, the sample will be monodisperse. The equation for reaching a maximum

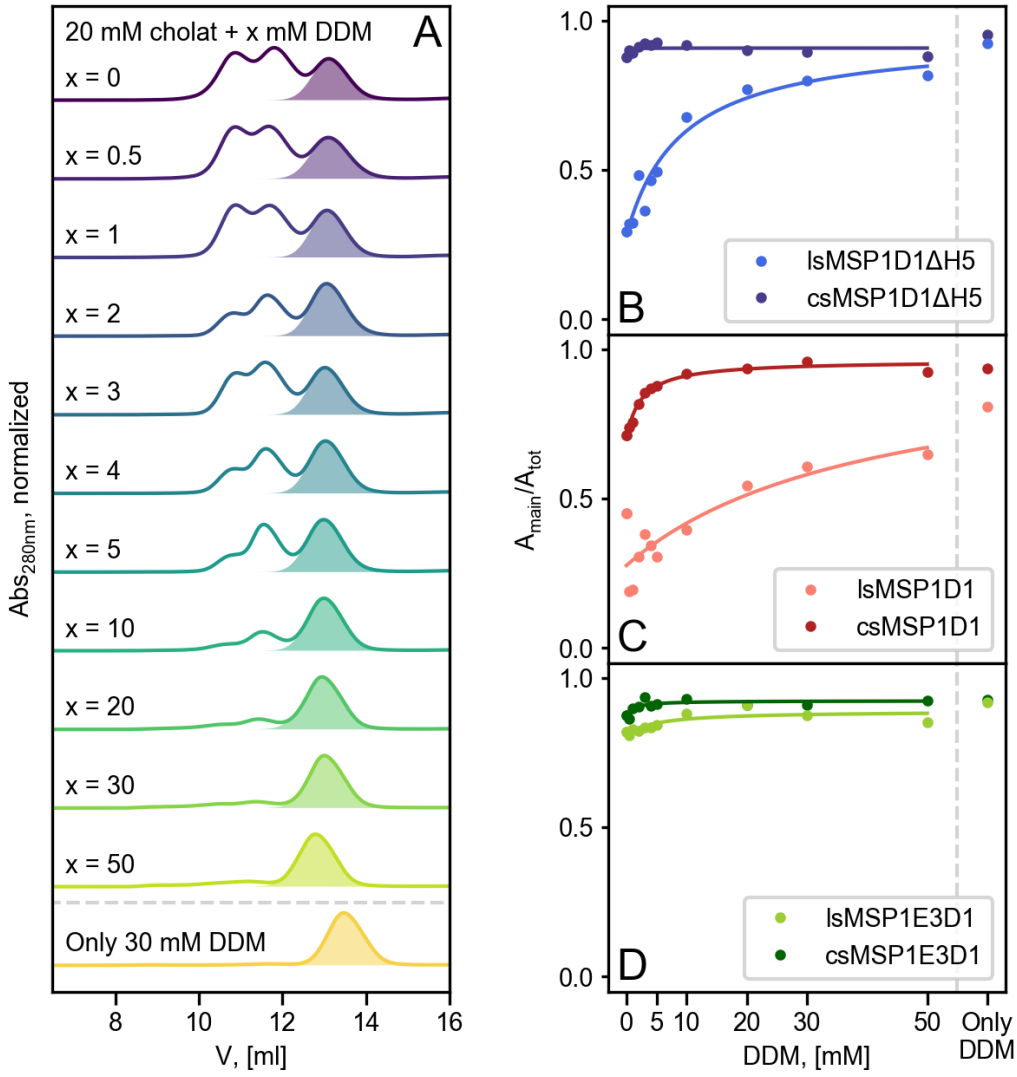


Figure 5.15: ***Panel A****: Fits of EMGs to the nanodisc peaks from SEC of lsMSP1D1ΔH5.* ***Panel B-D****: area ratio $^{A_{main}}/_{A_{tot}}$ between the area the main nanodisc peak and the total area of the SEC curve as function of the DDM concentration c_{DDM}.*

state of monodispersity with increasing DDM concentration is given as:

$$\frac{A_{main}}{A_{tot}} = \frac{(A_{max} - A_0) \cdot c_{DDM}}{K_{DDM} + c_{DDM}} + A_0 \tag{5.1}$$

where $^{A_{main}}/_{A_{tot}}$ is the area ratio between the area of the main nanodisc peak A_{main} and the total area of the SEC curve A_{tot}, A_{max} is the maximum possible area ratio, A_0 is the initial area ratio without DDM, c_{DDM} is the concentration of DDM, and K_{DDM} is a constant, similar to the Michaelis constant, determining the DDM concentration at which the area ratio is halfway between the initial and maximum obtainable area.

The fitted functions are presented in figure 5.15 panel B-D and the parameters can be seen in table 5.1.

The circularized nanodiscs have smaller K_{DDM}s as these samples already are quite monodisperse,

	A_{max}	A_0	K_{DDM}, [mM]
lsMSP1D1ΔH5	0.95	0.28	9.11
csMSP1D1ΔH5	0.91	0.86	0.08
lsMSP1D1	0.99	0.28	40.64
csMSP1D1	0.96	0.70	2.45
lsMSP1E3D1	0.89	0.81	5.00
csMSP1E3D1	0.92	0.87	1.04

Table 5.1: *Fit parameters from the fits of eq. 5.1 to the area ratios,* $^{A_{main}}/_{A_{tot}}$.

even without introducing DDM. Meanwhile, the lsMSP1D1 NDs would need a concentration of more than 110 mM to reach 80% monodispersity as suggested by the equation 5.1. From the fitted parameters, it can be seen that the "worst" maximal area ratio is for lsMSP1D1E3 NDs. However, this nanodisc also has a large initial area ratio and it decreases slightly at larger DDM concentrations.

The equation is useful for a quantitative overview of the effect of DDM, but it is reasonable to believe that it will break down eventually. The equation assumes the concentration of DDM can increase infinitely and eventually all nanodiscs will reach the maximum area ratio, but this assumption is probably wrong.

The concentration of DDM will presumably reach a state where it affects the reconstitution process to a great extent and prevents the formation of nanodiscs. There will probably be a point where the effect of DDM exceeds the effect of cholate and as with the samples with only DDM, the possibility of formation of lipid poor discs. Furthermore, one would also need additionally many detergent absorbing beads to remove the excessive amounts of DDM. But the equation could be useful to evaluate the effects of DDM on monodispersity at DDM concentrations below 50 mM.

5.8.5 Skewness of nanodisc peaks

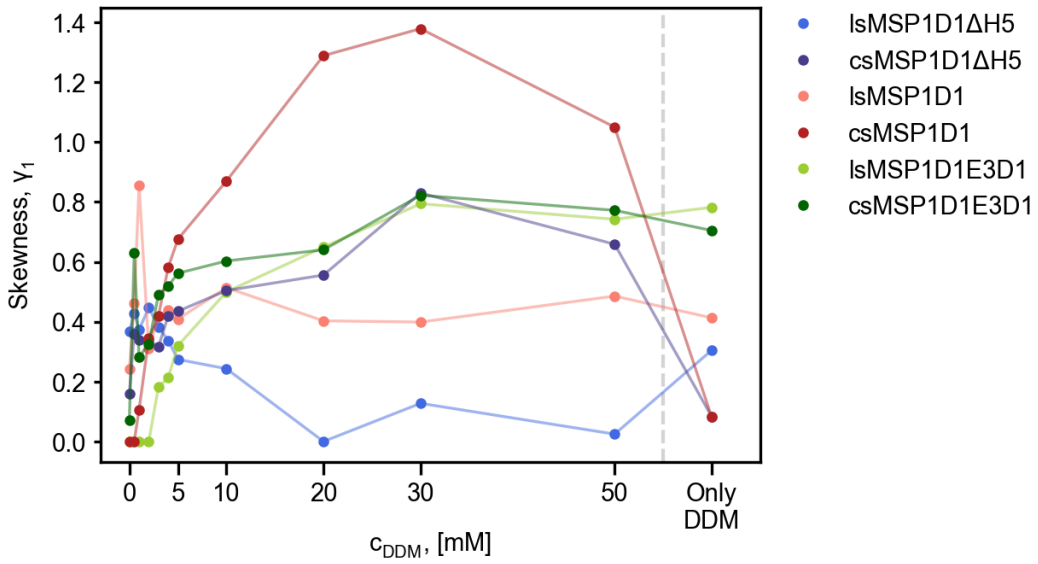


Figure 5.16: *Skewness,* γ_1*, of the EMGs as function of the DDM concentration.*

Out of curiosity, the skewness γ_1 of the EMGs was plotted as a function of the DDM concentration

in figure 5.16. The skewness is a measure of the asymmetry of the peaks and also an indication of the degree of tailing. Surprisingly, the skewness of the csMSP1D1 NDs seem to follow the same trend as the SAXS data; the skewness increases from 0 for the sample without DDM to almost 1.4 for the sample with 30 mM DDM and decreases again for the largest DDM concentrations.

It should be emphasized that as absorption in the SEC chromatograms is not measured with errorbars, the goodness of fit of the EMGs can only be assessed visually. Thus the skewness will be a rough estimate and the results cannot stand alone. However, they could indicate that something happens to the csMSP1D1 ND samples already during SEC. However, it is unknown what that "something" is. As all SEC experiments were performed under the same conditions it is odd that the skewnesses should be so different.

The two smallest discs with linear MSPs, lsMSP1D1ΔH5 and lsMSP1D1, were the ones most difficult to fit, due to the presence of the multiples peaks, so I would estimate these samples to have the poorest fits.

5.9 Proteorhodhopsin

After experimenting with the effect of DDM on empty nanodiscs, the effect was evaluated when integrating the transmembrane protein proteorhodopsin (PR). The samples were handled as described in the method section, and for each MSP, a sample with and a sample without 30 mM DDM were mixed and separated with SEC. As PR is almost entirely embedded within the lipid bilayer, and does not protrude much, integration of PR does not have a significant effect on the elution volume of PR-loaded discs compared to that of empty discs. As PR has an absorption peak at 530 nm, where the MSPs do not absorb, measurements of absorption at 530 nm simultaneously with 280 nm enables monitoring of nanodiscs with PR only.

The resulting SEC-chromatograms from absorption at 530 nm are presented in figure 5.17. To compare the amount of incorporated PR, the chromatograms have been normalized with respect to the sample with DDM, for each type of nanodisc. So for instance is lsMSP1D1ΔH5 without DDM normalized with respect to the height of the peak for lsMSP1D1ΔH5 with DDM.

For PR-nanodiscs reconstituted the conventional way without DDM, it is evident from figure 5.17 that larger particles have formed for all samples, as every chromatogram has peaks at smaller volumes than expected for nanodiscs. There is hardly any absorption for csMSP1D1ΔH5 ND at 530 nm with only a tiny signal at V≈8.5 mL, indicating that all PR has aggregated in larger particles. Absorption at 280 nm confirms that there indeed are nanodisc peaks for all samples, however, without integration of PR in csMSP1D1ΔH5 ND. The chromatograms for absorption at 280 nm can be seen in figure A.8 in the appendix section A.4.

Without DDM, the samples with the two largest discs, lsMSP1E3D1 ND and csMSP1E3D1 ND, are the only ones displaying slight peaks at the expected nanodisc elution volumes of V≈11.5 mL. Albeit, for both samples the peaks overlap to a great extend. Furthermore, csMSP1E3D1 ND has a small bump at V≈13.5 mL indicating some kind of PR-containing particle, that is smaller than a nanodisc.

The use of DDM changes the SEC-chromatograms substantially. LsMSP1D1ΔH5 ND and lsMSP1D1 ND demonstrate almost the same pattern; a large peak at V≈9.6 mL indicating larger particles and a small expected nanodisc peak at V≈12 mL. CsMSP1D1ΔH5 ND has a small peak at V≈10.5 mL and a large nanodisc peak at V≈13.2 mL. CsMSP1D1 ND has a chromatogram similar to that of csMSP1D1ΔH5 ND, however, the peak of the larger particles is taller and not as separated from the nanodisc peak as csMSP1D1ΔH5 ND. LsMSP1E3D1 ND displays a peak at V≈9.7 mL that has merged with the nanodisc

peak at V≈10.8 mL to form an almost indistinguishable double peak. CsMSP1E3D1 ND displays an almost perfect monodisperse nanodisc peak at V≈11.5 mL with only a slight shoulder to the left indicating a tiny amount of larger particles.

When dissolved in detergent PR will exist as both monomers and oligomers [36]. The general perception from previous studies has been that only PR monomers fit inside nanodiscs. Meanwhile, the oligomers would integrate in larger particles, estimating oligomeric PR as to large to fit in a nanodisc [19, 32].

The results for csMSP1E3D1 ND in figure 5.17 suggests that all PR has readily been incorporated in nanodiscs. The other circularized discs, csMSP1D1ΔH5 ND and csMSP1D1 ND, display a large degree of PR in discs, and minor parts of larger particles. As the elution volumes of the main nanodisc peaks match those of the empty discs, it is a clear indication that the addition of DDM increases the ability to incorporate PR in nanodiscs.

It is not possible from the SEC-chromatograms alone to confirm whether the integrated PR is of monomeric or oligomeric structure, as it might be possible that the oligomers have fallen apart and are integrated as monomers only. However, all PR have been integrated in csMSP1E3D1 NDs, so there is an indication that oligomers have been incorporated as well. This result determines a great improvement in the ability to integrate PR when DDM is added during the reconstitution process.

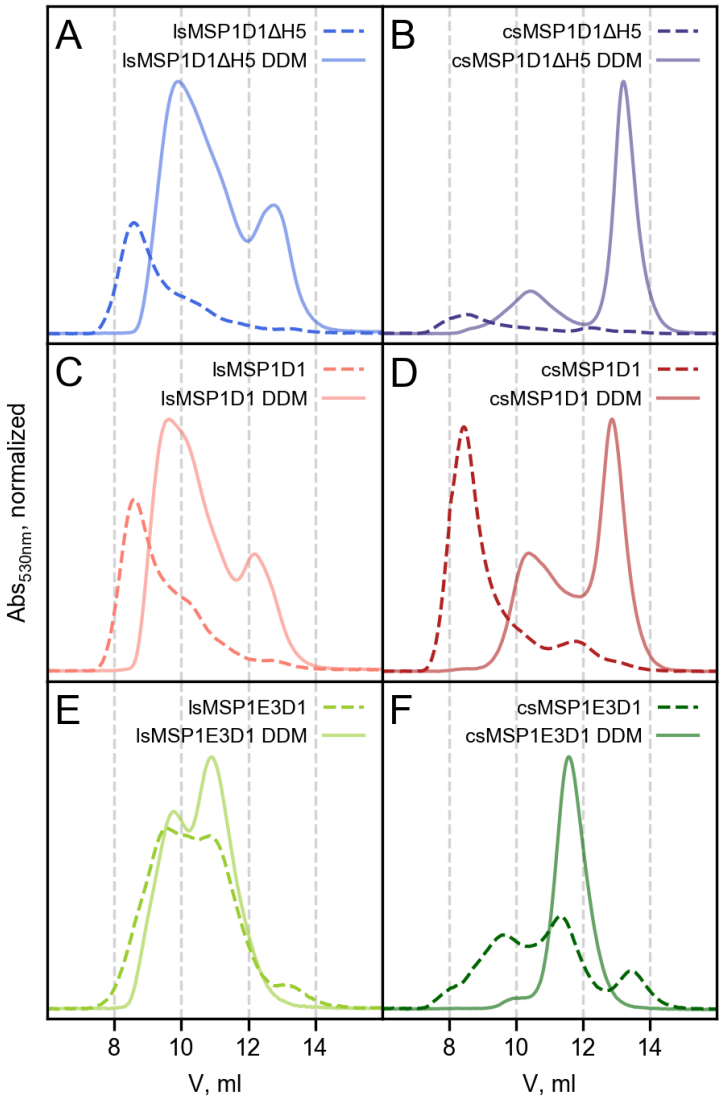


Figure 5.17: *SEC chromatograms measured at 530 nm of nanodiscs reconstituted with proteorhodopsin (PR), and with and without 30 mM DDM.* ***A****: lsMSP1D1ΔH5* ***B****: csMSP1D1ΔH5* ***C****: lsMSP1D1* ***D****: csMSP1D1* ***E****: lsMSP1E3D1* ***F****: csMSP1E3D1*

6 Conclusion and perspective

In this project, it was systematically investigated how the length of the MSP, the concentration of DDM, and circularization affects the structure of solubility enhanced nanodiscs by using size exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS).

It was found that circularization increases the monodispersity of the formed nanodiscs, but the nanodiscs are slightly smaller and contain slightly fewer lipids than their linear counterparts. The length of the MSP determines the size of the formed nanodiscs and affects monodispersity as well. Adding DDM in the reconstitution process induces monodispersity for nanodisc samples that displayed polydispersity when reconstituted with conventional methods. However, DDM alone yields highly elliptical and lipid-poor discs.

The csMSP1D1 NDs displayed systematically odd behaviour in the SAXS data, as the refined parameters estimated an immense number of lipids should be integrated in the discs. The strange behaviour of csMSP1D1 NDs is not explained, so it would be advisable to repeat the SAXS measurements for that nanodisc, as this could maybe shine a light on whether it was a systematic error or some unknown effect.

The incorporation of oligomeric proteorhodopsin (PR) showed an increase of PR integration in nanodiscs when adding 30 mM DDM to the reconstitution process. As implied by SEC measurements it was even possible to integrate a large fraction of PR in the smallest disc csMSP1D1ΔH5 ND. Furthermore, a completely monodisperse sample was obtained for the larger disc csMSP1D1E3 ND, suggesting that PR oligomers has also been integrated. However, it is not possible from SEC alone to determine whether the integrated PR is of oligomeric or monomeric state.

The effect of DDM on the monodispersity of the formed nanodiscs, and improving integration of PR, has the potential for integration of other and more unstable membrane proteins in nanodiscs. It could be interesting to investigate reconstitution with DDM and less cholate than the conventional 20 mM with the aim of integrating unstable membrane proteins, that do not tolerate the standard conditions under which nanodiscs are formed.

Furthermore, it should be confirmed whether or not oligomeric PR has been integrated in the nanodiscs, perhaps by doing SAXS measurement of the nanodiscs containing PR.

7 References

[1] *18: Amino Acids, Proteins, and Enzymes*. http://saylordotorg.github.io/text_the-basics-of-general-organic-and-biological-chemistry/s21-amino-acids-proteins-and-enzym.html. [Last accessed: February 14^{th}, 2021].

[2] *2L6X: Solution NMR Structure of Proteorhodopsin*. https://www.rcsb.org/structure/2L6X. [Last accessed: February 15^{th}, 2021].

[3] Bruce Alberts. "Chapter 10: Membran Structure". In: *Molecular Biology of the Cell*. 6th ed. Taylor & Francis Group, 2014. Chap. 10, pp. 565–596. DOI: 10.1038/216864a0.

[4] Bruce Alberts. "Chapter 3: Proteins". In: *Molecular Biology of the Cell*. 6th ed. Taylor & Francis Group, 2014. Chap. 3, pp. 109–172. DOI: 10.1088/0031-9112/24/12/030.

[5] Jens Als-Nielsen and Des McMorrow. *Elements of Modern X-ray Physics: Second Edition*. 2011. ISBN: 9780470973950. DOI: 10.1002/9781119998365.

[6] Christian Bamann et al. "Proteorhodopsin". In: *Biochimica et Biophysica Acta - Bioenergetics* 1837.5 (2014), pp. 614–625. ISSN: 18792650. DOI: 10.1016/j.bbabio.2013.09.010.

[7] Timothy H. Bayburt, Yelena V. Grinkova, and Stephen G. Sligar. "Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins". In: *Nano Letters* 2.8 (2002), pp. 853–856. ISSN: 15306984. DOI: 10.1021/nl025623k.

[8] Timothy H. Bayburt and Stephen G. Sligar. "Membrane protein assembly into Nanodiscs". In: *FEBS Letters* 584.9 (2010), pp. 1721–1727. ISSN: 00145793. DOI: 10.1016/j.febslet.2009.10.024. URL: http://dx.doi.org/10.1016/j.febslet.2009.10.024.

[9] *Bayes App*. https://somo.chem.utk.edu/bayesapp/. [Last accessed: February 4^{th}, 2021].

[10] Stefan Bibow et al. "Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I". In: *Nature Structural and Molecular Biology* 24.2 (2017), pp. 187–193. ISSN: 15459985. DOI: 10.1038/nsmb.3345.

[11] J. P. Busnel et al. "Investigation and interpretation of band broadening in size exclusion chromatography". In: *Journal of Chromatography A* 930.1-2 (2001), pp. 61–71. ISSN: 00219673. DOI: 10.1016/S0021-9673(01)01159-1.

[12] Nathan P. Cowieson et al. "Beamline B21: High-throughput small-angle X-ray scattering at Diamond Light Source". In: *Journal of Synchrotron Radiation* 27 (2020), pp. 1438–1446. ISSN: 16005775. DOI: 10.1107/S1600577520009960.

[13] Stefano Da Vela and Dmitri I. Svergun. "Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution". In: *Current Research in Structural Biology* 2.August (2020), pp. 164–170. ISSN: 2665928X. DOI: 10.1016/j.crstbi.2020.08.004. URL: https://doi.org/10.1016/j.crstbi.2020.08.004.

[14] P. Debye and A. M. Bueche. "Scattering by an inhomogeneous solid". In: *Journal of Applied Physics* 20.6 (1949), pp. 518–525. ISSN: 00218979. DOI: 10.1063/1.1698419.

[15] I. G. Denisov et al. "Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size". In: *Journal of the American Chemical Society* 126.11 (2004), pp. 3477–3487. ISSN: 00027863. DOI: 10.1021/ja0393574.

[16] *Diamond Light Source: status and perspectives*. https://royalsocietypublishing.org/doi/10.1098/rsta.2013.0161. [Last accessed: March 13^{th}, 2021].

[17] *Equipment Description*. https://www.diamond.ac.uk/Instruments/Soft-Condensed-Matter/small-angle/B21/description.html. [Last accessed: March 6th, 2021].

[18] *Experimental noise in small-angle scattering can be assessed and corrected using the Bayesian Indirect Fourier Transformation*. https://arxiv.org/abs/2012.04247. [Last accessed: March 25th, 2021].

[19] Srividya Ganapathy et al. "Membrane matters: The impact of a nanodisc-bilayer or a detergent microenvironment on the properties of two eubacterial rhodopsins". In: *Biochimica et Biophysica Acta - Biomembranes* 1862.2 (2020), p. 183113. ISSN: 18792642. DOI: 10.1016/j.bbamem.2019.183113. URL: https://doi.org/10.1016/j.bbamem.2019.183113.

[20] P. Garidel et al. "Thermodynamic characterization of bile salt aggregation as a function of temperature and ionic strength using isothermal titration calorimetry". In: *Langmuir* 16.12 (2000), pp. 5267–5275. ISSN: 07437463. DOI: 10.1021/la9912390.

[21] Elisabeth Gasteiger et al. "Protein Identification and Analysis Tools on the ExPASy Server". In: *The Proteomics Protocols Handbook*. Ed. by John M. Walker. Humana Press Inc., 2005. Chap. 52, pp. 571–608. DOI: 10.1385/1592598900.

[22] Otto Glatter. "A New Method for the Evaluation of Small-Angle Scattering Data". In: *J. Appl. Cryst.* 10? (1977), pp. 415–421.

[23] A. Golubev. "Exponentially Modified Peak Functions in Biomedical Sciences and Related Disciplines". In: *Computational and Mathematical Methods in Medicine* 2017 (2017). ISSN: 17486718. DOI: 10.1155/2017/7925106.

[24] Pontus Gourdon et al. "Optimized in vitro and in vivo expression of proteorhodopsin: A seven-transmembrane proton pump". In: *Protein Expression and Purification* 58.1 (2008), pp. 103–113. ISSN: 10465928. DOI: 10.1016/j.pep.2007.10.017.

[25] Eli Grushka. "Characterization of Exponentially Modified Gaussian Peaks in Chromatography". In: *Analytical Chemistry* 44.11 (1972), pp. 1733–1738. ISSN: 15206882. DOI: 10.1021/ac60319a011.

[26] Steen Hansen. "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering". In: *Journal of Applied Crystallography* 33.6 (2000), pp. 1415–1421. ISSN: 00218898. DOI: 10.1107/S0021889800012930.

[27] Steen Hansen. "BayesApp: A web site for indirect transformation of small-angle scattering data". In: *Journal of Applied Crystallography* 45.3 (2012), pp. 566–567. ISSN: 00218898. DOI: 10.1107/S0021889812014318.

[28] Annegret Hildebrand et al. "Temperature dependence of the interaction of cholate and deoxycholate with fluid model membranes and their solubilization into mixed micelles". In: *Colloids and Surfaces B: Biointerfaces* 32.4 (2003), pp. 335–351. ISSN: 09277765. DOI: 10.1016/j.colsurfb.2003.08.001.

[29] *Injection system and linac*. https://www.diamond.ac.uk/Science/Machine/Components/linac.html. [Last accessed: March 16th, 2021].

[30] Jacob N. Israelachvili. *Part III: Self-Assembling Structures and Biological Systems*. 2011, pp. 503–534. ISBN: 978-0-12-391927-4.

[31] A J De Jesus, H Yin, and United States. *Supramolecular Membrane Chemistry*. Second Edi. Vol. 5. Elsevier, 2017, pp. 311–328. ISBN: 9780124095472. DOI: 10.1016/B978-0-12-409547-2.12572-7. URL: http://dx.doi.org/10.1016/B978-0-12-409547-2.12572-7.

[32] Nicolai Tidemand Johansen et al. "Introducing SEC–SANS for studies of complex self-organized biological systems". In: *Acta Crystallographica Section D: Structural Biology* 74.12 (2018), pp. 1178–1191. ISSN: 20597983. DOI: 10.1107/S2059798318007180.

[33] Nicolai Tidemand Johansen et al. "Circularized and solubility-enhanced MSPs facilitate simple and high-yield production of stable nanodiscs for studies of membrane proteins in solution". In: *FEBS Journal* 286.9 (2019), pp. 1734–1751. ISSN: 17424658. DOI: 10.1111/febs.14766.

[34] Yuri Kalambet et al. "Reconstruction of chromatographic peaks using the exponentially modified Gaussian function". In: *Journal of Chemometrics* 25.7 (2011), pp. 352–356. ISSN: 08869383. DOI: 10.1002/cem.1343.

[35] Toshimitsu Kawate and Eric Gouaux. "Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins". In: *Structure* 14.4 (2006), pp. 673–681. ISSN: 09692126. DOI: 10.1016/j.str.2006.01.013.

[36] Adriana L. Klyszejko et al. "Folding and Assembly of Proteorhodopsin". In: *Journal of Molecular Biology* 376.1 (2008), pp. 35–41. ISSN: 00222836. DOI: 10.1016/j.jmb.2007.11.030.

[37] Anders Krogh et al. "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes". In: *Journal of Molecular Biology* 305.3 (2001), pp. 567–580. ISSN: 00222836. DOI: 10.1006/jmbi.2000.4315.

[38] Norbert Kučerka, Mu Ping Nieh, and John Katsaras. "Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature". In: *Biochimica et Biophysica Acta - Biomembranes* 1808.11 (2011), pp. 2761–2771. ISSN: 00052736. DOI: 10.1016/j.bbamem.2011.07.022.

[39] Andreas Haahr Larsen and Martin Cramer Pedersen. "Experimental noise in small-angle scattering can be assessed and corrected using the Bayesian Indirect Fourier Transformation". In: (2020), pp. 1–13. URL: http://arxiv.org/abs/2012.04247.

[40] S. Le Vent. "Simulation of chromatographic peaks by simple functions". In: *Analytica Chimica Acta* 312.3 (1995), pp. 263–270. ISSN: 00032670. DOI: 10.1016/0003-2670(95)00220-T.

[41] Anthony G. Lee. "How lipids affect the activities of integral membrane proteins". In: *Biochimica et Biophysica Acta - Biomembranes* 1666.1-2 (2004), pp. 62–87. ISSN: 00052736. DOI: 10.1016/j.bbamem.2004.05.012.

[42] Kim Lefmann. "Neutron Scattering: Theory, Instrumentation, and Simulation". In: (2012).

[43] Thomas G. Mayerhöfer and Jürgen Popp. "Beer's law derived from electromagnetic theory". In: *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy* 215 (2019), pp. 345–347. ISSN: 13861425. DOI: 10.1016/j.saa.2019.02.103. URL: https://doi.org/10.1016/j.saa.2019.02.103.

[44] *Micelle*. [Last accessed: February 16^{th}, 2021]. URL: https://en.wikipedia.org/wiki/Micelle#/media/File:Phospholipids_aqueous_solution_structures.svg.

[45] *mpstruc: Membrane Proteins of Known 3D Structure*. https://blanco.biomol.uci.edu/mpstruc/. [Last accessed: March 25^{th}, 2021].

[46] *n-Dodecyl β-D-maltoside*. https://www.sigmaaldrich.com/catalog/product/sigma/d4641?lang=en&region=DK. [Last accessed: March 9^{th}, 2021].

[47] Mahmoud L. Nasr et al. "Covalently circularized nanodiscs for studying membrane proteins and viral entry". In: *Nature Methods* 14.1 (2016), pp. 49–52. ISSN: 15487105. DOI: 10.1038/nmeth.4079.

[48] James E. Noble. *Quantification of protein concentration using UV absorbance and coomassie dyes*. 1st ed. Vol. 536. Elsevier Inc., 2014, pp. 17–26. ISBN: 9780124200708. DOI: 10.1016/B978-0-12-420070-8.00002-7. URL: http://dx.doi.org/10.1016/B978-0-12-420070-8.00002-7.

[49] Doris Orthaber, Alexander Bergmann, and Otto Glatter. "SAXS experiments on absolute scale with Kratky systems using water as a secondary standard". In: *Journal of Applied Crystallography* 33.2 (2000), pp. 218–225. ISSN: 00218898. DOI: 10.1107/S0021889899015216.

[50] Joanne L. Parker and Simon Newstead. "Current trends in α-helical membrane protein crystallization: An update". In: *Protein Science* 21.9 (2012), pp. 1358–1365. ISSN: 09618368. DOI: 10.1002/pro.2122.

[51] *PDB Statistics: Protein-only Structures Released Per Year*. https://www.rcsb.org/stats/growth/growth-protein. [Last accessed: March 25^{th}, 2021].

[52] Jan Skov Pedersen. "Chapter 16: Modelling of Small-Angle Scattering Data from Colloids and Polymer Systems". In: *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. Ed. by Thomas Zemb and Peter Lindner. Elsevier Science B.V, 2002, pp. 391–420. ISBN: 978-0-444-51122-5.

[53] Martin Cramer Pedersen, Lise Arleth, and Kell Mortensen. "WillItFit: A framework for fitting of constrained models to small-angle scattering data". In: *Journal of Applied Crystallography* 46.6 (2013), pp. 1894–1898. ISSN: 00218898. DOI: 10.1107/S0021889813026022.

[54] *Phase Transition Temperatures for Glycerophospholipids*. https://avantilipids.com/tech-support/physical-properties/phase-transition-temps. [Last accessed: February 11^{th}, 2021].

[55] *Polypeptide Chain (Molecular Biology)*. http://what-when-how.com/molecular-biology/polypeptide-chain-molecular-biology/. [Last accessed: February 14^{th}, 2021].

[56] *ProtParam*. https://web.expasy.org/protparam/. [Last accessed: February 4^{th}, 2021].

[57] Jean Louis Rigaud et al. "Detergent removal by non-polar polystyrene beads: Applications to membrane protein reconstitution and two-dimensional crystallization". In: *European Biophysics Journal* 27.4 (1998), pp. 305–319. ISSN: 01757571. DOI: 10.1007/s002490050138.

[58] Aiman Sadaf et al. *Amphipathic agents for membrane protein study*. 1st ed. Vol. 557. Elsevier Inc., 2015, pp. 57–94. DOI: 10.1016/bs.mie.2014.12.021. URL: http://dx.doi.org/10.1016/bs.mie.2014.12.021.

[59] Nicholas Skar-Gislinge and Lise Arleth. "Small-angle scattering from phospholipid nanodiscs: Derivation and refinement of a molecular constrained analytical model form factor". In: *Physical Chemistry Chemical Physics* 13.8 (2011), pp. 3161–3170. ISSN: 14639076. DOI: 10.1039/c0cp01074j.

[60] Nicholas Skar-Gislinge et al. "Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the lipids and the protein". In: *Journal of the American Chemical Society* 132.39 (2010), pp. 13713–13722. ISSN: 00027863. DOI: 10.1021/ja1030613.

[61] Nicholas Skar-Gislinge et al. "Comprehensive Study of the Self-Assembly of Phospholipid Nanodiscs: What Determines Their Shape and Stoichiometry?" In: *Langmuir* 34.42 (2018), pp. 12569–12582. ISSN: 15205827. DOI: 10.1021/acs.langmuir.8b01503.

[62] *Structural Classification of Membrane Proteins*. https://www.creative-biolabs.com/blog/index.php/membrane-protein-overview/. [Last accessed: February 15^{th}, 2021].

[63] W.-J. Sun et al. "Order and disorder in fully hydrated unoriented bilayers of gel-phase dipalmitoylphosphatidylcholine". In: *Phys. Rev. E.* 49.5 (1994), pp. 4665–4676. DOI: 10.5840/socphiltoday1991555.

[64] Dmitri I. Svergun et al. *Chapter 4: Monodisperse systems*. May. 2013.

[65] Dmitri I. Svergun et al. *Chapter 6: Static structural studies*. Vol. 1. May. 2013, pp. 171–219. DOI: 10.1093/acprof:oso/9780199639533.003.0007.

[66] Dmitri I. Svergun et al. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Ed. by J. Bernstein et al. 1st ed. May. Oxford University Press, 2013. ISBN: 978–0–19–963953–3.

[67] Charles Tanford. "Micelle shape and size". In: *Journal of Physical Chemistry* 76.21 (1972), pp. 3020–3024. ISSN: 00223654. DOI: 10.1021/j100665a018.

[68] *Tips & Tricks GPC/ SEC: What Are the Differences Between GPC, SEC, and GFC, and How Do You Get Started with the Technique?* https://alfresco-static-files.s3.amazonaws.com/alfresco_images/pharma/2018/10/23/2daed5ee-8a69-46e1-9e12-a305965c30bc/figure%201%20L.png. [Last accessed: February 16^{th}, 2021].

[69] Saul R. Trevino, J. Martin Scholtz, and C. Nick Pace. "Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa". In: *Journal of Molecular Biology* 366.2 (2007), pp. 449–460. ISSN: 00222836. DOI: 10.1016/j.jmb.2006.10.026.

[70] *WillItFit?* https://sourceforge.net/projects/willitfit/. [Last accessed: February 4^{th}, 2021].

A Appendix

A.1 First phosphate analysis

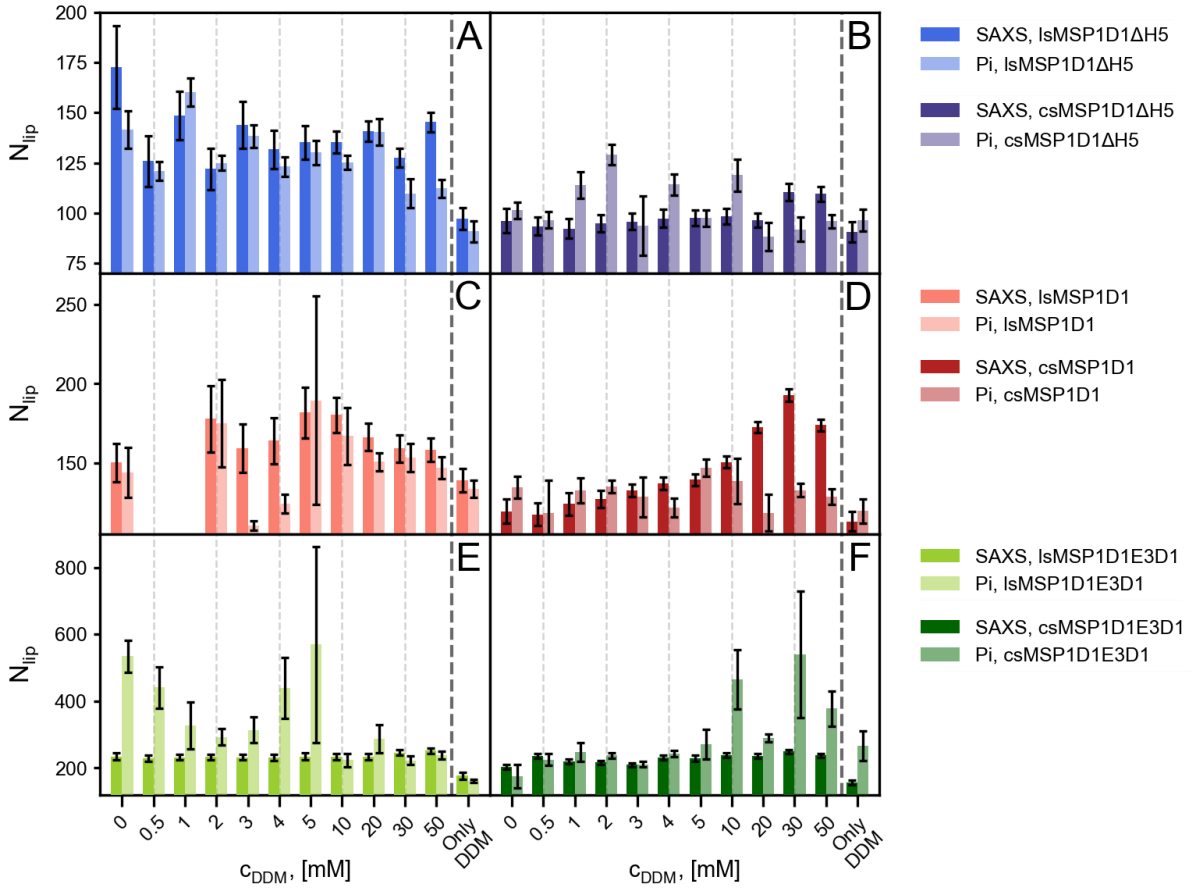


Figure A.1: *Bar chart of the number of lipids, N_lip, determined from SAXS (WillItFit) and the first phosphate analysis as function of the concentration of DDM, c_{DDM}.* ***A****: lsMSP1D1ΔH5* ***B****: csMSP1D1ΔH5* ***C****: lsMSP1D1* ***D****: csMSP1D1* ***E****: lsMSP1E3D1* ***F****: csMSP1E3D1*

The first phosphate analysis can be seen in figure A.1. The phosphate analyzes for lsMSP1D1E3D1 are extremely large, and much larger compared to the same numbers determined by SAXS. As all these very large phosphate measurements were determined on the same day, it seemed like something systematically had gone wrong, so the measurements for that particular day was discarded, and done again.

A.2 Fit parameters from WillItFit

See the next pages.

Table A.1 *Fit parameters with uncertainties for lsMSP1D1ΔH5 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.62±0.18	51.0±4.5	172±20	15514±3028	1259±729	2.0±2.6	(-0.9±2.2)e-4	4.3
0.5 mM DDM	1.65±0.2	60.5±4.2	125±12	22290±1309	1272±462	4.5±0.9	(-1.1±1.3)e-4	1.8
1 mM DDM	2.37±0.21	66.1±4.8	148±12	18410±1161	1303±374	4.7±0.8	(-1.2±1.5)e-4	3.4
2 mM DDM	1.5±0.16	58.7±3.3	121±10	22644±1131	1266±401	4.3±0.7	(-1.6±1.4)e-4	2.0
3 mM DDM	2.16±0.2	64.1±4.3	143±11	19506±1161	1290±374	4.6±0.8	(-1.3±1.7)e-4	4.0
4 mM DDM	1.57±0.14	59.0±2.7	131±9	22071±997	1265±326	4.3±0.6	(-1.7±1.4)e-4	2.9
5 mM DDM	1.55±0.12	58.4±2.3	135±8	21786±891	1261±279	4.3±0.5	(-1.9±1.3)e-4	3.1
10 mM DDM	1.44±0.08	58.2±1.5	135±5	22222±582	1259±181	4.3±0.4	(-3.1±1.3)e-4	3.3
20 mM DDM	1.52±0.07	58.1±1.3	140±5	21654±546	1253±159	4.1±0.3	(-2.8±1.2)e-4	5.9
30 mM DDM	1.42±0.08	58.6±1.3	127±4	22745±506	1253±162	4.3±0.3	(-3.3±1.3)e-4	3.0
50 mM DDM	1.44±0.07	57.4±1.2	145±4	21667±551	1240±147	4.0±0.4	(-3.2±1.6)e-4	2.9
only 30 mM DDM	1.59±0.12	58.0±2.5	97±5	24074±694	1255±297	4.5±0.5	(-3.3±1.6)e-4	2.2

Table A.2 *Fit parameters with uncertainties for csMSP1D1ΔH5 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.45±0.11	61.3±2.9	96±6	21916±619	1286±291	4.5±0.5	(-0.1±1.5)e-4	1.4
0.5 mM DDM	1.46±0.08	62.1±2.3	93±4	21846±455	1290±223	4.4±0.4	(-2.4±1.4)e-4	1.7
1 mM DDM	1.46±0.09	61.9±2.6	92±4	21968±493	1289±244	4.5±0.4	(-2.6±1.5)e-4	1.4
2 mM DDM	1.45±0.07	61.3±2.1	94±4	21862±446	1285±211	4.4±0.4	(-2.5±1.3)e-4	1.6
3 mM DDM	1.43±0.07	61.2±2.0	95±4	21958±428	1283±199	4.4±0.3	(-3.2±1.4)e-4	1.5
4 mM DDM	1.42±0.08	60.8±2.1	97±4	21685±480	1281±217	4.3±0.4	(-3.1±1.6)e-4	1.4
5 mM DDM	1.41±0.07	61.0±1.9	97±3	21937±423	1277±188	4.4±0.3	(-3.1±1.4)e-4	1.5
10 mM DDM	1.41±0.07	60.5±1.8	98±3	21718±436	1275±189	4.3±0.3	(-3.5±1.5)e-4	1.5
20 mM DDM	1.43±0.07	61.0±1.7	96±3	21881±391	1273±172	4.3±0.3	(-3.1±1.4)e-4	1.4
30 mM DDM	1.38±0.08	59.5±1.6	110±4	21466±484	1255±172	4.1±0.4	(-3.9±2.0)e-4	1.3
50 mM DDM	1.4±0.06	59.9±1.4	109±3	21349±401	1265±150	4.2±0.3	(-3.4±1.5)e-4	1.9
only 30 mM DDM	1.53±0.1	62.3±3.1	90±5	21644±545	1287±271	4.4±0.5	(-2.2±1.6)e-4	1.5

Table A.3 *Fit parameters with uncertainties for lsMSP1D1 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.5±0.16	59.2±2.7	150±12	25543±1074	1268±350	4.5±0.7	(-1.8±1.1)e-4	1.9
2 mM DDM	2.07±0.29	62.7±5.2	177±20	21487±1875	1283±530	4.4±1.2	(-1.0±1.6)e-4	2.6
3 mM DDM	1.8±0.23	63.0±3.8	159±15	24956±1265	1275±398	4.9±0.8	(-1.7±1.3)e-4	2.3
4 mM DDM	1.77±0.21	61.9±3.4	163±14	24545±1259	1270±374	4.8±0.8	(-1.7±1.3)e-4	2.6
5 mM DDM	1.88±0.23	61.7±3.3	181±16	23811±1412	1265±367	4.7±0.8	(-1.7±1.3)e-4	3.2
10 mM DDM	1.68±0.15	60.4±2.1	180±11	24099±951	1262±247	4.5±0.6	(-1.7±1.2)e-4	3.7
20 mM DDM	1.44±0.12	58.7±1.7	166±8	24986±771	1257±215	4.4±0.5	(-2.4±1.3)e-4	2.1
30 mM DDM	1.44±0.12	60.3±1.8	159±8	26058±715	1259±210	4.8±0.5	(-2.6±1.3)e-4	2.1
50 mM DDM	1.44±0.1	58.9±1.6	158±7	25363±666	1259±196	4.4±0.4	(-2.5±1.1)e-4	2.2
only 30 mM DDM	1.59±0.11	60.8±2.0	138±7	26134±671	1268±236	4.7±0.4	(-2.6±1.3)e-4	3.3

Table A.4 *Fit parameters with uncertainties for csMSP1D1 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.54±0.11	62.6±2.7	119±7	24361±642	1283±271	4.5±0.5	(-1.9±1.4)e-4	2.5
0.5 mM DDM	1.54±0.11	63.3±2.7	117±7	24691±621	1284±268	4.6±0.5	(-2.0±1.4)e-4	2.4
1 mM DDM	1.48±0.11	61.2±2.3	124±7	24344±645	1278±255	4.5±0.5	(-2.4±1.5)e-4	2.0
2 mM DDM	1.44±0.08	60.7±1.6	127±5	24327±488	1274±185	4.4±0.4	(-2.9±1.4)e-4	2.3
3 mM DDM	1.42±0.06	59.7±1.1	132±4	24342±386	1268±136	4.3±0.3	(-3.3±1.2)e-4	3.2
4 mM DDM	1.4±0.06	59.5±1.0	136±3	24320±376	1263±125	4.3±0.3	(-3.6±1.2)e-4	2.8
5 mM DDM	1.4±0.06	59.2±1.0	139±3	24128±367	1260±119	4.2±0.3	(-3.6±1.2)e-4	2.2
10 mM DDM	1.4±0.06	59.0±0.9	150±3	23999±361	1252±104	4.3±0.2	(-4.8±1.4)e-4	2.3
20 mM DDM	1.41±0.05	58.2±0.7	172±3	23088±339	1241±81	4.0±0.2	(-4.6±1.3)e-4	4.0
30 mM DDM	1.41±0.05	57.7±0.6	192±3	22514±371	1230±75	3.8±0.2	(-4.3±1.4)e-4	5.0
50 mM DDM	1.41±0.05	58.3±0.7	173±3	23180±353	1240±83	4.0±0.2	(-4.3±1.3)e-4	3.8
only 30 mM DDM	1.59±0.1	61.0±2.4	112±6	24358±615	1277±263	4.3±0.5	(-2.0±1.4)e-4	2.2

Table A.5 *Fit parameters with uncertainties for lsMSP1D1E3D1 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.55±0.1	60.0±1.1	234±9	35573±580	1262±156	4.8±0.3	(-2.6±0.9)e-4	6.1
0.5 mM DDM	1.55±0.11	60.0±1.2	228±9	35805±596	1261±164	4.8±0.4	(-2.9±1.0)e-4	5.3
1 mM DDM	1.55±0.09	60.0±1.0	232±8	35544±523	1262±142	4.8±0.3	(-2.8±0.9)e-4	7.6
2 mM DDM	1.52±0.09	59.8±0.9	232±8	35748±497	1260±133	4.8±0.3	(-3.2±0.9)e-4	7.1
3 mM DDM	1.51±0.09	59.9±1.0	231±8	35599±530	1261±143	4.8±0.3	(-3.1±1.0)e-4	6.3
4 mM DDM	1.52±0.1	59.7±1.1	230±9	35505±567	1260±154	4.7±0.3	(-3.3±1.0)e-4	5.2
5 mM DDM	1.47±0.11	59.9±1.2	233±10	35824±646	1259±171	4.8±0.4	(-3.3±1.3)e-4	3.7
10 mM DDM	1.46±0.1	59.4±1.1	234±9	35464±580	1257±152	4.7±0.4	(-3.4±1.2)e-4	3.2
20 mM DDM	1.46±0.09	59.4±1.0	233±8	35431±522	1256±136	4.6±0.3	(-3.5±1.2)e-4	2.8
30 mM DDM	1.42±0.09	59.4±0.9	246±8	35270±529	1255±130	4.6±0.3	(-3.7±1.3)e-4	2.7
50 mM DDM	1.42±0.09	59.0±0.9	251±8	34816±539	1255±129	4.5±0.3	(-3.3±1.2)e-4	2.9
only 30 mM DDM	1.79±0.17	61.9±2.3	176±10	36251±818	1262±285	4.9±0.6	(-2.6±1.4)e-4	2.8

Table A.6 *Fit parameters with uncertainties for csMSP1D1E3D1 from WillItFit.*

	Axis ratio	A_{head}, [Å^2]	N_{lip}	V_{belt}, [Å^3]	V_{lip}, [Å^3]	Roughness, [Å]	Background, [cm$^{-1}\cdot\mu$M^{-1}]	χ^2_{red}
0 mM DDM	1.66±0.09	61.3±1.1	202±7	34402±459	1269±143	4.9±0.3	(-3.2±1.0)e-4	6.4
0.5 mM DDM	1.43±0.08	59.7±0.9	236±7	33991±455	1263±119	4.8±0.3	(-3.2±1.0)e-4	4.0
1 mM DDM	1.5±0.07	60.5±0.9	220±6	34471±401	1266±115	4.9±0.2	(-3.2±0.8)e-4	6.8
2 mM DDM	1.54±0.07	60.1±0.9	216±6	34030±409	1266±118	4.8±0.3	(-3.5±0.9)e-4	6.8
3 mM DDM	1.54±0.07	60.8±0.8	209±5	34645±363	1264±107	5.0±0.2	(-3.9±0.9)e-4	9.0
4 mM DDM	1.47±0.07	59.5±0.8	231±6	33865±417	1262±111	4.7±0.3	(-3.7±1.0)e-4	5.8
5 mM DDM	1.46±0.09	60.1±1.0	229±8	34256±513	1262±137	4.8±0.3	(-3.8±1.2)e-4	3.3
10 mM DDM	1.45±0.07	59.2±0.8	237±7	33531±446	1260±114	4.6±0.3	(-4.4±1.1)e-4	3.5
20 mM DDM	1.44±0.06	59.4±0.7	236±6	33533±394	1260±101	4.6±0.2	(-3.4±1.0)e-4	4.0
30 mM DDM	1.43±0.06	58.9±0.7	249±6	33121±383	1254±89	4.5±0.2	(-3.5±1.0)e-4	3.3
50 mM DDM	1.46±0.06	59.3±0.6	237±5	33360±344	1257±86	4.5±0.2	(-3.9±0.9)e-4	4.9
only 30 mM DDM	1.92±0.13	62.5±2.1	156±7	33938±627	1270±247	4.7±0.5	(-3.8±1.2)e-4	4.7

A.3 Fits of exponentially modified Gaussians to the SEC-chromatograms

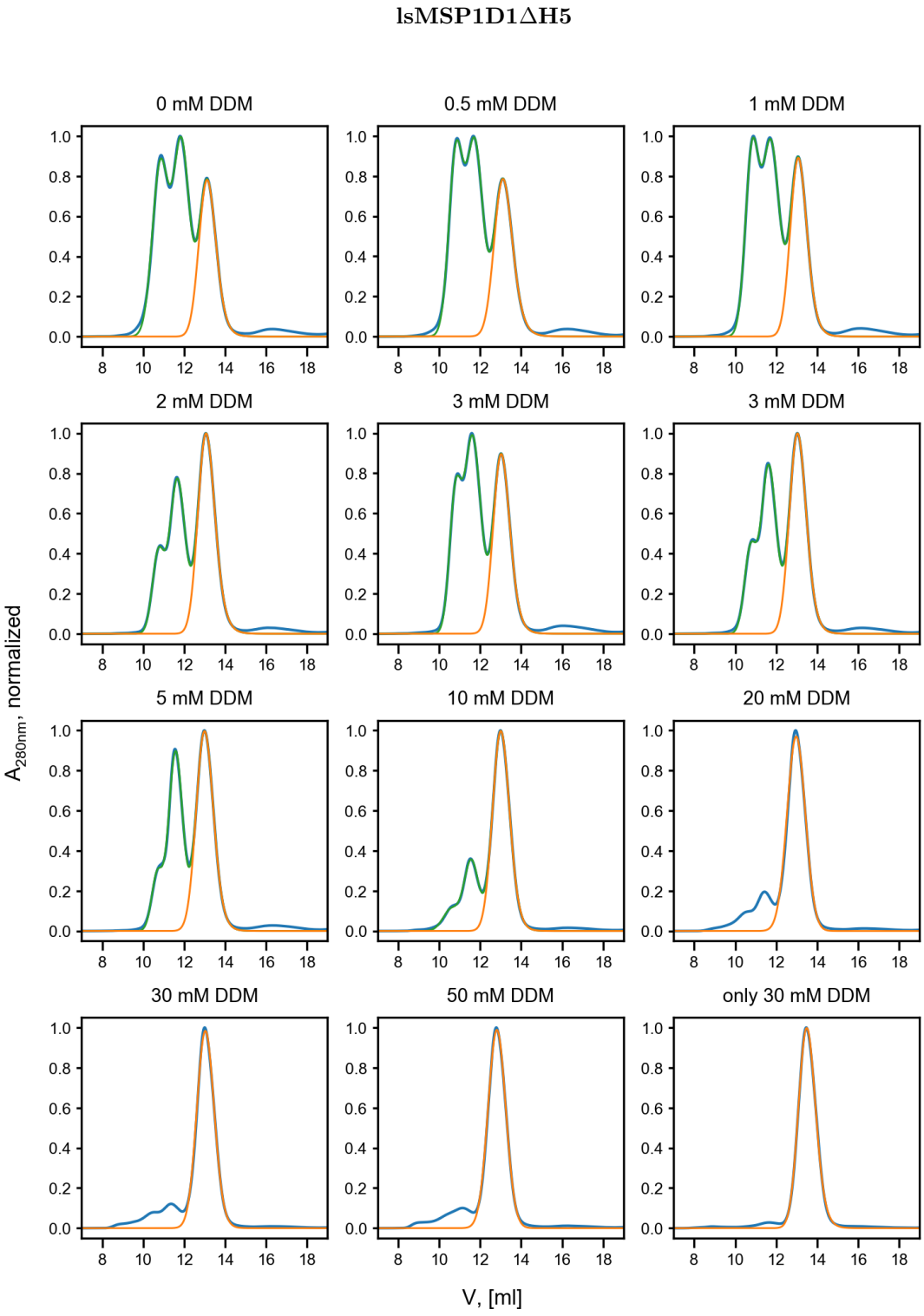


Figure A.2: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of lsMSP1D1ΔH5 with normalized absorption as function of the elution volume. For some curves it was necessary to fit a sum of two Gaussians and the EMG, to account for the other peaks and obtain a proper fit to the nanodisc peak. The blue curves are the measured SEC-data, the green curves are the multi-Gaussians, and the orange curves are the EMGs of the nanodisc peaks.*

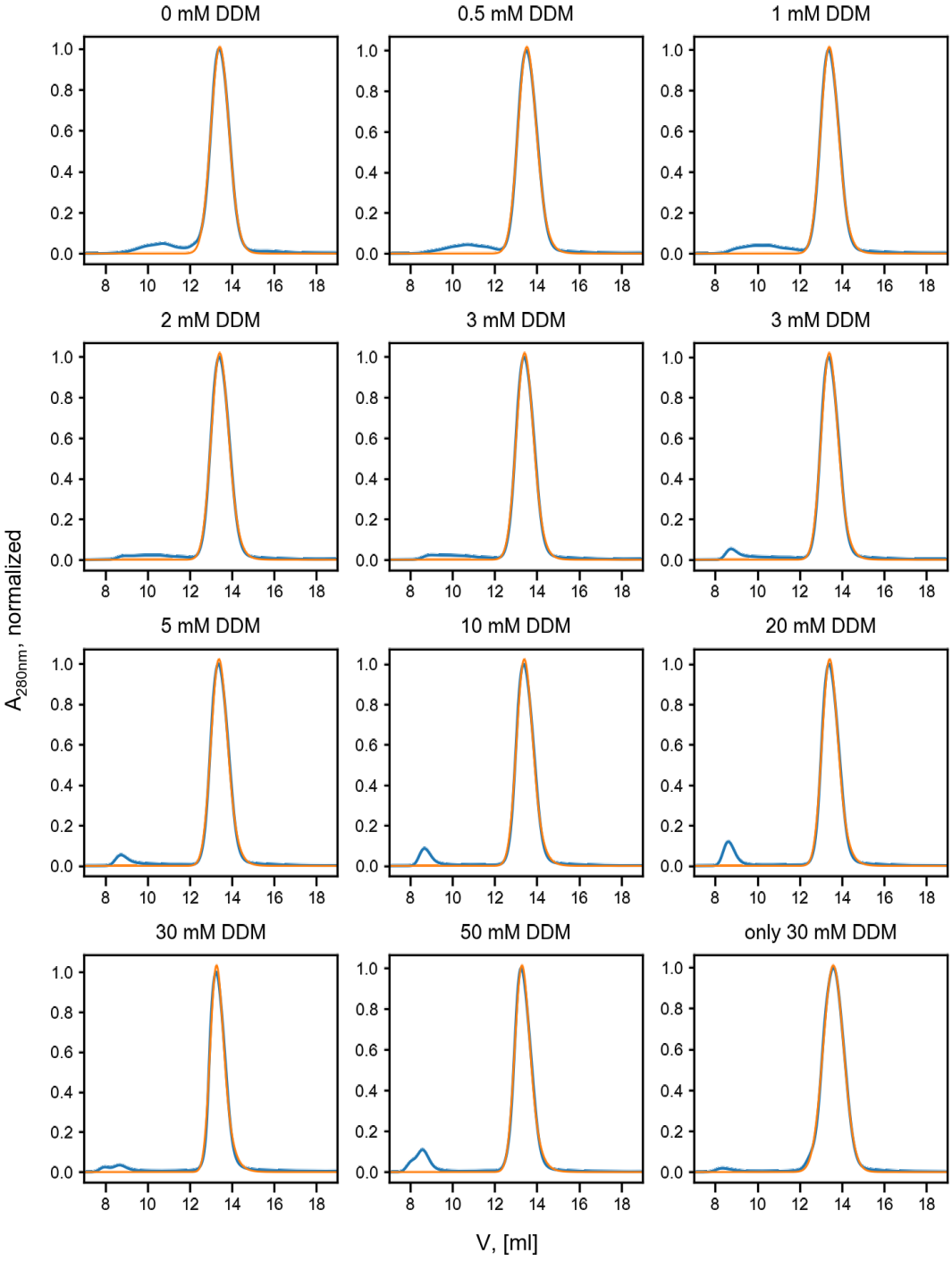


Figure A.3: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of csMSP1D1ΔH5 with normalized absorption as function of the elution volume. The blue curves are the measured SEC-data, and the orange curves are the EMGs of the nanodisc peaks.*

lsMSP1D1

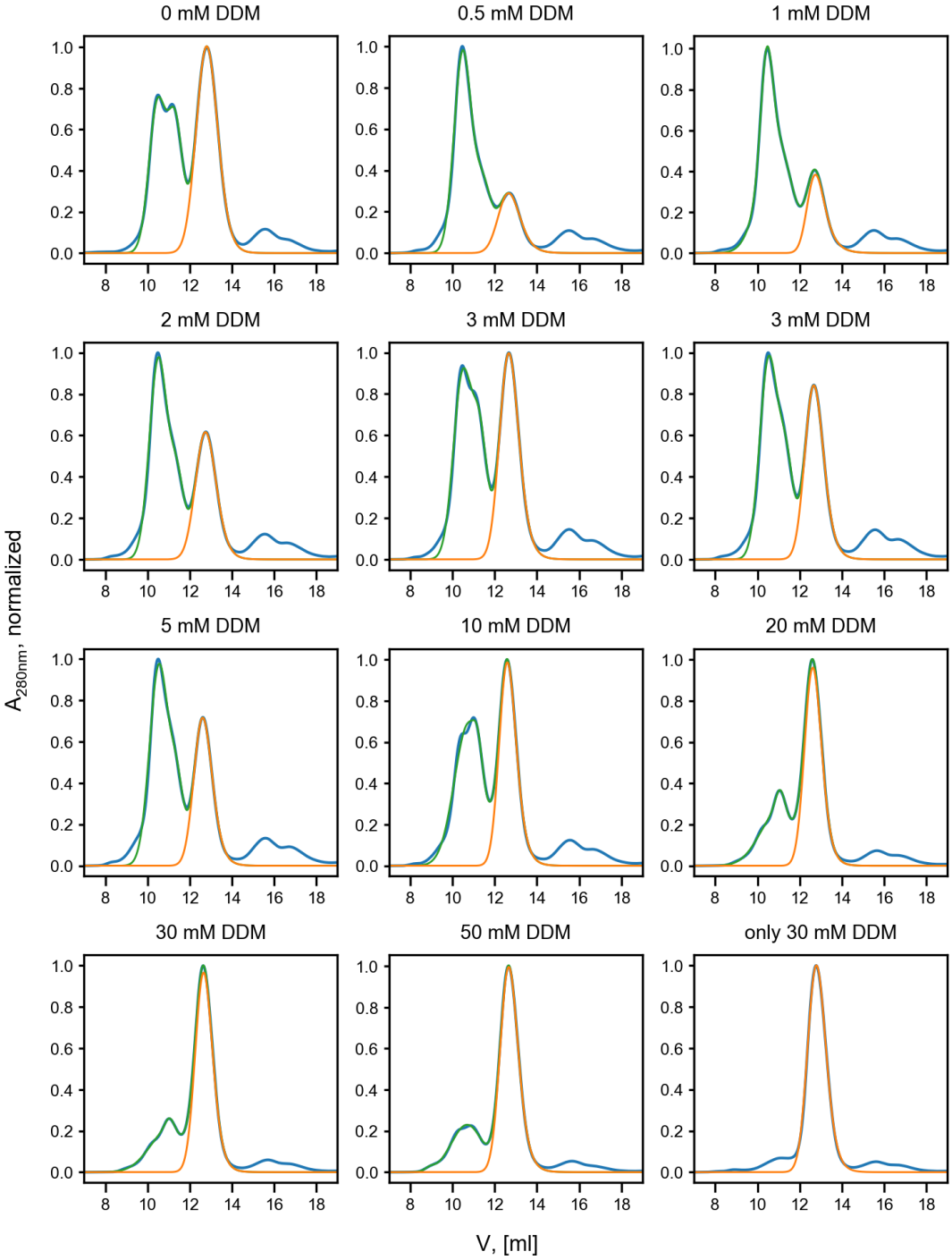


Figure A.4: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of lsMSP1D1 with normalized absorption as function of the elution volume. For some curves it was necessary to fit a sum of two Gaussians and the EMG, to account for the other peaks and obtain a proper fit to the nanodisc peak. The blue curves are the measured SEC-data, the green curves are the multi-Gaussians, and the orange curves are the EMGs of the nanodisc peaks.*

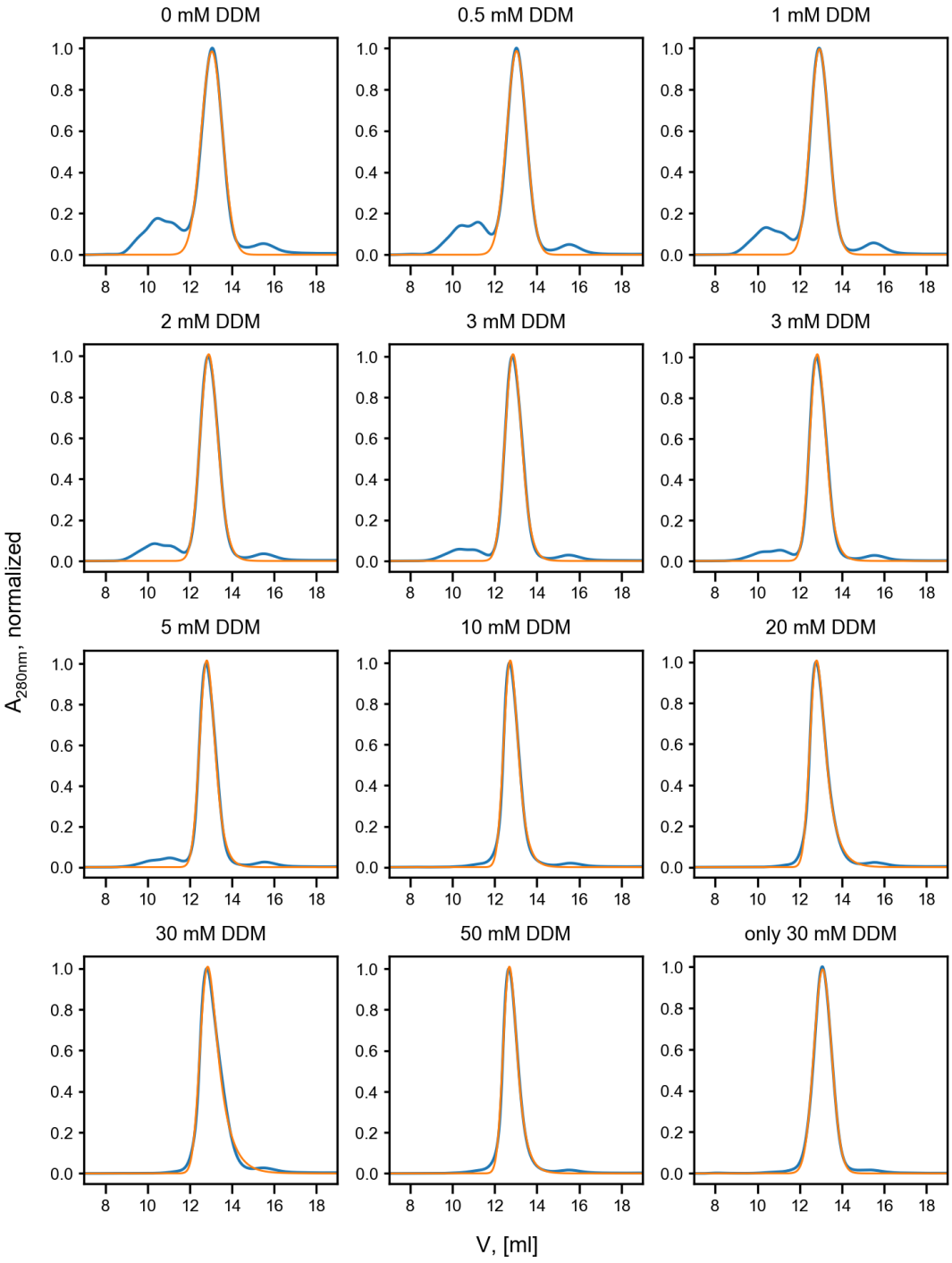


Figure A.5: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of csMSP1D1 with normalized absorption as function of the elution volume. The blue curves are the measured SEC-data, and the orange curves are the EMGs of the nanodisc peaks.*

lsMSP1D1E3

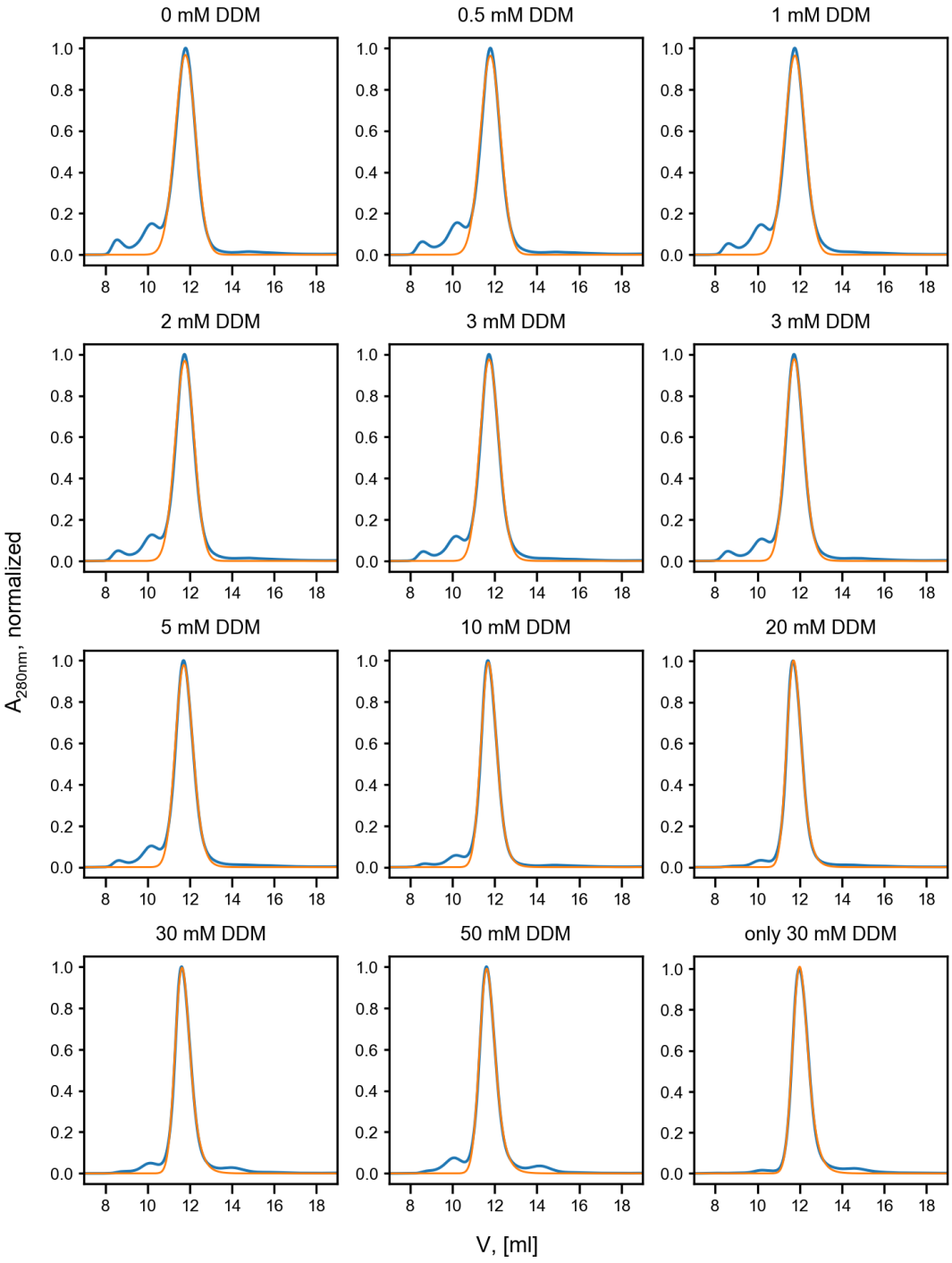


Figure A.6: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of lsMSP1D1E3 with normalized absorption as function of the elution volume. The blue curves are the measured SEC-data, and the orange curves are the EMGs of the nanodisc peaks.*

csMSP1D1E3

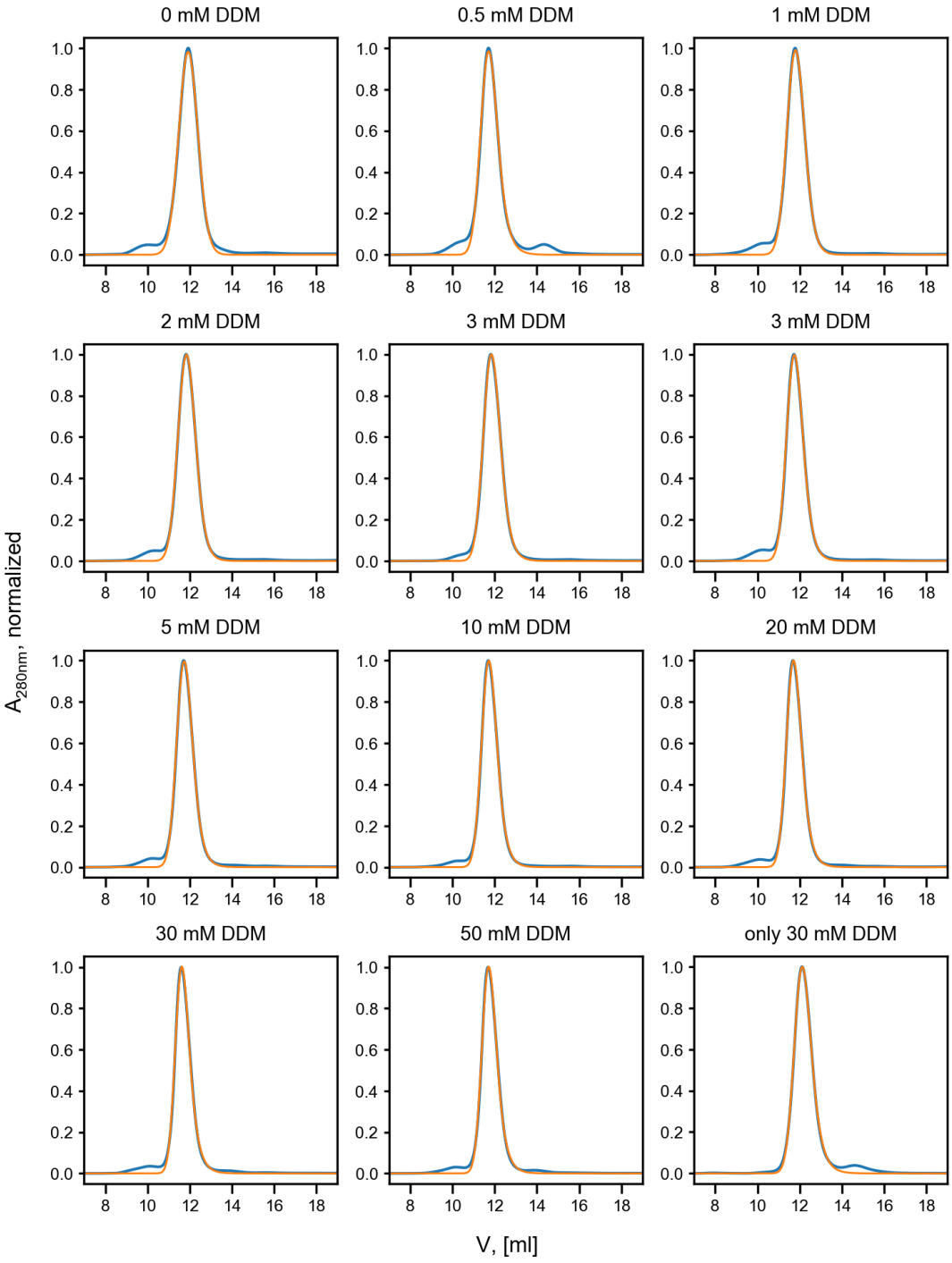


Figure A.7: *Fits of exponentially modified Gaussians (EMG) to the nanodisc peaks in the SEC curves of csMSP1D1E3 with normalized absorption as function of the elution volume. The blue curves are the measured SEC-data, and the orange curves are the EMGs of the nanodisc peaks.*

A.4 SEC of proteorhodopsin at 280 nm

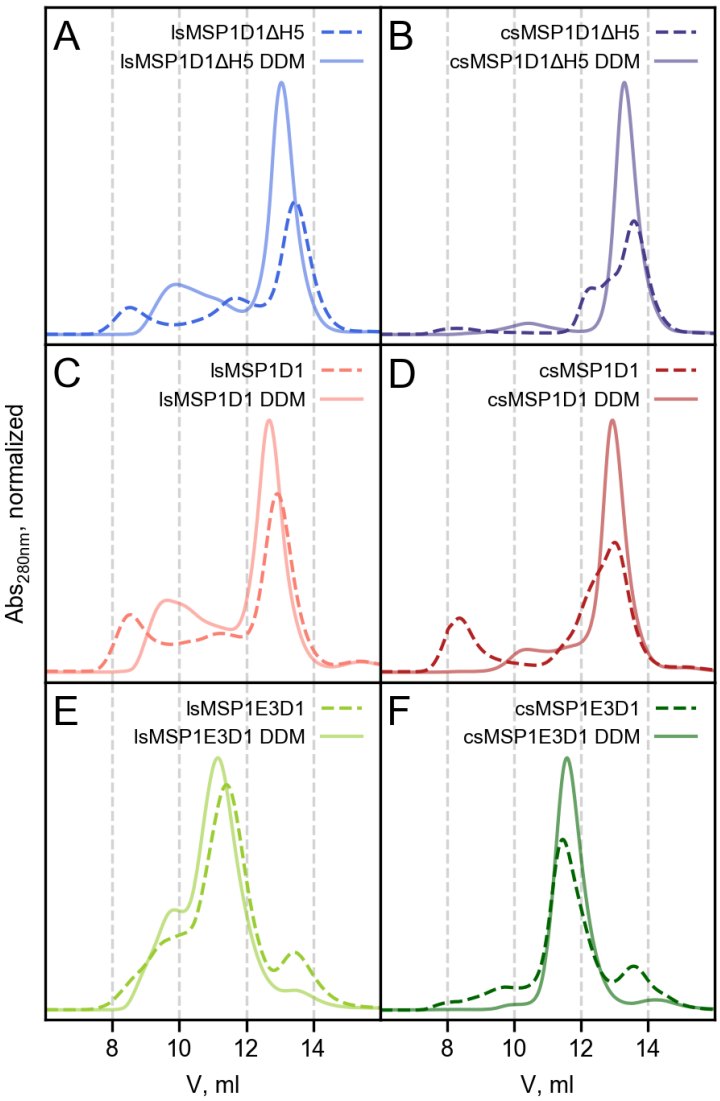


Figure A.8: *SEC chromatograms at 280 nm of nanodiscs reconstituted with proteorhodopsin (PR) and with and without 30 mM DDM.* ***A****: lsMSP1D1ΔH5 ND.* ***B****: csMSP1D1ΔH5 ND.* ***C****: lsMSP1D1 ND.* ***D****: csMSP1D1 ND.* ***E****: lsMSP1E3D1 ND.* ***F****: csMSP1E3D1 ND.*