Modelling and analysis of membrane proteins with flexible structures as investigated by small-angle scattering.

Master project

Written by:

Jakob Holmsted Kruse

June 30, 2022

Supervised by:

Heloisa Nunes Bordallo

University of Copenhagen

UNIVERSITY OF COPENHAGEN

FACULTY:	Science	
Niels Borh institute		
AUTHOR:	Jakob Holmsted Kruse	KU-ID: KLQ995
TITLE:	Modelling and analysis of membrane proteins with flexible structures as investigated by small-angle scattering.	
SUPERVISOR:	Heloisa Nunes Bordallo	bordallo@nbi.ku.dk
HANDED IN:	20^{th} of June 2022	
DEFENDED:	30^{th} of June 2022	

NAME _____

SIGNATURE _____

DATE _____

Abstract

Small-Angle Scattering studies of flexible membrane protein structures: Many membrane proteins have large flexible domains which make them very challenging to study with classical structural biology methods such as CryoEM and protein crystallography. Modern Small-angle X-ray and Neutron Scattering (SAXS and SANS) instruments through their combination with SEC in so-called SEC-SAXS and SEC-SANS setups have the potential of providing a wealth of information about such systems but in order to interpret the data a detailed modelling of the hypothesised structure is required, that integrates the available information about the system. This must be done case by case. In the MSc project, SAS data from Nanodiscs will be interpreted and refined, with emphasis on the Tissue-factor complex in a Nanodisc, using point cloud modeling and Fast Debye Sum(FDS). The task of the MSc student is to perform a full analysis of the data including testing and developing different structural hypotheses. The data will be reduced and afterwards analysed using various locally developed Python and C++ software which the MSc student will have to master and be able to adapt for the purpose. The goal of the MSc project is to be able to propose a SAXS-based low resolution structure for the studied membrane protein and to describe its flexibility.

Acknowledgement

First of all, I would like to gratefully acknowledge my two supervisors. Lise Arleth for introducing me to the field and making this project possible. Heloisa Nunes Bordallo for her insightful perspectives and for supporting me in my last year as a master student. I would also like to thank Martin Cramer Pedersen and Abigail Barclay for their useful comments and suggestions throughout the whole three years. Further, I would like to thank Copenhagen University for having a supporting system when I experienced personal challenges finishing my thesis. Finally, I would like to thank my family who have cheered and encouraged me all the way.

Acronym	Meaning
PDB	Protein Data Bank
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
$p(r)$	Pair-distance distribution
R_g	Radius of gyration
SAXS	Small-angle X-ray scattering
SEC	Size-exclusion chromatography
ESRF	European Synchrotron Radiation Facility
FDS	Fast Debye Sum
TF	Tissue Factor
MSD	Membrane Scaffold Protein

Symbol	Meaning
b	Scattering length
$I(q)$	Scattering intensity
$P(q)$	Particle form factor
ρ	Scattering length density

Contents

Chapter 1

Motivation

The reason why I stated at the Physics studies back in 2004 was deepfelt yearning after working at the frontier of knowledge production and exploration new discoveries. The journey alone the way has been bumpy the say the least. However the studies into scattering and working with rethinking the way we are able to use modeling for hypothesise testing, have been very rewording.

When we look at the Protein Data Bank 86 percent of all entries are produced with X-rays. Close to every one of these are found using methods not able to say anything about the flexibility in the protein and how they function in their natural environment. When using nanodiscs with embedded membrane proteins we are able to make experiments close to the proteins in vivo state. The aim of this thesis is to further this development using point could modeling and Fast Debye Sum (FDS).

In this thesis I will do a full analysis on a data set recorded on a beamtime at ESRF in 2018. I will code scripts to perform data reduction and analysis. Then i will fit the data against existing PDB's and use this to determine the structure of the particles in the sample. Then i will produce a point cloud model that can be used to analyse the data further together with the FDS software. During this work mastering the different coding languages is a point of emphasis.

Chapter 2

Introduction

2.1 Introduction to Tissue Factor and Nanodiscs

A huge part of the studies done in the biophysics field are about proteins. Proteins are vital part of how the human can function as a biologycal system, they are macromolecules with a large variety in their functions, maintain the shape of the cell membrane, make our muscles work and Hemoglobin is part of oxygen transport. There are more then 20,000 protein-coding genes in the human genome that each code for many protein[1] and each protein have a different function.

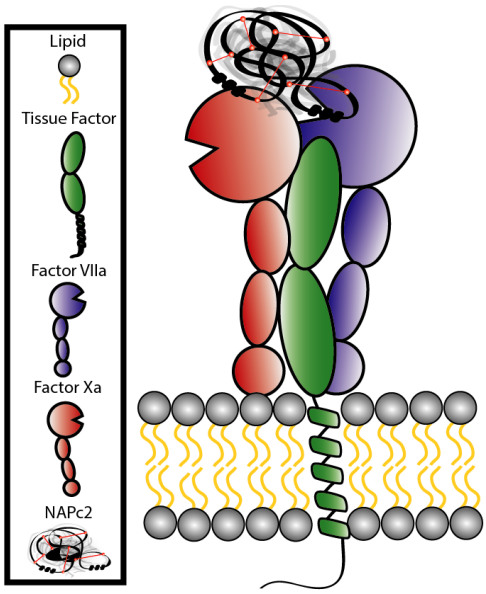


Figure 2.1: Tissue factor complex situated in the membrane. Two activation proteins FactorVII_a and FactorX_a interacts with TF and the complex is held together with the inhibitor protein $NAPc2$[2]. Figure from[2]

One of the protein we are actually able to spectate at work are the membrane protein Tissue Factor(TF), see Figure 2.1, that enables cells to initiate coagulation in the human blood. Tissue factor has an smaller transmembrane domain embedded in the cell membrane and a main extracellular domain outside the cell, it will activate and initiate coagulation when the extracellular domain interacts with other proteins. The link between the transmembrane domain and extracellular domain is flexible and will be affected by

the protein interaction[2].

When you want to study a membrane protein an important part is the understanding of its function in vivo. You have to study how it is embedded in the membrane, how it interacts with other proteins and that can be incredible challenging in vitro, especially when a flexible link or domain is present. A way to make this possible is to use a nanodisc, which has been a main focus of the work done in Professor Lise Arlehts biophysics group.

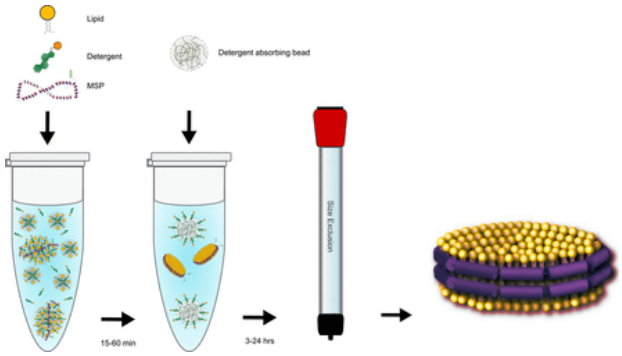


Figure 2.2: To assemble the nanodisc, the two phospholipids are mixed with the membrane scaffold protein(MSP) and reconstitution detergent in a buffer solution and left to equilibrate. To remove the detergent from the solution, detergent-absorbing resin beads are added and the sample will then contain nanodiscs of phospholipid and MSP plus some other aggregates. This aggregates can be removed by running the sample through an Size-exclusion chromatography(SEC) column [3].Figure adapted from[3]

A nanodisc are basically a small(nano size) disc of cell membrane made of phospholipids held together by a scaffold protein, see Figure 2.2, where the membrane protein can be embed and studied close to its in vivo state.

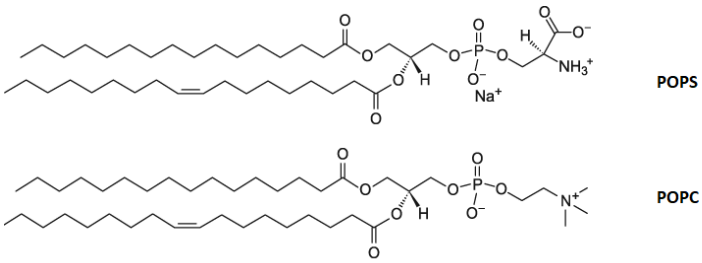


Figure 2.3: Structure of the two utilized phospholipids[4].

The data chosen for the full analysis in this thesis, was made by Professor Lise Arleths Biophysics group during a beamtime at the European Synchrotron Radiation Facility(ESRF) in Grenoble 2018. The focus of this beamtime was to study a Nanodisc and how the tissue factor complex is situated in the Nanodisc.

The nanodisc assemble was done with two phospholids, 20 % POPS and 80 % POPC, see Figure 2.3. Earlier studies have shown POPS to influence the activity of TF [2] and the 20 % was chosen with this in mind. Individual data sets where made with the nanodisc alone, TF in nanodisc, TF with FactorVII_a in nanodisc, the TF complex in nanodisc and finaly soluble tissue fatcor alone, that is TF without the transmembrane domain. All of these data set where analysed in this thesis.

2.2 Introduction to Small-Angle X-ray Scattering

Small-Angle X-ray Scattering (SAXS) are a very useful tool when studies strive to find the structure of proteins or other macromolecules. X-rays are able to probe structures with the size around $1 - 100nm$ so it is possible to measure the shape and size of each particle in a sample but not the finer structure on the atomic level, SAXS find usefulness in studies of diverse areas like pharmaceuticals, polymers, food-science, metallurgy and biological systems [5].

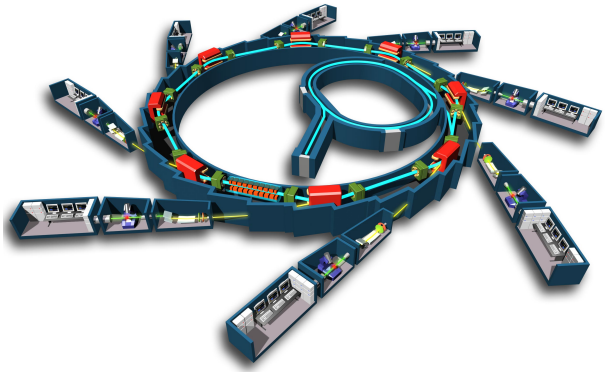


Figure 2.4: Schematic of the ESRF synchrotron in Grenoble, Figure from [6]. At ERSF the production of electromagnetic radiation begin with bundles of electrons emitted by an electron gun getting accelerated with electric fields through a linear accelerator into the booster synchrotron where the electron bundles will be further accelerated before they are sent into the large storage ring. In the storage ring the electrons travel close to the speed of light and when the electrons get bended out of their straight path by large bending magnets and undulaters (a series of smaller bending magnets) the electrons will emit electromagnetic radiation (Synchrotron Radiation), which includes X-rays.

The European Synchrotron Radiation Facility(ESRF) in Grenoble, see Figure 2.4, is dedicated to the production of electromagnetic radiation (photons), that can be focused and manipulated into a beam usable in experiments. The wavelength λ of the emitted electromagnetic radiation range from gamma rays to infrared and can be calculated with the energy of the emitted radiation E, Plank's constant h and the speed of light c.

$$\lambda = \frac{hc}{E} \tag{2.2.0.1}$$

At ESRF most of the emitted electromagnetic radiation have an energy around $8keV$ that corresponds to X-rays with a λ around $0.1nm$.

When X-rays are used for scattering experiments the interaction with the sample are mainly with the electrons of the atoms as x-rays are electromagnetic. The photons from the x-rays have a energy much smaller than the mass energy of electrons $\nu_p << \frac{m_e c^2}{h}$ or $8keV << 511kev$, therefore the scattering are Thomson scattering (elastic scattering) and the scattering length off each atom can be calculated with the Thomson radius constant $r_0 = 2.82 \cdot 10^{-13} cm$. For x-rays the scattering length of an atom $b = Z \cdot r_0$, since r_0 is a constant and Z is the number of electrons, the x-ray scattering length are proportional to the number of electrons [7], see Table 2.1.

Atom	Z	X-ray scattering length
H	1	$2.82 \cdot 10^{-13} cm$
C	6	$16.92 \cdot 10^{-13} cm$
N	7	$19.74 \cdot 10^{-13} cm$
O	8	$22.56 \cdot 10^{-13} cm$
Na	11	$31.02 \cdot 10^{-13} cm$
P	15	$42.30 \cdot 10^{-13} cm$
S	16	$45.12 \cdot 10^{-13} cm$

Table 2.1: Table of relevant elements and their scattering lengths.

2.2.1 Experimental set-up

When SAXS experiments are done at ESRF the typical experimental set-up will be as shown in Figure 2.5. The beam from the synchrotron will pass through a monochromater and then a series of apertures that collimate the sample, when the beam travel through the sample most of it pass without any interference and to protect the detector a beam stop is put in place. Some of the x-rays will interact with sample and be incoherently scattered in all directions, creating a constant background radiation. So it will only be a small part of the x-rays that interact with electrons in the sample and create Thomson scattering at an angle of 2θ onto the detector. The detector counts the number of x-rays it absorbs and over time a 2-d image is built up showing the scattering intensity [7].

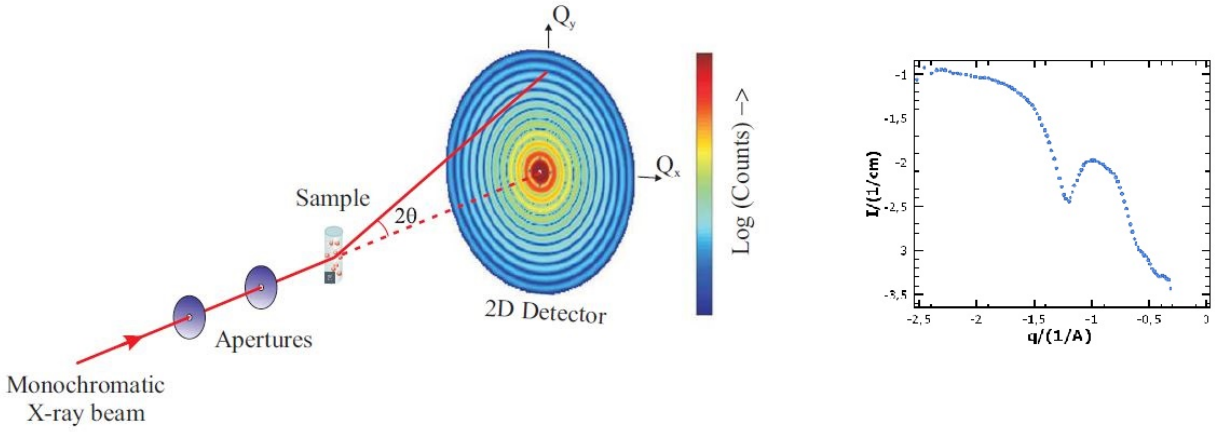


Figure 2.5: Small angle x-ray scattering scheme and the 1-d curve of the detected scattering intensity plotted on a log-log scale with respect to q. Data from the sample containing only nanodiscs, Figure adapted from [7].

The typical sample in a biological SAXS experiment contains a very large number of particles, around 10^{14} [8], situated totally independent to each other. This results in an isotropic pattern on the 2-d image that can be described with a 1-d curve of scattering intensity, see Figure 2.5.

In the five different samples examined during the beamtime at ESRF, the nanodisc and proteins was kept in a buffer solution. The x-ray beam will interact with the electrons in the buffer solution and contribute significantly to the Thomson scattering. So for each individual sample the pattern of scattering intensity was recorded with buffer solution only and subtracted from the total scattering pattern [5].

2.2.2 Scattering Theory

At a synchrotron facility with a monochromatic x-ray beam, the simplest scattering experiment to execute are the scattering of a single electron. The monochromatic beam can be described with the wave vector $\vec{k}$ and when it scatter of a point scatterer like a single electron it will be in a spherical wave with the wave vector $\vec{k}'$, see Figure 2.6, the scattering are Thomson scattering so the monochromatic beam don't loose any energy, therefore $\mid \vec{k} \mid = \mid \vec{k}' \mid$. The scattering vector $\vec{q} = \vec{k} - \vec{k}'$ and its magnitude can be expressed as equation 2.2.2.1.

$$\mid \vec{q} \mid = 2 \mid \vec{k} \mid sin\theta = \frac{4\pi sin\theta}{\lambda} \tag{2.2.2.1}$$

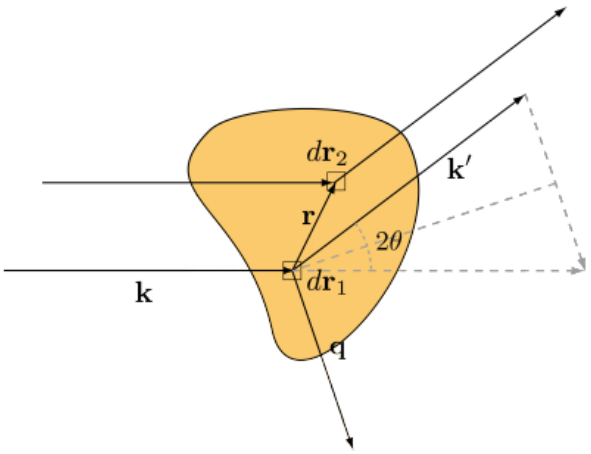


Figure 2.6: Schematic diagram depicting a monochromatic beam $\vec{k}$ scattering of two point scatterers $d\vec{r}_1$ and $d\vec{r}_2$, Figure from [9].

The scattering off two point scatterer separated by $\vec{r}$ will result in a phase difference between the two scattered beams $(\vec{k} - \vec{k}') \cdot \vec{r} = \vec{r} \cdot \vec{q}$ and the combined scattered amplitude $A(\vec{q}) = b_{r_1} + b_{r_2} e^{-i\vec{r}\cdot\vec{q}}$ [9]which means that the scattering amplitude is angle-dependent. The intensity of the combined scattered beam can be expressed as equation 2.2.2.3 [8].

$$I(\vec{q}) = \mid A(\vec{q}) \mid^2 \tag{2.2.2.2}$$

Through extensive derivation the Derbye equation can be derived [8]. Also when the particles are in a solution free to rotate to every orientation will the scattering pattern measured be centrosymmetric around the incoming x-ray beam and therefore the orientation of the scattering vector $\vec{q}$ is irrelevant.

$$I(q) = 4\pi \int p(r) \frac{sin(qr)}{qr} dr \tag{2.2.2.3}$$

$$p(r) = \langle \gamma(r) \rangle_{\Omega} r^2$$

Where $\gamma(r)$ is the distance correlation, Ω is the solid angle, $p(r)$ is known as the pair-distance distribution and quantifies the distances between each pair of scatterers. Hence a SAXS curve for a sample containing protein will show different structural information at different q, see Figure 2.7.

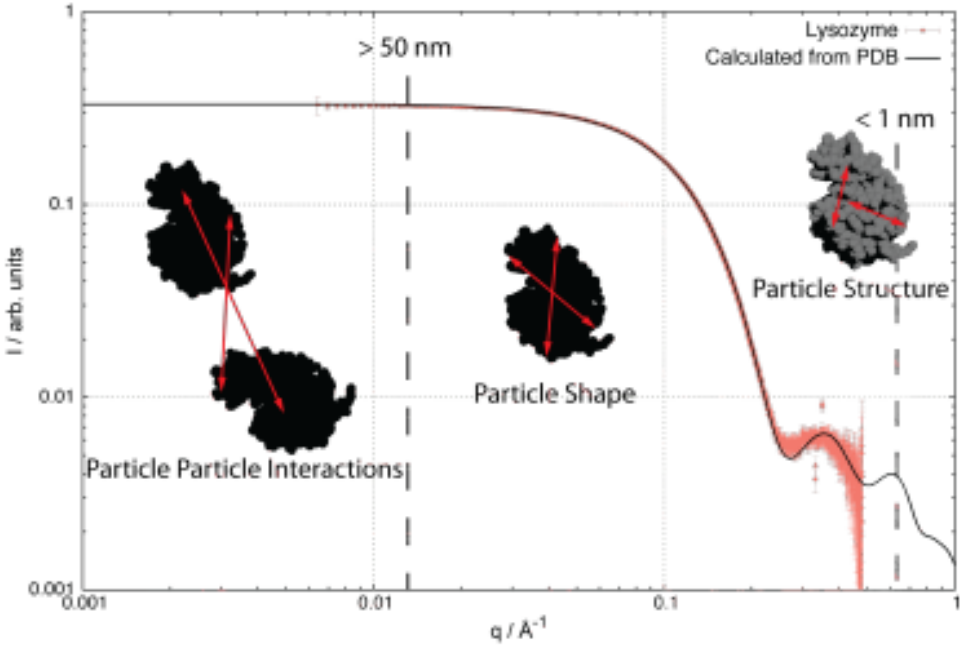


Figure 2.7: General SAXS curve showning different structural information at different q. At low-q the SAXS curve contains information about the interaction between the particles, the intermediate-q describe the particle shape and at high-q information about the internal structure of the proteins begin to appear. Figure from [9].

In very diluted samples the particle inter-particle interactions are negligible so $I(q)$ can be rewritten as:

$$I(q) = n\Delta\rho^2 V^2 P(q) \tag{2.2.2.4}$$

Where n is the number density of the particles, ρ is the scattering length density, the excess scattering length density $\Delta\rho = \rho_{particle} - \rho_{solvent}$, $P(q)$ is the particle form factor.

$$P(q) = \langle \mid F(\vec{q}) \mid^2 \rangle = \langle \mid \frac{1}{V} \int_V exp(i\vec{q} \cdot \vec{r}) dV \mid^2 \rangle \tag{2.2.2.5}$$

Where $F(\vec{q})$ is called the form factor amplitude.

The form factor describes in detail the shape and size of the particles in a sample and it is comparable to the information contained in the intermediate to high q-range of the SAXS curve. For simple structures the form factor can be evaluated analytically, but for more complex structures evaluation can become exceedingly challenging because the form factor has to calculated numerically and this is where point-cloud modelling can be a really useful tool.

Chapter 3

Data Analysis

A very important aspect when an application for time at a beamline is accepted and the group arrive at facility to execute the experiments at the designated time-slot, are that each sample are well prepared and ready for data sampling. When experimenting on samples of nanodiscs with or without embedded membrane-proteins this can be a challenge in itself. The samples are fragile and deteriorate quickly, so as soon as the assemble and purification shown in Figure 2.2 are concluded, the samples are frozen and prepared for transport. When performing the SAXS experiment the goal is to get as pure data as possible of the target molecule without the background noise from the buffer solution, deteriorated particles and larger aggregates leftover from the purification process. To achieve this size-exclusion chromatography(SEC) is coupled with SAXS experiment, see Figure 3.1.

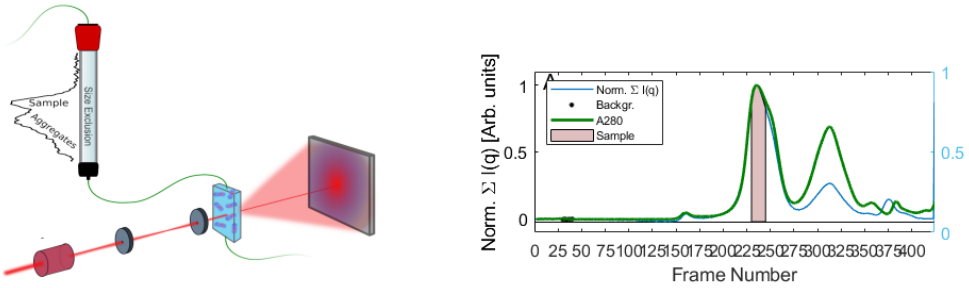


Figure 3.1: Experimental SEC-SAXS setup [10]. The SEC column are constructed with porous beads for the sample to pass through, the beads allow larger particle to pass straight through, but the smaller the particle gets the longer it will take to pass through. The normalized plot (a Chromatogram) depict the Scattering intensity $I(q)$ (blue) and Absorbance A (green) for each frame number. The sample area(red) are chosen as is corresponds with the peak in both Scattering intensity and Absorbance. In first frames of the experiment the plotted graph are constant and only buffer solution flow through the capillary. 10 frames are selected in this area and used to calculate the scattering intensity of the buffer solution. This background noise are subtracted from the recorded scattering pattern.

During a SEC-SAXS experiment the normal sample holder is replaced with capillary. The sample are transferred to the SEC column and buffer solution is added to press the sample through the column at the defined flow speed, as it elutes it travels through the capillary. The x-ray beam are directed through the capillary and the scattering pattern

can be recorded periodically.

For each of the five different sample analysed in this thesis a data set where recorded during a SEC-SAXS experiment. Beyond the computational corrections and reductions during the measurement no other manipulation of the data sets was performed.

3.1 Reduction of the data

With the raw data sets finally in hand, the analysis and the journey towards a new modelling regime begin. This was a very non trivial task, with no direct guidelines on how to approach the analysis and resent comparable work done case by case. A series of steps was set in place under guidance of phD-student Abigail Barclay and adjunkt Martin Cramer Pedersen, both previously part of Professor Lise Arleths Biophysics group.

First step was to code scripts that could reduce and plot the data sets, the coding was done in Matlap with information from the beamtime at ESRF. During the SEC-SAXS experiment the flow speed through the capillary where defined to make sure that scattering pattern of the target protein was recorded, in Figure 3.1 the duration of the or the plotted experiment was set to 400 frames. A frame are one run of the SAXS detector, the detector run for 10 seconds and each second a measurement is made, the measurement are combined in a scattering pattern that are recorded for the corresponding frame.

Parallel with the SEC-SAXS experiment a control experiment is made. The amino acids Tryptophane and Tyrosine play a vital role as building blogs in the TF-complex and MS-protein. They are also know to have a high absorbance at $280nm$, so when ultraviolet light with $\lambda = 280nm$ are transmitted through the capillary, an absorption peak will appear in the transmitted light when the proteins move through the capillary. The equation for calculating absorbance A are:

$$A = \epsilon \cdot c \cdot l \tag{3.1.0.1}$$

Where ϵ is molar absorptivity, c is molar concentration, l is the thickness of that material light travels through.

Implementations

Matlab Selecting frames, Avgdata, Pair-distance distribution, Guinierplot, Chromatogram and Radius of gyration.

Gnom Radius of gyration fitting PDB.

Crysol fitting PDB

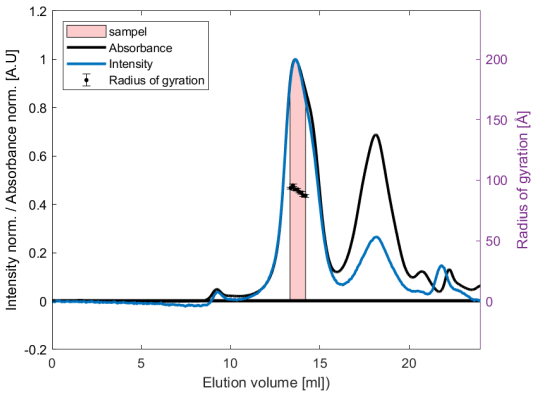
Bayes app Pair-distance distribution.

Pymol Visualization PDB

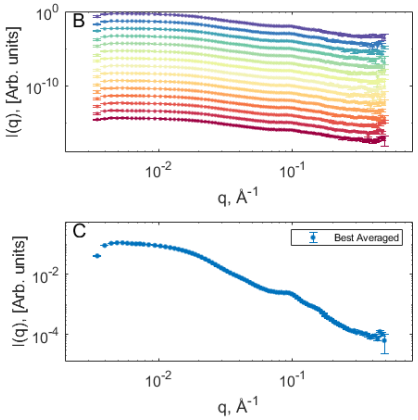
Willitfit Fitting PDB finding parameters used when running the FDS software.

Python Coding the FDS software

Opencl Coding the secondary software called into FDS.



(a) 10 frames are chosen in the sample area. The Radius of gyration are calculated for each and plotted in the chromatogram.



(b) With the eye test the 10 selected frames seem very similar. To ensure this is also the the case, the derivation from the mean are calculated for the 10 frames 3.3.

Figure 3.2: Finding the data set to put into the model.

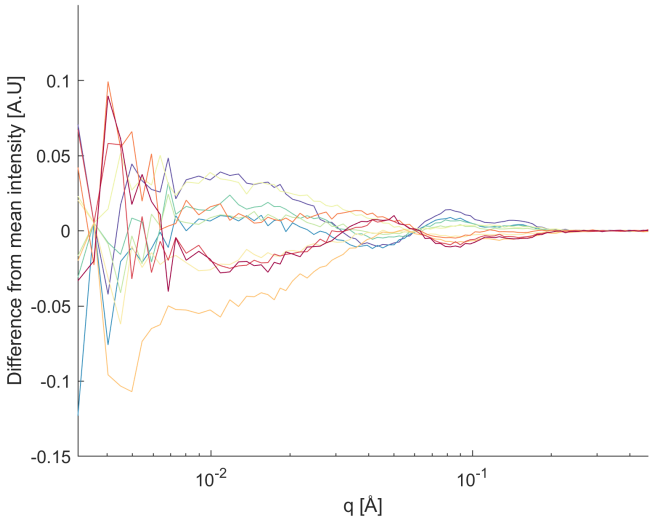


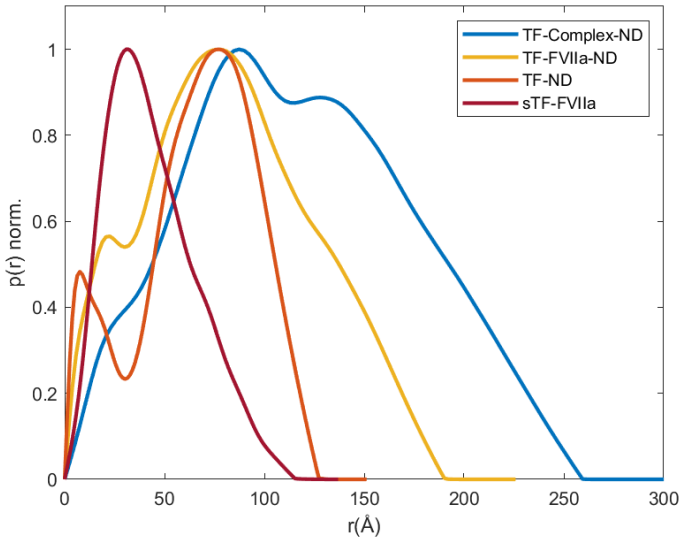
Figure 3.3: Deviation from the mean of the the chosen sample frames. Nanodisc sample.

3.2 Data Analysis

The radius of gyration R_g and Pair-distance distribution, can be used to say something about the shape and size of the particles in the sample.

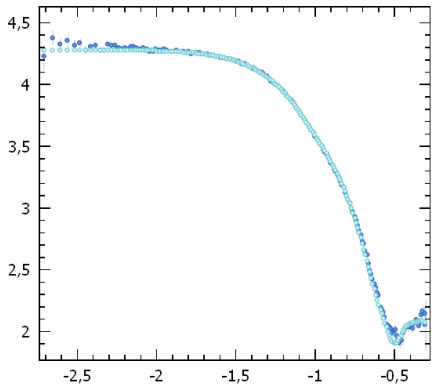
Sample	R_g(Å)
Nanodisc	52.01 Å
TF-ND	54.41 Å
TF-FVIIa-ND	66.17 Å
TF-complexND	92.18 Å
sTF-FVIIa-ND	34.24 Å

(a) R_g

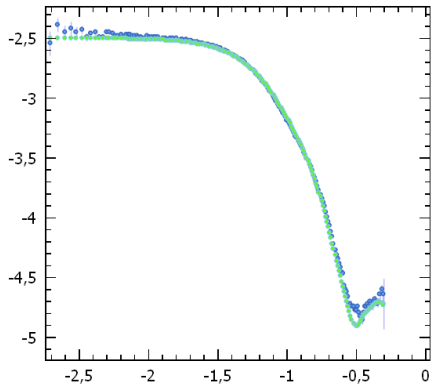


(b) Pair-distance distribution.

Table 3.1

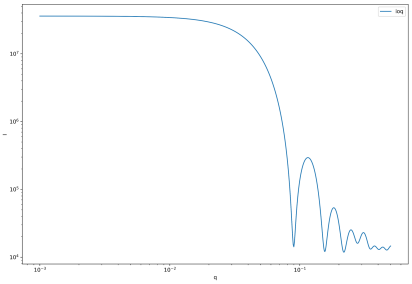


(a) Chi$^2 = 3.2 sTf - pdb$

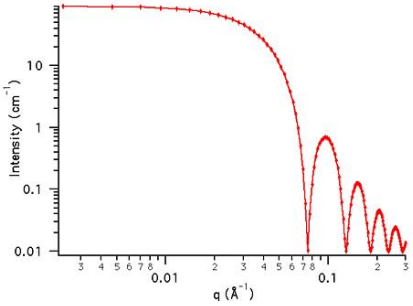


(b) Chi$^2 = 17.11 boy - pdb$

Figure 3.4: Fitting the PDB in Crysol



(a) FDS sphere



(b) P(q) formfactor sphere[11].

Figure 3.5: Model not accurate in high q, the number of points affect the scattering curve in high q.

The nanodisc is made with 20 % POPS with $(C40H75NO10PNa)$ and 80 % POPC with $(C42H82NO8P)$

Chapter 4

Models

4.1 Models

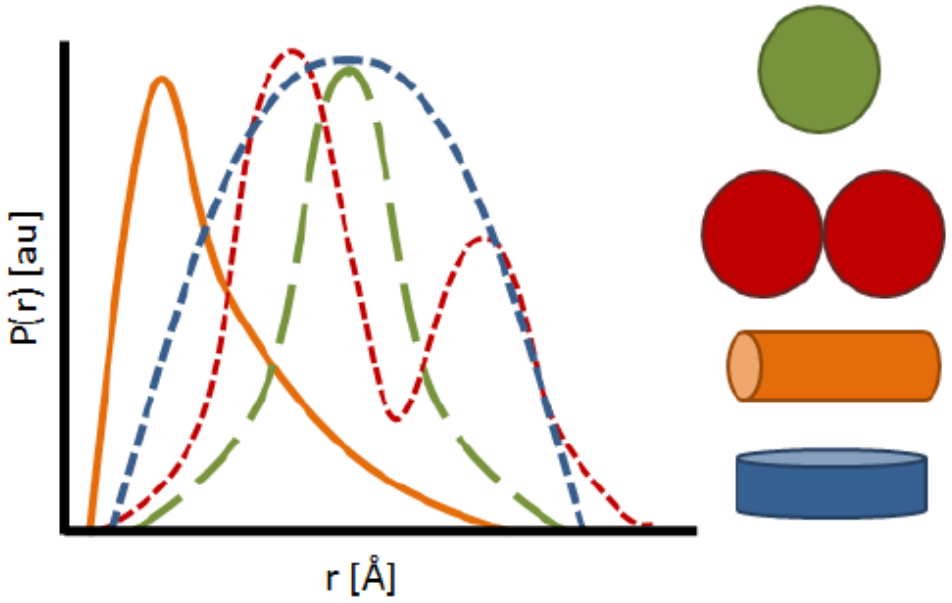


Figure 4.1: Pair distance distribution functions indicative of particle shape[12]

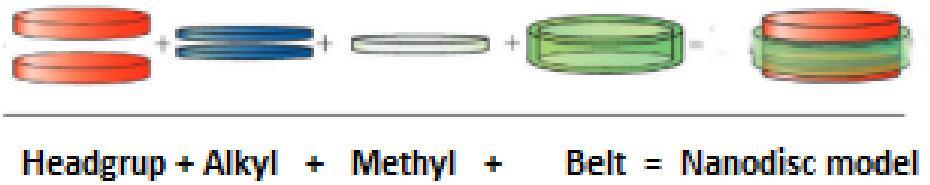


Figure 4.2: Making the geometry of the nanodiscmodel [10].

Chapter 5

Structural Analysis

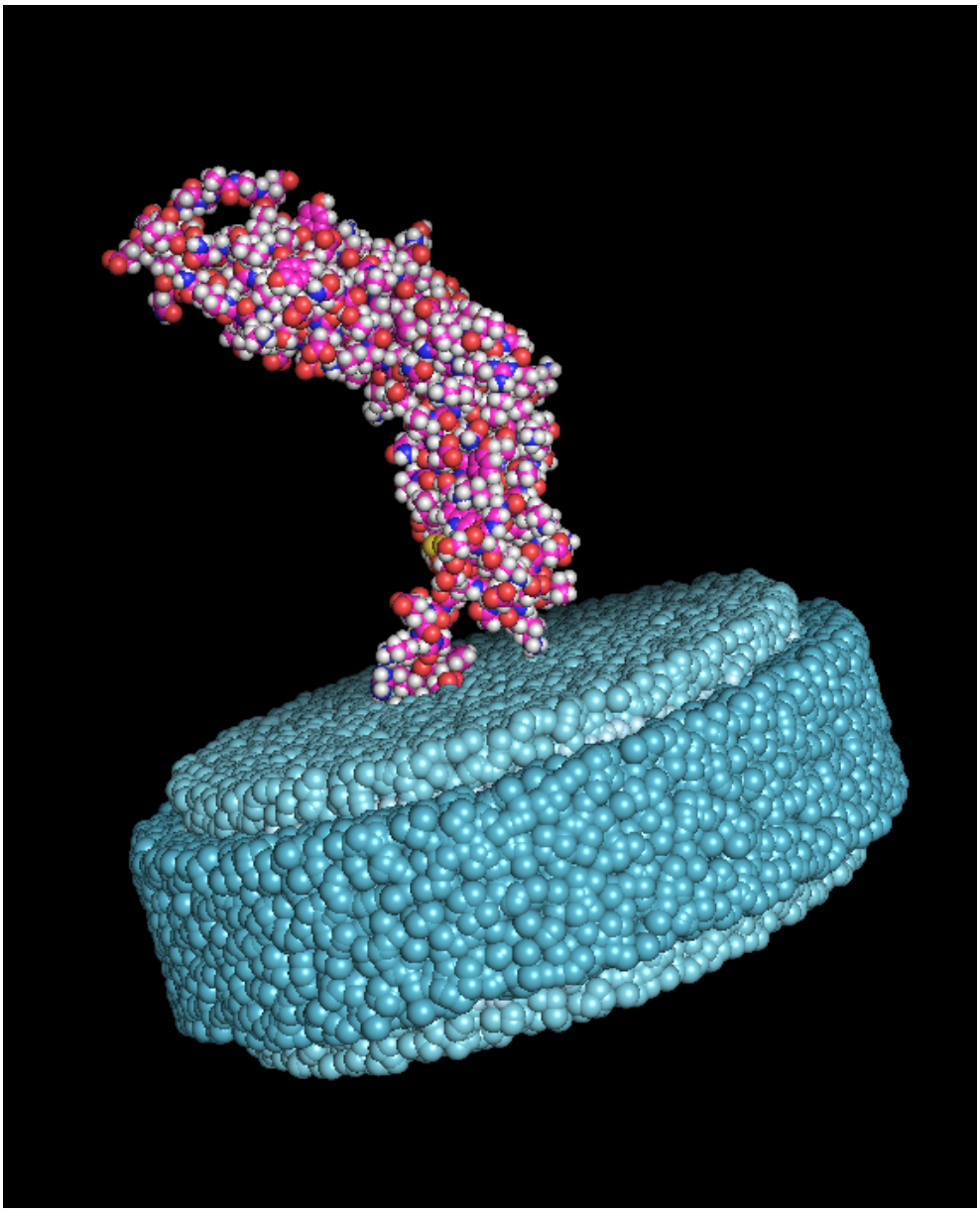


Figure 5.1: Tissue factor in the nanodisc

Chapter 6

Conclusion

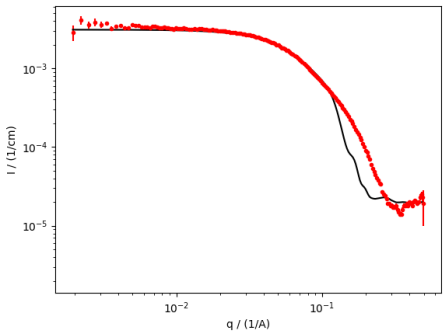
The aim this thesis was to perform data analysis, structural analysis and develop new modelling tools using point clouds and alone the hone my coding skills.

I have constructed numerous scripts in matlab that can be used as a general fitting toolbox, when working with scattering data. I have used this to reduce and analyse the given data sets and then fitted the data against different PDB's. I have constructed a point cloud model of a nanodisc with or with an embedded protein and made a proof of concept where i have shown that the scattering curve of point cloud model corresponds perfectly to that of the analytical model prediction. Before this project my skills at coding where close to none existing but along the way this have obviously developed and I am now comfortable working in Matlab and Python, plus experience in many other programs in the field of scattering.

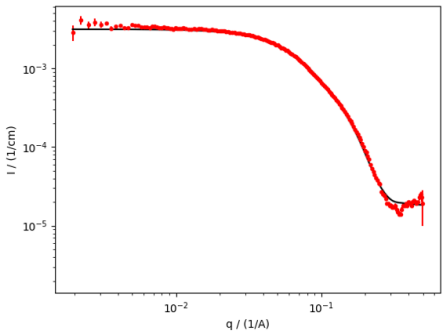
It is glaringly obviously that i ran out time writhing this thesis. The last chapters has close to no text and this was off cause not how I intended it.

Beyond that, if I had more time I would like to assemble the whole tissue factor complex in the point cloud model and run the week long simulations off how the complex will move on the flexible link between the complex and the trans-membrane domain.

Appendices



(a) sTf-pdb



(b) 1boy-pdb

Figure 1: Fitting the PDB in Willitfit

Chapter 7

Bibliography

[1] Ilgisonis EV Pyatnitskiy MA Kopylov AT Zgoda VG Lisitsa AV Archakov AI. Ponomarenko EA, Poverennaya EV. The size of the human proteome: The width and depth. *Int J Anal Chem.*, 2016.

[2] Frederik Tidemand. *Structural investigation of the tissue factor:Factor VIIa:Factor Xa complex*. PhD thesis, 2019.

[3] Nicholas Skar-Gislinge, Nicolai Tidemand Johansen, Rasmus Høiberg-Nielsen, and Lise Arleth. Comprehensive study of the self-assembly of phospholipid nanodiscs: What determines their shape and stoichiometry? *Langmuir*, 34(42):12569–12582, 2018.

[4] Avanti polar lipids. `https://avantilipids.com/product/850457`. Accessed: 2022-06-26.

[5] H. Schnablegger and Y. Singh. *The SAXS Guide, 4^th edition*. Anton Paar GmbH, 2017.

[6] Bellec Ewen. *Study of charge density wave materials under current by X-ray diffraction*. PhD thesis, 2019.

[7] Jens Als-Nielsen and Des McMorrow. *Elements of Modern X-ray Physics, Second Edition*. John Wiley & Sons, 2011.

[8] Abigail Barclay. Structural analysis and modeling of small-angle scattering data for membrane proteins and their carrier systems, 2019.

[9] Nicholas Skar-Gislinge. *Structural investigation of the tissue factor:Factor VIIa:Factor Xa complex*. PhD thesis, 2019.

[10] Skar-Gislinge and Lise Arleth. Small-angle scattering from phospholipid nanodiscs: derivation and refinement of a molecular constrained analytical model form factor. *Phys Chem Chem Phys.*, pages 3161–70, 2011.

[11] Model sphere. `https://www.ncnr.nist.gov/resources/sansmodels/Sphere.html`. Accessed: 2022-06-27.

[12] May Nyman and Lauren McQuade. *Small Angle X-ray Scattering of Group V Polyoxometalates*, pages 151–170. 06 2015. ISBN 978-1-63482-656-3.