# COMPUTATIONAL MODELLING IN STRUCTURAL BIOLOGY: INSIGHTS FROM SMALL-ANGLE SCATTERING AND MOLECULAR DYNAMICS SIMULATIONS

UNIVERSITY OF
COPENHAGEN

This thesis has been submitted to the PhD school of The Faculty of Science, University of Copenhagen, Denmark

For the PhD degree by

**Abigail Barclay**

Niels Bohr Institute
University of Copenhagen

August 2023

ABSTRACT

---

Proteins play a diverse and crucial role in essential physiological processes, intricately interacting with other proteins and biomolecules such as lipids or ligands. Understanding the structures and mechanisms of these biomolecular systems is crucial to understanding their specific functions. Experimental techniques are constantly improving. It is imperative to simultaneously develop computational methods to bridge the gap between raw data and meaningful results. The main focus of this thesis is on the development of analytical models and integrative computational tools to fully exploit the wealth of structural information that can be extracted from small-angle scattering (SAS) data. The thesis also explores how molecular dynamics (MD) simulations of proteins can enhance the interpretation of experimental data by providing insights which are not accessible through experiments alone.

Firstly, the thesis focuses on the advancement of size-exclusion chromatography coupled with small-angle x-ray scattering (SEC-SAXS) and introduces a novel procedure to investigate underlying structural distributions within a single species. It shows how an analytical model can be refined against many frames from the same SEC-SAXS data sets simultaneously to provide more robust fit results. The procedure is applied to study populations of nanodiscs. This thesis also explores different methods for modelling flexible membrane proteins embedded in nanodiscs. Flexible particles pose a challenge for SAS analysis, since the scattering signal is averaged over an ensemble of conformations. Therefore, an advanced semi-analytical model accounting for conformational diversity was built for the human growth hormone receptor (GHR) in a nanodisc and refined against SAXS data. The thesis goes on to discuss methods for ensemble modelling of membrane proteins in nanodiscs. In the case of the GHR, a simulated ensemble of protein structures was placed in an analytical nanodisc model with pre-determined parameters. The averaged theoretical scattering from the ensemble was in good agreement with the SAXS data. It is then shown how a novel method based on point-cloud models and Fast Debye Sums can be used to refine nanodisc parameters for an entire ensemble of protein structures in a more accurate and computationally efficient manner.

Furthermore, this thesis delves into a comprehensive SAS study on the interaction of $\alpha$-Synuclein and negatively charged lipid structures. The data suggest that the amphipathic properties of the protein can induce a break down of the lipid structures into smaller disc- or rod-like particles.

Detailed model-free analysis as well analytical models were used to characterise the structural transformations.

Finally, the focus is shifted away from SAS towards all atom simulations of amyloid fibrils. Experimental Φ-values were employed to guide the simulations to sample the transition state of amyloid fibril elongation. The crucial interactions sites between the incoming monomer and fibril end were identified to help shed light on the mechanisms of fibril formation.

## RESUMÉ PÅ DANSK

Proteiner spiller en mangfoldig og afgørende rolle i essentielle fysiologiske processer, hvor de indgår i et kompliceret samspil med andre proteiner og biomolekyler såsom lipider eller ligander. Det er afgørende at forstå strukturerne og mekanismerne i disse biomolekylære systemer for at forstå systemernes specifikke funktioner. Eksperimentelle teknikker bliver konstant forbedret, hvilket er vigtigt for at kunne studere så komplekse systemer. For at bygge bro mellem rådata og meningsfulde resultater er det yderst nødvendigt samtidig at udvikle nye beregningsmetoder. Det primære fokus i denne afhandling er på udvikling af analytiske modeller og integrerende beregningsværktøjer, der muliggør udnyttelse af det væld af strukturelle oplysninger, som kan udledes fra data fra småvinkelspredning (SAS). Afhandlingen undersøger også, hvordan molekylærdynamiske (MD) simuleringer af proteiner kan udbygge fortolkningen af eksperimentelle data ved at give indsigt, som ikke er tilgængelig fra eksperimenter alene.

Afhandlingen fokuserer først på udviklingen af størrelseskromatografi kombineret med røntgenstrålebaseret småvinkelspredning (SEC-SAXS) og introducerer en ny metode til at undersøge underliggende strukturelle fordelinger inden for en enkelt biomolekylær struktur. Den viser, hvordan en analytisk model kan forfines mod mange frames fra det samme SEC-SAXS-datasæt samtidigt for at give mere robuste resultater. Metoden anvendes på nanodiske. Afhandlingen undersøger også forskellige metoder til modellering af fleksible membranproteiner indlejret i nanodiske. Fleksible partikler udgør en udfordring for SAS-analyse, da spredningssignalet er gennemsnitligt over et ensemble af konformationer. En avanceret semi-analytisk model, der tager højde for konformationel diversitet, blev derfor opbygget for den humane væksthormonreceptor (GHR) i en nanodisk og forfinet i forhold til SAXS-data. Afhandlingen diskuterer derefter metoder til ensemble-modellering af membranproteiner i nanodiske. For GHR blev et simuleret ensemble af proteinstrukturer placeret i en analytisk nanodiskmodel med forudbestemte parametre. Den gennemsnitlige teoretiske spredning fra ensemblet var i god overensstemmelse med SAXS-dataene. Derefter vises det, hvordan en ny metode baseret på punktsky-modeller og Fast Debye Sums kan bruges til at forfine nanodisk-parametre for et helt ensemble af proteinstrukturer på en mere præcis og beregningsmæssigt effektiv måde.

Denne afhandling indeholder desuden et omfattende SAS-studie af interaktionen mellem $\alpha$-Synuclein og negativt ladede lipidstrukturer. Data tyder på, at proteinets amfipatiske egenskaber kan fremkalde en

nedbrydning af lipidstrukturerne til mindre skive- eller stavlignende partikler. Detaljerede model-fri analyser såvel som analytiske modeller blev brugt til at karakterisere de strukturelle ændringer.

Endelig flyttes fokus fra SAS til atomistiske simuleringer af amyloid-fibriller. Eksperimentelle $\Phi$-værdier blev brugt til at guide simuleringerne for at indsamle data, der beskriver overgangstilstanden ved forlængelse af amyloid-fibrillerne. De vigtige interaktionspunkter mellem den indkommende monomer og fibrilenden blev identificeret for at kaste lys over mekanismerne bag fibrildannelse.

**Translated with help from Thea Schulze.**

# PREFACE

My PhD was carried out over the course of 3.5 years at the University of Copenhagen. I started in the structural biophysics group at the section for Neutron and X-ray Science where I also conducted my masters thesis, with Lise Arleth and Martin Cramer Pedersen as my supervisors. The research aim was to develop analytical methods for exploiting the wealth of structural information held in small-angle scattering (SAS) data, specifically for nanodiscs and membrane proteins embedded in nanodiscs. Half-way through my PhD I moved to Kresten Lindorff-Larsen's group at the Department of Biology, where I continued to work on SAS, and was also given the opportunity to delve into molecular dynamics simulations.
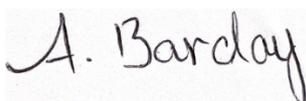
# ACKNOWLEDGMENTS

*A. Barclay*

Copenhagen, August 2023

# CONTENTS

# LIST OF ABBREVIATIONS

$\alpha$-Syn ................. $\alpha$-Synuclein

BME ................. Bayesian maximum entropy

ECD ................. Extracellular domain

FDS ................. Fast Debye sum

GFP ................. Green fluorescent protein

GHR ................. Growth hormone receptor

GHSR ................ Growth hormone secretagogue receptor

GRC ................. Gaussian random coil

ICD ................. Intracellular domain

IDP ................. Intrinsically disordered protein

IDR ................. Intrinsically disordered region

IFT ................. Indirect Fourier transform

MD ................. Molecular dynamics

MP ................. Membrane protein

MSP ................ Membrane scaffold protein

ND ................. Nanodisc

PDB ................. Protein Data Bank

$R_g$ ................. Radius of gyration

SEC ................. Size-exclusion chromatography

SANS ................ Small-angle neutron scattering

SAS ................. Small-angle scattering

SAXS ................ Small-angle x-ray scattering

TMD ................ Transmembrane domain

TS ................. Transition state

TSE ................. Transition state ensemble

SUV ................ Small unilamaller vesicles

Symbols:

$\Delta b$ ................. Excess scattering length

$\Delta \rho$ ................. Excess scattering length density

$P(q)$ ................. Form factor

$F(q)$ ................. Form factor amplitude

$D_{max}$ ................ Maximum dimension

$p(r)$ ................. Pair-distance distribution function

$I(q)$ ................. Scattering intensity

$b$ ..................... Scattering length

$\rho$ ..................... Scattering length density

$S(q)$ .................. Structure factor

# LIST OF PUBLICATIONS

This thesis is based on work presented in the following publications and manuscript. The publications and manuscripts are included in Part II, accompanied by contribution statements.

**Paper I**:

**Abigail Barclay**, Nicolai Tidemand Johansen, Frederik Grønbæk Tidemand, Lise Arleth, Martin Cramer Pedersen (2022). Global fitting of multiple data frames from SEC-SAXS to investigate the structure of next-generation nanodiscs. *Acta Crystallographica Section D*, 78, 483-493. DOI: 10.1107/S2059798322001838

**Paper II**:

Noah Kassem, Raul Araya-Secchi, Katrine Bugge, **Abigail Barclay**, Helena Steinocher, Adree Khondker, Yong Wang, Aneta J. Lenard, Jochen Bürck, Cagla Sahin, Anne S. Ulrich, Michael Landreh, Martin Cramer Pedersen, Maikel C. Rheinstädter, Per Amstrup Pedersen, Kresten Lindorff-Larsen, Lise Arleth, Birthe B. Kragelund (2021). Order and disorder - An integrative structure of the full-length human growth hormone receptor. *Science Advances*, 7, DOI: 10.1126/sciadv.abh3805

**Paper III**:

**Abigail Barclay**, Birthe B. Kragelund, Lise Arleth, Martin Cramer Pedersen (2023). Modeling of flexible membrane-bound biomolecular complexes for solution small-angle scattering. *Journal of Colloid and Interface Science*, 635, 611-621. DOI: 10.1016/j.jcis.2022.12.024

**Manuscript I**:

Céline Galvagnion, **Abigail Barclay**, Katarzyna Makasewicz, Frederik Ravnkilde Marlet, Martine Moulin, Juliette Devos, Sara Linse, Anne Martel, Lionel Porcar, Emma Sparr, Martin Cramer Pedersen, Felix Roosen-Runge, Lise Arleth, Alexander K. Buell (2023). Structural characterisation of $\alpha$-synuclein lipid-membrane interactions and the resulting aggregation using small-angle scattering. *Under review*

**Manuscript II**:

Jacob Aunstrup Larsen, **Abigail Barclay**, Nicola Vettore, Louise K. Klausen, Lena Mangels, Kresten Lindorff-Larsen, Alexander K. Buell (2023). $\Phi$-value analysis of PI3-kinase SH3 amyloid fibril elongation. *In preparation*

Part I

OVERVIEW OF THEORY AND RESEARCH

# INTRODUCTION

Proteins are the driving force behind most biological processes in the cell. Their intricate structures enable them to carry out diverse functions, from catalyzing reactions, ensuring cellular function, to transmitting signals. Understanding their structures and function in combination with other proteins or biomolecules helps us uncover the mysteries of how living organisms work. By studying proteins, we gain insights into diseases, advance biotechnologies and enrich our fundamental scientific knowledge of biology.

Particularly important is studying the many proteins which can commonly mutate and manifest in diseases. Amyloid fibrils, for example, are closely associated with neurodegenerative disorders such as Alzheimer's, Parkinson's, and Huntington's diseases (Soto and Estrada, 2008; Spillantini and Goedert, 2000; Sweeney et al., 2017), as well as type 2 diabetes (Opie, 1901; Westermark, Andersson, and Westermark, 2011). Understanding the structure and interactions of amyloid fibrils can guide the development of targeted therapies to disrupt or prevent their formation. Furthermore, the location of membrane proteins mean they are prime targets for many existing drugs which can correct abnormal signaling, transport across cell membranes, and other vital functions. Therefore, investigating the structure and function of membrane proteins can lead to the development of new therapies.

The landscape of structural biology is moving towards more complex samples, including macromolecular assemblies, dynamics, flexibility and interactions. This trend is facilitated by improving experimental techniques and the dawn of integrative structural biology, where a 'divide and conquer' strategy of breaking down convoluted problems and merging information from different sources is employed to create one cohesive and information-rich model (Araya-Secchi et al., 2023; Kim et al., 2018; Schuller et al., 2021).

Biological small-angle scattering (SAS) is one experimental technique which is growing in popularity and performance. Brighter x-ray and neutron sources, specialised SAS beamlines, standardised analysis protocols and the development of SAS data repositories like SASBDB (Valentini et al., 2015) are all contributing to the improvement of the field (Brosey and Tainer, 2019). SAS is one of the few structural techniques that can probe structures on the nanoscale under native solution conditions, including dynamics, disordered states and interacting assemblies of molecules.

However, as the complexity of samples increases, interpreting SAS data becomes less straightforward. The inverse scattering problem means there is no 'one size fits all' protocol and analysis often requires special expertise. Therefore, it is critical to continue developing computational methods that can bridge the gap between raw data and meaningful results. Usually analytical or integrative models which incorporate prior knowledge about the sample can help. Now that larger and higher-quality SAS data sets are accessible, it is important to build tools that optimise the amount of structural information that can be extracted from them.

Molecular dynamics (MD) simulations are computer techniques which can model complex biological systems at the atomic level. Newton's equations of motions are solved in order to simulate the interactions and movements of atoms. MD simulations are becoming increasingly popular in their power to predict the underlying structural distributions and dynamics of a system, which are often not possible to uncover experimentally. Experimental data can be used to guide and verify MD simulations, in order to obtain detailed and biologically relevant insights to the system.

In this thesis, some examples of complex biological systems which can only be tackled with modern or integrative methods are investigated. This includes membrane proteins with flexible domains in nanodiscs, complexes of flexible proteins and lipid structures, and amyloid fibrils. Although this thesis contains four independent projects which have resulted in five research articles, the underlying motivation is the same: to establish methods or tools that utilise experimental data to the fullest, and build detailed structural models of biomolecular complexes. I hope the red thread that connects the works will be clear. The first four research articles aim to push the limits of SAS analysis. The final research article takes a different path, where MD simulations are combined with data from protein engineering experiments.

The first section of the thesis aims to provide the background theory necessary to understand the motivation and methods behind the five research articles. Starting with Chapter 2 which gives a brief overview of protein complexes and demonstrates how we can use integrative structural biology to reveal the structure-function relationships driving cell biology. This chapter briefly covers intrinsically disordered proteins, membrane proteins, nanodiscs and amyloid proteins.

Chapter 3 introduces the basics of MD simulations. It also describes how experimental data can be utilised in simulations, particularly the attractive combination with experimental $\Phi$-values to sample the transition state of protein folding. It then discusses how the same principle can be applied to simulate the transition state of amyloid fibril elongation.

Chapter 4 focuses on SAS theory, the inverse scattering problem and what kind of information can be extracted from the data. The chapter also describes some of the exciting recent technological advancements in the field and how these will greatly benefit structural studies of biological samples.

In chapter 5, various methods for modeling SAS data from protein: lipid complexes are outlined and discussed in detail. This includes the analytical model for nanodiscs, building up to analytical models for flexible membrane proteins in nanodiscs, and finally to integrating MD simulations of membrane proteins with nanodisc models. This chapter also contains a case-study for a novel method for modelling SAS data from flexible membrane proteins in nanodiscs. This example is not included in any of the research articles but can hopefully be investigated further in the future. Finally, Chapter 5 presents analytical models which were built to model co-structures of lipids and the intrinsically disordered proteins.

Chapter 6 closes Part I with some discussion and overall conclusions of the PhD thesis as well as some future perspectives.

The main research and results are presented in Part II of the thesis in the form of five research articles, three published, one under review and one under preparation. More specifically, Paper I focuses on a very popular advancement in instrumentation at bioSAXS beamlines: size-exclusion chromatography coupled with SAXS (SEC-SAXS), where the sample is separated by particle size as the SAXS data is collected. While some simple and advanced tools exists for analysing SEC-SAXS data sets, this paper outlines a novel procedure to investigate the underlying structural distribution present in nanodisc populations. The procedure can be used to maximise the amount of structural information extracted from a SEC-SAXS data set.

Paper II demonstrates the power of integrative structural biology, as a plethora of experimental and computational techniques were used to build a high-resolution model of the human growth hormone receptor. The human growth hormone receptor represents a challenge in structural biology since it is a small membrane protein with 50 % order and 50 % disorder. Results from different experiments were combined into one structural model and a conformational ensemble of the protein was simulated by MD. The simulated ensemble was validated against SAXS data from the protein embedded in a nanodisc. Therefore, to first obtain the structural parameters of the nanodisc, a novel semi-analytical model had to be developed. (I did not perform the MD simulations in this study. My role was to model the SAS data.)

Paper III is a detailed documentation of the semi-analytical model developed for Paper II. The model combines traditional form factor modeling with the widely used spherical harmonics-based approach for

atomic structures. The framework is important since it can accommodate flexible and ordered domains, as well as a marker-protein which does not have a fixed position with respect to the rest of the system. We also investigated how the model would look for different scattering situations and discuss how experimentalists can get the most out of their SAXS or SANS experiments on protein:lipid complexes.

Manuscript I is a comprehensive SAXS and SANS study on $\alpha$-synuclein and negatively charged lipid structures. This is a challenging case for SAS analysis, where the sample undergoes large-scale morphological changes. These transformations are visible in the raw data, and then analytical models are built to extract additional information in order to characterise the resulting structures better. Studying $\alpha$-synuclein in the presence of lipids is important as it has been shown that this greatly accelerates pathogenic fibril formation, though the exact underlying mechanism is not clear.

Finally, Manuscript II investigates the mechanisms of fibril formation for PI3K-SH3 fibrils. For the first time, experimental $\Phi$-values were collected on amyloid fibrils, where the $\Phi$-value informs on whether a specific residue is structured or unstructured in the transition state. An MD simulation was performed, guided by the experimental data, to simulate the transition state ensemble of amyloid elongation. This provides far greater insight than the data can alone. The simulation helps to identify the primary contacts between the incoming monomer and the fibril end.

# BIOMOLECULAR COMPLEXES

Biomolecular complexes are assemblies of multiple molecules, such as proteins, nucleic acids (DNA/RNA) and lipids, that interact and play a crucial role in countless biological processes.These complexes are formed by non-covalent interactions, including hydrogen bonding, electrostatic interactions, hydrophobic interactions and van der Waals forces. These are dynamic structures that undergo conformational changes and interact with other molecules in response to cellular signals and environmental conditions. Studying biomolecules is essential for unraveling the complexities of life: understanding fundamental biological processes, evolution and disease mechanisms. Gaining insight into structurs on the nanoscale is not an easy task, but by combining expertise there can be far-reaching implications in fields such as drug discovery, agriculture and biotechnology.

Of particular importance are proteins, which are the workhorses driving many biological process occurring in living organisms. Proteins are intricate structures made up of smaller units called amino acid residues. An amino acid consists of a central carbon atom ($C_\alpha$), covalently bonded to a carboxyl group (COOH), an amino group (NH2), a hydrogen and a chemical group, usually called the side chain. Side chains define the chemical properties of the amino acid. Two amino acids are bound through a covalent bond, called peptide bond, between the amino and carboxyl groups. The number of amino acid residues varies widely from protein to protein. For example, human insulin, a protein that controls blood sugar levels, is among the smallest, containing only 50 residues. The giant titin protein contains around 30,000 residues and is responsible for the elasticity in muscles. The sequence of amino acids completely determines the unique 3-dimensional structure (or lack of) of each protein and therefore its specific function (Anfinsen, 1973).

## 2.1 PROTEINS

### 2.1.1 *Folded proteins*

Folded proteins spontaneously fold into well-defined 3D structures in solution. These protein structures are often referred to in terms of four levels, where each level describes a particular type of organisation, as depicted in Figure 2.1. The primary sequence is the order that the amino acid residues are linked in. Secondary structure (Linderstrøm-Lang,
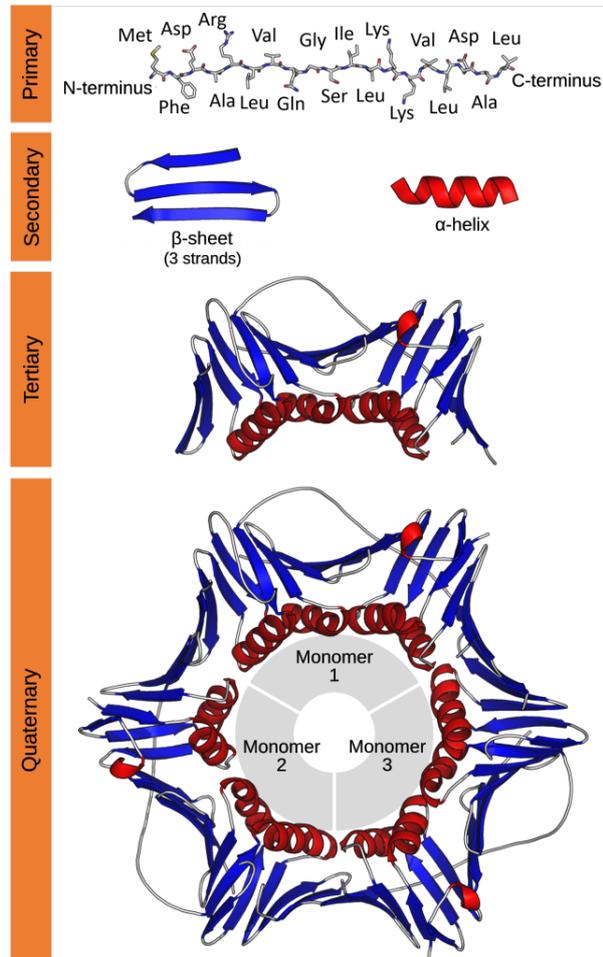
**Figure 2.1:** The four levels of protein structure: primary, secondary, tertiary and quaternary. Attribution to Jmarchn, from Thomas Shafee, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons, August 2023.

1952) refers to repeating patterns in the protein chain that are stabilised by hydrogen bonds. These patterns can be in the form of an $\alpha$-helix or a $\beta$-sheet. The overall 3D shape of the protein chain is called the tertiary structure of the protein and often dictates the function of the protein. The quaternary structure is formed if more than one polypeptide chain assembles into one unit. The arrangement and interactions of these subunits contribute to the overall function of the protein complex (Berg, Tymoczko, and Stryer, 2002).

Predicting the structure of a protein from its primary sequence is a challenging problem. However, the 'structure-function' relationship is very important for understanding the workings of living organisms. For instance, enzymes have evolved to catalyze chemical reactions by binding substrates to their precisely shaped active sites. Any disruption to this structure may prevent the enzyme from binding the correct substrate, leading to alterations or inhibition of its function. Furthermore, protein related diseases often arise from mutations and misfolding. By studying structure and interactions, the molecular basis of diseases can be uncovered and potential drug targets can be identified to pave the way for the design of new medicines.

Historically, x-ray crystallography has been an extremely important method for determining the atomic structure of proteins (Dauter and Wlodawer, 2016). The first atomic protein structures of myoglobin and haemoglobin were determined successively by John Kendrew and Max Perutz in 1958 and 1960, respectively. Kendrew and Perutz went on to win the Nobel prize in chemistry. The discovery of protein diffraction revolutionised molecular biology, giving the first insights into the role of the arrangement of amino acids and opening the door to understanding the structure-function relationships. X-ray crystallography continues to be a foundational technique in structural biology, but the field is evolving. X-ray crystallography requires high-quality crystals which often can be extremely challenging to produce for certain protein systems, as will be discussed in the following sections.

### 2.1.2 *Intrinsically disordered proteins*

Intrinsically disordered proteins (IDPs) and regions (IDRs) lack the rigid, fixed 3D structure that folded proteins posses. IDPs can adopt many, many conformations, make transient interactions and exhibit dynamic behaviour, as illustrated in Figure 2.2. The discovery of IDPs in the late 90s challenged the traditional 'lock and key' notion that proteins have to be well-folded in their unique structure to perform its function (Uversky, Gillespie, and Fink, 2000). However, it has become clear that the high conformational flexibility of IDPs is to their own advantage, and they can engage in functions that folded proteins would not be able to (Uversky, 2019). IDPs are particularly prevalent in signaling
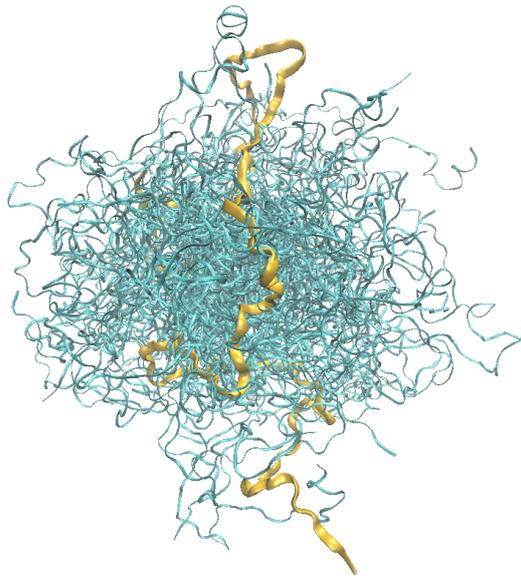
**Figure 2.2:** Spaghetti-like behaviour of intrinsically disordered proteins. The structural ensemble is of a hybrid synuclein protein and was generated with MD simulations (Allison et al., 2014). The ensemble is entry 00006 from the Protein Ensemble Database.

and regulatory pathways where the protein often needs to interact with multiple binding partners (Wright and Dyson, 2015).

Since IDPs and multidomain proteins with IDRs cannot not be described with a single set of coordinates, more advanced experiments and analysis are needed to characterise their ensemble of structures. Flexible domains do not produce coherent x-rays for x-ray crystallography and are poorly resolved with cryo-EM. Integrative methods, particularly experiments combined with simulations, are gradually making IDPs more accessible (Evans et al., 2023).

### 2.1.3 *Membranes and membrane proteins*

Membrane proteins (MPs) are another family of proteins that should be tackled with interdisciplinary methods, since they pose their own challenges in terms of experimentation and modelling.

Integral membrane proteins are permanently attached to the membrane, including monotopic proteins which are permanently attached to one side, and transmembrane proteins which span the membrane one or more times, as illustrated Figure 2.3. . These proteins have hydrophobic regions which like to be shielded from aqueous environments by interacting with the hydrophobic interior of the membrane. Peripheral membrane proteins are only temporally attached to the membrane. They
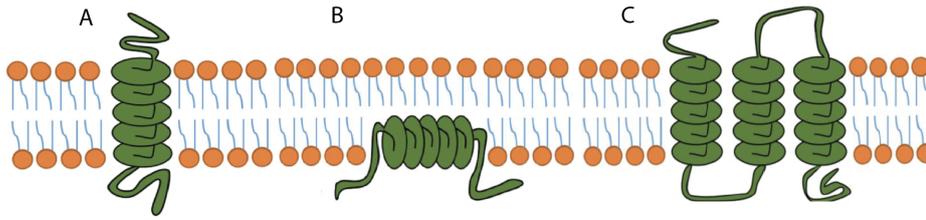
**Figure 2.3:** Examples of membrane proteins. (A) a single-pass transmembrane protein. (B) a a monotopic protein (C) a multi-pass transmembrane protein where the polypeptide chain spans the membrane two or more times. Figure is adapted from Zhou, Wang, and Yuan, 2022 under https://creativecommons.org/licenses/by/4.0/

either interact with the membrane by binding with integral membrane proteins or by associating with the lipid polar headgroups.

MPs are important for communication between cells and for transporting ions, water molecules, nutrients, and other substrates and drugs through special channels. Membrane proteins are therefore among the most common targets for therapeutic drugs. However, structural studies of MPs are notoriously challenging since they must be held in close-to native lipid environments in order to remain folded in their physiological form. This makes them more difficult to crystallise than soluble proteins. Therefore, other experimental techniques must be employed and a membrane mimetic system must be chosen on a case-by-case basis. While MPs make up $\sim$ 30% of the human proteome, they are underrepresented with less than 3% of entries in the Protein Data Bank (PDB) being annotated with 'membrane protein' (https://www.rcsb.org/, 2023). However, the landscape of membrane proteins in structural biology is improving. 50% of MP structures in the PDB were deposited in the last three years alone. This surge can be attributed to the emergence of cryo-EM which is now the dominant experimental method for solving the structure of membrane proteins down to atomic resolution.

For cryo-EM, MPs first need to be solubilised in lipids or other amphipathic molecules. The complexes are flash frozen to preserve their native state, and 2D micrographs are collected and then computationally manipulated to form a 3D representation. The final protein model represents an ensemble average from across the grid and, due to their heterogeneity compared to the rigid protein structure, the signal from the carrier-system is averaged away. This also means cryo-EM cannot capture flexible regions of MPs.

**Membrane mimetics**

In order to remain soluble and in a close-to native state, membrane proteins must be held in carrier systems for structural and biophysical studies. This adds a layer of complexity to sample handling and, often, data interpretation. All carrier systems are made up of amphiphilic
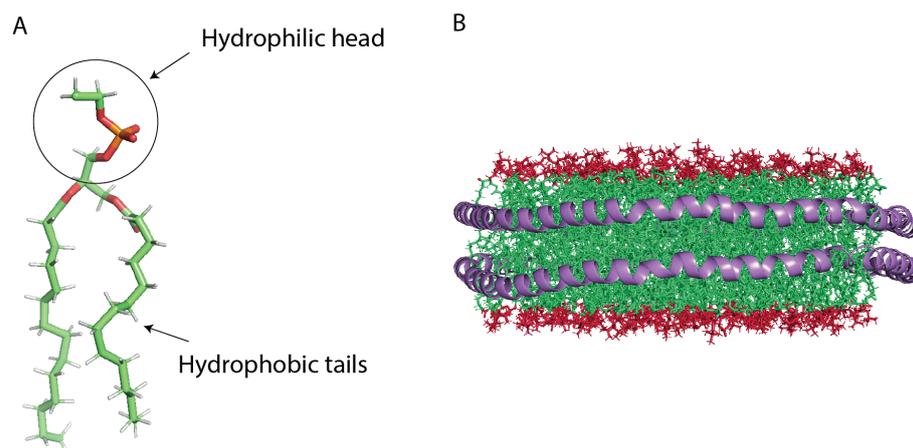
**Figure 2.4:** (A) Molecular model of a DMPC lipid. The atomic color coding is white: hydrogen, red: oxygen and orange: phosphate. (B) Molecular model of a nanodisc with DMPC lipids and MSP1D1Δh5 represented in purple, built with CHARMM-GUI nanodisc builder (Qi et al., 2019).

molecules which self-assemble to shield the hydrophobic transmembrane domains of MPs, usually detergents or lipids (Johansen et al., 2023).

Lipids are the primary building blocks for cell membranes (Figure 2.4). In aqueous environments, they spontaneously form stable bilayers or micelles, where their hydrophilic heads face outward, interacting with water, and their hydrophobic tails are shielded in the interior away from water.

Membrane scaffold proteins (MSPs) are derivatives of the Apo-A1 protein. MSP nanodiscs (NDs) are stable discoidal bilayer patches, where the lipid tails are encompassed by two MSPs stacked on top of each other (Bayburt, Grinkova, and Sligar, 2002). An example is shown in Figure 2.4B. When a MP is added to the mix, the system self-assembles to encompass the MP inside the ND. Since the utility of NDs was first demonstrated, their popularity has skyrocketed for structural and functional experiments of MPs (Sligar and Denisov, 2021). One of the main advantages of MSP NDs is that they form highly homogeneous and stable particles compared to other types of carrier system. Their size is controlled by the length of the MSP, so they can accommodate MPs of different sizes.

In the recent years, their stability and homogeneity was further improved. Circularised NDs see the MSP's N- and C-terminal covalently linked (Nasr et al., 2017), and supercharged NDs contain solubility-increasing mutations in the MSP (Johansen et al., 2019). Taken together, circularised and supercharged nanodiscs (csND) are particularly advantageous for SAS studies, which is a technique extremely sensitive to size and shape polydispersity of the sample.
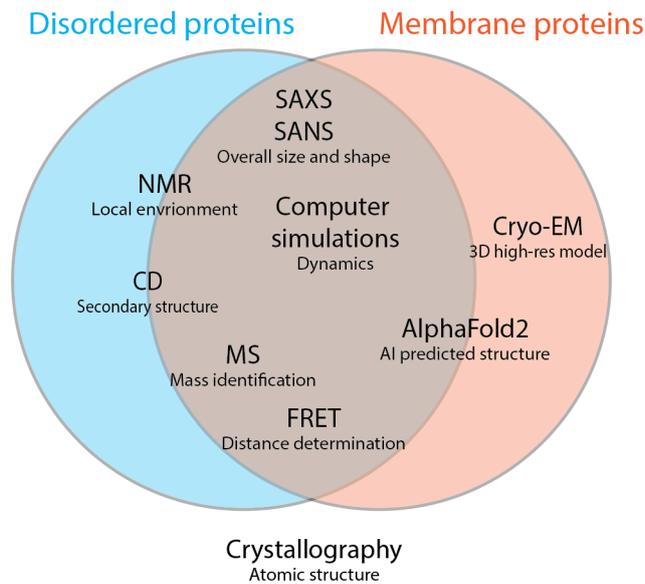
**Figure 2.5:** A non-exhaustive list of the diverse array of techniques in integrative structural biology that can be applied to study membrane proteins and/or disordered proteins. Crystallography is outside of the diagram to illustrate that it is generally not applicable to the study of either protein family. NMR and CD are placed on the edge of MPs since, although still valuable, there are often challenges associated with these techniques when applied to membrane proteins. While cryo-EM is pioneering structure determination of membrane proteins, it is limited when it comes to flexible regions. AlphaFold2 is placed on the edge of disordered proteins, since it can do little more than predict which regions are disordered.

### 2.1.4   *Integrative structural biology*

There are many experimental techniques available to study biomolecular complexes. Each technique offers unique insights but cannot give a full picture on structure, dynamics and interactions. Integrative structural biology is when information from more than one technique is combined to piece together one coherent and accurate picture of the molecule (Ward, Sali, and Wilson, 2013). Usually a computational model is built, including information from multiple sources, to bridge the gap between experiments and results.

Among the first leaps into integrative structural biology came from Alber et al., 2007, who used data from seven different sources as restraints to build a structure of the Nuclear Pore Complex. The complex, which contains 400 protein subunits and a high degree of flexibility, represented a significant challenge for conventional structure determination methods. The study included sedimentation analysis for information on the shape of the isolated protein units, overlay assays to identify protein pairs which interact, and electron microscopy to provide the overall shape and

symmetry of the complex. A bead-based computational model could then be optimised to satisfy the criteria from all of the experiments and provide a unified model of the structure. Since then, the 'divide and conquer' approach has gained momentum and provided novel insights into a number of biomolecular complexes (Araya-Secchi et al., 2023; Farrell et al., 2020; Johansen et al., 2022; Kim et al., 2018; Schuller et al., 2021; Stella et al., 2018; Whitford et al., 2011).

IDPs and membrane proteins are particularly intriguing cases for integrative structural biology. Many MPs are not rigid in nature or contain IDRs and therefore also require techniques which can probe conformational diversity (Wang et al., 2018). For example, Nuclear Magnetic Resonance (NMR) reports on the local environment of nuclei in the protein and is highly effective for studying the dynamics of IDPs and IDRs (Jensen, Ruigrok, and Blackledge, 2013; Prestel et al., 2018). NMR can also be used to study the structure and dynamics of MPs in membrane mimetics, as long as the complex is small enough (Andreas et al., 2015; Cho et al., 2014). Furthermore, paramagnetic relaxation enhancement (PRE) can even provide information on MP orientations and the depth of insertion in a membrane (Danmaliki and Hwang, 2020).

SAS can easily differentiate between folded and flexible proteins, and give information on the overall size and arrangements of IDPs and proteins containing IDRs (Bernado and Svergun, 2012; Kikhney and Svergun, 2015). The SAS signal from flexible regions is not reduced, like in other techniques. SAS is well-suited to studying MPs in carrier systems (Denisov and Sligar, 2016). Although, SAS provides low-resolution information on the average shape and size of the complex, it has some advantages compared to cryo-EM for MPs. The particles are held in solution meaning the sample preparation is more practical and data is recorded in a native environment. SAS can also be used to detect how proteins react to changes in environment, for example induced by adding binding partners or changes in temperature or pH.

Both SAS and NMR report on ensemble averaged properties from a very large number of conformations in the sample. Combining computer simulations, such as as molecular dynamics (MD) simulations, with these techniques offers atomic-level insights into the specific properties and distributions underlying IDPs and other non-rigid systems (Henriques, Cragnell, and Skepo, 2015). The conformational ensemble obtained from MD should always be validated against experimental data (Orioli et al., 2020; Thomasen et al., 2023), as will be discussed in Chapter 3. Simulations can also help to integrate data from different experiments to obtain a cohesive overview of an ensemble, e.g. by combining overall shape and size information from SAXS and secondary structure propensity from NMR (Gomes et al., 2020; Mertens and Svergun, 2017).

Force fields are available which are compatible with lipids and proteins, and thus all-atom and course-grain simulations of MPs in lipid bilayers, (Goossens and De Winter, 2018) and even in nanodiscs (López et al., 2019), can be performed.

For both IDPs and MPs with IDRs, single-molecule approaches, such as Förster Resonance Energy Transfer (FRET), can deliver insights into nanoscale distances and dynamics. Advanced computational tools are emerging to validate MD simulations against experimental FRET data (Lerner et al., 2021; Montepietra et al., 2023).

Another important experimental technique for studying flexible proteins is circular dichroism (CD), which gives a signal unique to different secondary structural elements. CD is useful for studying IDP dynamics as IDPs often move through $\alpha$-helices and $\beta$-sheets as they change conformation. CD has been applied to structural studies of MPs in various carrier systems, although challenges arise since the carrier system itself can give rise to artifacts (Wallace, 2010). Native mass spectrometery (MS) measures mass-to-charge ($m/z$) ratios, providing insights into protein composition, structure, and interactions. MS is applicable including any membrane assemblies.

Any thesis on structural biology in 2023 would not be complete without a mention of AlphaFold2 (Jumper et al., 2021). AlphaFold2 is an artificial intelligence structure prediction tool which can accurately predict the 3D structure of two thirds of the human proteome (Tunyasuvunakool et al., 2021). Furthermore, often regions in proteins that are predicted by AlphaFold2 with very low confidence can be attributed to flexible or dynamic regions (Ruff and Pappu, 2021). The performance of AlphaFold2 on membrane proteins is excellent. This is impressive and somewhat surprising, considering that AlphaFold2 does not explicitly consider lipid bilayers in its predictions and was trained on the few MPs present in the PDB (Hegedűs et al., 2022). Structures from AlphaFold2 can be used directly as structural models or indirectly as aids for experiments.

However, the pursuit of structural biology is far from over. AlphaFold2 cannot give information about the mechanisms behind protein folding. It does not provide information on protein misfolding, stability or dynamics. Furthermore, the prediction accuracy is low for protein complexes with more than two subunits (Bryant et al., 2022). Therefore, AlphaFold2 should be added to the structural biology toolbox, while the field continues to push experiments and integrative methods to study complex biological systems.

Since the first structure of the Nuclear Pore Complex was published (Alber et al., 2007), progress in experimental techniques and integrative methods means the structure has been modelled with increasingly precise data, ultimately resulting in near-atomic models (Akey et al.,
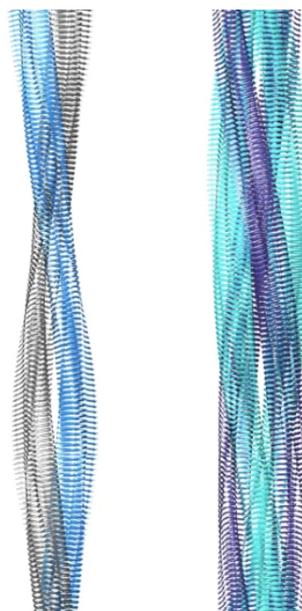
**Figure 2.6:** Cryo-EM structures of two morphologies of amyloid fibrils from serum amyloid A protein (two or four intertwined filaments). Figure is from Bansal et al., 2021 without changes under https://creativecommons.org/licenses/by/4.0/

2022; Kim et al., 2018; Mosalaganti et al., 2022; Petrovic et al., 2022). One of these studies which exemplifies how cutting-edge computational techniques can be employed to accelerate the structural determination of large biomolecular complexes is by Mosalaganti et al., 2022. In the study they used AlphaFold to generate models of single proteins and subcomplexes of the Nuclear Pore Complex. The AlphaFold structures were first validated against x-ray crystallography and cryo-EM structures, and then assembled to fit cryo-EM density maps of the entire complex. The resulting model covered various domains which were previously structurally uncharacterised, which enabled a course-grained MD simulation of the Nuclear Pore Complex scaffold to capture its dilation and constriction movements.

To conclude, there are many methods capable of investigating both membrane proteins and flexible proteins and there is more power in combining more methods. Membrane proteins with disordered regions are among the most ambitious samples to investigate in structural biology. Figure 2.5 showcases integrative approaches which could be used in unraveling the complexities of these kinds of biomolecules.

## 2.2   AMYLOID FIBRILS

Amyloid fibrils are formed by the abnormal self-association of proteins into large, highly stable, fibrillar structures (Dobson, 2003). Over 50 different proteins form amyloid fibrils which are associated with a range

of major diseases, including Alzheimer's, Parkinson's and Huntington's disease (Soto and Estrada, 2008; Spillantini and Goedert, 2000; Sweeney et al., 2017),. The amyloid fold is usually substantially different in structure to the native fold of the protein. However, over the last two decades, non-toxic amyloids with a biological functional role have been identified as the sole native fold for some proteins, for example as structural components in biofilms (Erskine, MacPhee, and Stanley-Wall, 2018; Van Gerven et al., 2018). Amyloid fibrils are structurally diverse and the same protein has been shown to aggregate into a variety of species that differ in size and morphology under different conditions (Bansal et al., 2021; Gosal et al., 2005; Radamaker et al., 2021) (Figure 2.6). A detailed understanding of the fibrillation mechanism is therefore of great importance to understanding amyloid disease development and what determines cytotoxicity. While ensemble techniques such as ThT fluorescence, NMR and CD can inform of the kinetic properties of amyloid fibrillation, cryo-EM is leading the way for imaging static structures of amyloid (Gremer et al., 2017; Iadanza et al., 2018; Radamaker et al., 2019) and pre-amyloid (Claridge et al., 2023) states.

### 2.2.1  *Amyloid fibril formation*

Amyloid formation is typically understood as a nucleation-dependent mechanism, where monomers undergo a structural reorganisation and assemble into a protofibril nucleus. The protofibril nucleus is thought to be thermodynamically unstable and therefore nucleation is the rate-limiting step. The nucleus then serves as a platform for the rapid elongation of monomers into amyloid fibrils, without forming stable intermediates. Vettore and Buell, 2019 showed that amyloid elongation is a highly cooperative process for their systems. They showed that a cooperative model, where there is a slower rate constant for nucleation and a higher rate constant for elongation, is a successful model for describing amyloid fibrillation. The three phases, nucleation, elongation and thermodynamic equilibrium, are illustrated in Figure 2.7.

### 2.2.2  *α-Synuclein*

α-Synuclein (α-Syn) is an abundant small protein in the brain, composed of three domains: an N-terminal lipid-binding domain, a non-amyloid-component (NAC) and a negatively charged C-terminal tail. α-Syn appears monomeric and intrinsically disordered in solution but adopts an α-helical structure in the N-terminal upon binding to a lipid membrane. The exact physiological function of α-Syn remains unknown, however, it is involved in synaptic activity (Ghiglieri, Calabrese, and Calabresi, 2018). The NAC domain is highly hydrophobic, giving its high propensity for aggregation and leading to fibril formation. Although monomeric α-Syn is non-toxic, the presence of α-Syn fibrils in Lewy bod-
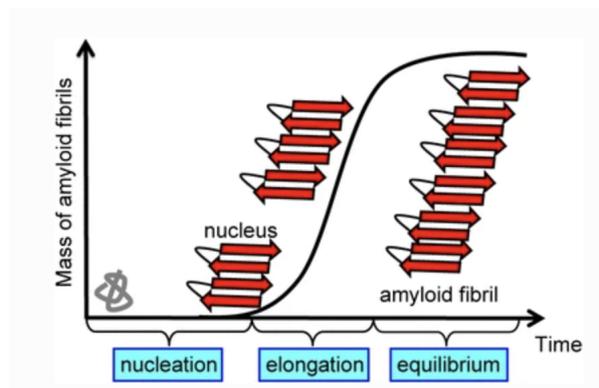
**Figure 2.7:** Model for the kinetics of amyloid formation. Figure from Chatani and Yamamoto, 2018 with permission from the publisher, *Springer Nature*.

ies is a hallmark of Parkison's disease (Neumann et al., 2004; Peelaerts et al., 2015). Over-expression of the protein seems to lead to accumulation, first into prefibrillar forms and then into larger aggregates which are toxic, possibly because of interaction with vesicles or disrupting membrane curvature. Current therapeutic approaches for Parkinon's disease focus on inhibiting $\alpha$-Syn aggregation or using small molecules to dissociate aggregates. However, further research is required to understand $\alpha$-Syn pathology, starting with finding the precise underlying mechanisms which lead to fibrillation and the progression of the disease (Fields, Bengoa-Vergniory, and Wade-Martins, 2019).

Earlier this year Ray et al., 2023 showed that at high concentrations (>100 $\mu M$), $\alpha$-Syn fibrils can form inside condensate droplets via liquid-liquid phase separation under physiological conditions. However, there are multiple pathways leading to $\alpha$-Syn amyloid formation. The better studied way is by nucleation on lipid surfaces (Auluck, Caraveo, and Lindquist, 2010; Butterfield and Lashuel, 2010; Fink, 2006). In fact Galvagnion et al., 2015 reported, that the presence of small unilamellar vesicles (SUV) of DMPS accelerates fibril formation by three orders of magnitude, compared to in bulk solution. They propose a model where, at high $\alpha$-syn:SUV ratios, the primary nucleation occurs on the surface of the vesicle where fibrils subsequently build on top. Interestingly not all of the monomeric $\alpha$-Syn is consumed. The amount that is converted to fibrils is directly proportional with SUV concentration. At low $\alpha$-Syn:SUV ratios, fibrils do not form, indicating lipid-bound $\alpha$-Syn is more thermodynamically stable than the fibrillar state under those conditions. However, the exact mechanism between protein and lipids which facilitates the aggregation process is not properly understood.

Electrostatic interactions between negatively charged polar lipid headgroups and the positively charged N-terminal of $\alpha$-Syn promote binding to the outer leaflet of anionic vesicles (Cholak et al., 2019; Davidson et al., 1998; Galvagnion et al., 2015; Middleton and Rhoades, 2010). However,

$\alpha$-Syn has also been shown to remodel anionic PG membranes into other kinds of structure such as cylindrical micelles (Jiang et al., 2018; Mizuno et al., 2012) and bilayer discs (Varkey et al., 2013), which in some cases even promoted fibrillation. The amino acid sequence for $\alpha$-Syn contains seven 11-residue repeats that are predicted to form amphipathic $\alpha$-helices that mediate its interaction with membranes; in this respect, it is reminiscent of apolipoprotein. Apolipoprotein has also been shown to have the ability to induce membrane curvature, as well as spontaneously stabilise disk-shaped bilayer patches by shielding the hydrophobic interior from solvent (Varkey et al., 2010). A similar phenomena have been observed for an amyloid-$\beta$ peptide, where the peptide was shown to induce a disruption of vesicles into discoidal structures (Ivankov et al., 2021).

Membrane fluidity and lipid alcyl chain length affects the propensity for $\alpha$-Syn fibrillation, as it has been shown that DLPS (12:0 (12 carbons in the alcyl chains, 0 double bonds)) has an increased aggregation rate compared to DMPS (14:0), and DOPS (18:1) does not facilitate aggregation at all (Galvagnion et al., 2016). This can be put down to increased solubility of shorter acyl chains. This shows that lipids type and possibly membrane remodelling could play a larger role in the fibrillation process than just providing a stable interface.

Hoover et al., 2021 investigated the disruption of DLPS large unilamellar vesicles upon mixing with $\alpha$-Syn via transmission emlectron microscopy, and observed a change in morphology to discs, tubes and ribbons over several hours. The same phenomena is studied with an extensive set of SAXS and SANS experiments in Manuscript I.

# 3

## MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics simulations have emerged as a powerful computational method in structural biology (Hollingsworth and Dror, 2018; McCammon, Gelin, and Karplus, 1977; Perilla et al., 2015). Through MD, the motions and physical properties of biomolecular systems can be studied with a high-level of detail, which is often not accessible through experiments where the temporal- and spatial-resolution is limited. While almost all experimental techniques report on the ensemble average of the sample, the ensemble average is not necessarily representative of the individual molecules. MD simulations provide distributions and time-series of conformations which gives invaluable insight into the underlying biophyscial properties of a protein or biomolecular complex. MD simulations can study a wide-range of protein dynamics, from individual atom fluctuations, side-chain fluctuations, flexible movements, and large-scale conformational changes (Zwier and Chong, 2010).

For example, Heller et al., 2017 used MD to simulate the binding of a small molecule to a disordered peptide and elucidated sequence specificity associated with the binding mechanism. Large flexible nucleosomes have been simulated with MD to study the dynamics of mobile subunits and DNA within the complex (Roccatano, Barthel, and Zacharias, 2007). Conformational ensembles of the antibody light chain, which has two domains connected by a flexible linker, show that certain inter-domain contacts stabilise the relative orientration of the two domains which is impacts amyloidogenicity (Weber et al., 2018). As a final example, Dedmon et al., 2005 simulated the conformational ensemble of $\alpha$-Syn which revealed that the native state is composed of a more compact ensemble than would be expected for a random coil state. They also identified specific residues involved in the contacts formed between the C-terminus and NAC regions.

MD simulations employ Newton's equations of motion to simulate the interactions and movements of atoms in a biomolecular system. The force field is a set of mathematical equations and parameters which describe the potential energy between atoms. Newton's equations are integrated to update the atomic positions in time and explore the conformational landscape of the system. The background presented in this section is mainly taken from Leach, 2001.

Of course, the accuracy of the simulated trajectory is dependent on, and limited by, the quality of the starting models, inaccuracies in the force
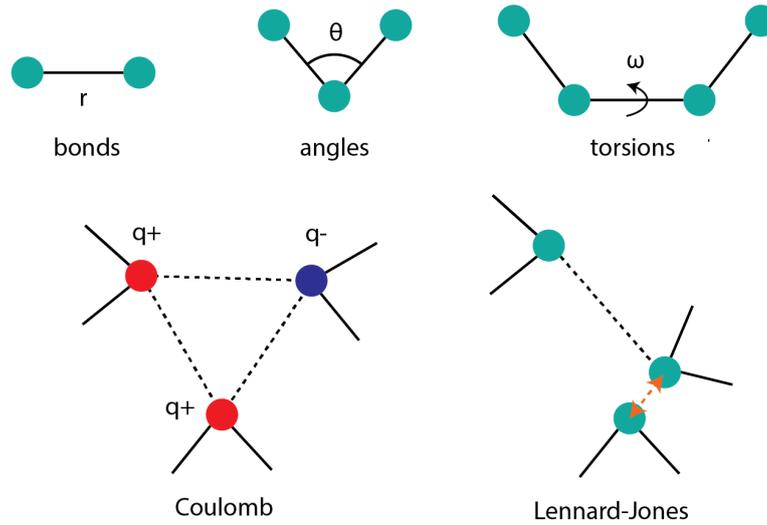
**Figure 3.1:** Schematic representation of the five terms contributing to the functional form of a force field. Bonded terms are on the top and are represented with solid lines. Non-bonded terms are on the bottom and are represented with dotted lines. The orange arrow in the Lennard-Jones diagram indicates repulsion between two atoms which are too close to each other. No line indicates no interaction between atoms that are too far apart.

field and insufficient sampling. The results from simulations should always be verified by experimental data, either by incorporating experimental data into the simulation as restraints, or by calculating observables from the trajectory afterwards and comparing it to the data (Orioli et al., 2020). Furthermore, coarse-grained simulations are a possibility for simplifying the system's complexity by representing groups of atoms as single particles. However, this approach comes with a trade-off in terms of accuracy and resolution (Ingólfsson et al., 2014).

## 3.1 SIMULATING PROTEINS

MD simulations start with an initial atomic structure of the biomolecular system. Understanding how the atoms interact with each other, the forces that they attract and repel each other with, is the key to understanding how the structure evolves. The force $\vec{F}$ on atom $i$ is the derivative:

$$\vec{F}_i = -\frac{\partial U}{\partial \vec{r}_i} \qquad (3.1)$$

where the energy $U$ depends on the atom type and the position with respect to the origin $\vec{r}_i = (\vec{x}_i, \vec{y}_i, \vec{z}_i)$ of the atom.

The total interaction energy of a system can be described as the sum of atomic interactions. The interaction energy for individual atoms in the system can be described with:

$$U = \underbrace{U_{\text{bond}} + U_{\text{angle}} + U_{\text{torsions}}}_{\text{bonded}} + \underbrace{U_{\text{Coulomb}} + U_{\text{Van der Waals}}}_{\text{non-bonded}} \qquad (3.2)$$

Where bonded interactions are between atom neighbours in the same molecule and non-bonded interactions are between all atoms in the system. The Coulomb force accounts for long-range electrostatic attraction and repulsion, while the Van der Waals force accounts for weaker, short-range interactions and is described with the Lennard-Jones potential.

The functional form of a force field that can be used to model assemblies of $N$ atoms is therefore:

$$U(\vec{r}) = \sum_{\text{bonds}} \frac{a_i}{2}(r_i - r_{i,0})^2 + \sum_{\text{angles}} \frac{b_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{c_i}{2}(1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N} \sum_{j=1+1}^{N} \left( \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[ \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \right)$$

$$(3.3)$$

Where the first and second terms describe the deformation energies in bond length $r$ and bond angle $\theta$ from the equilibrium position $r_0$ and $\theta_0$ respectively. The third term describes the deformation energy from rotations around the chemical bond where $n$ is the periodicity. $\epsilon_i j$ and $\sigma_i j$ describe the Lennard-Jones potential.

Force fields differ in their parameterisation, e.g. in the various constants such as $a_i, b_i$ and $c_i$ which denote the strength of an interaction depending on the atom type. Force fields are developed empirically, mostly through quantum mechanical calculations and experimental data like x-ray diffraction or infrared spectroscopy, to obtain bond lengths, angles, molecular vibrations etc. Force fields are constantly being updated by optimisation against experimental data, but force field parameters should not be system-specific. The same parameters should be able to model a series of related molecules, so that predictions on new systems can be made. Still, some force fields are better at capturing the properties of certain kinds of structures than others, e.g. $\alpha$-helical, intrinsically disordered, nucleic acids etc. Therefore the choice of force field may be considered on a case-by-case basis depending on the system of interest (Robustelli, Piana, and Shaw, 2018).

### 3.1.1 *Time steps*

In order to describe the dynamics of the system, the coordinates and momenta, $\vec{p} = m\vec{v}$ of all atoms must be known. The temperature determines the kinetic energy of the system. The final equation needed to run a simulation is Newton's second law of motion:

$$\vec{F_i}(t) = m_i \vec{a_i} = m_i \frac{\partial^2 \vec{r_i}(t)}{\partial t^2} \tag{3.4}$$

By solving this equation for the acceleration $a$ of each atom, the positions and momenta can updated with each time step. The function $\vec{F}$ is highly non-linear and therefore the differential equations cannot be solved analytically. Different algorithms exist to integrate Newton's equations numerically. In all of them, time is discretized by dividing it into small intervals referred to as time steps, $\Delta t$.

If the time step is too small, computational resources are wasted. If it is too large, the proper dynamics of the system will not be observed and errors are encountered. Time steps of 1 or 2 fs are usually used, slightly shorter than the fastest vibrations in biomolecular simulations. Therefore, simulations are limited by computational power. Typically, simulations may reach μ second length scales, requiring on the order of $10^8$ to $10^9$ steps, but this depends on the properties of the system that are under investigation and the size of the conformational space.

## 3.2 BIASING SIMULATIONS

As mentioned, force fields are mathematical approximations of the interactions in biomolecular systems. Accuracy in simulations must be balanced with computational simplicity, and therefore MD simulations cannot be expected to perfectly recreate the biophysical properties of the system. Methods have been developed to improve the synergy of simulated conformational ensembles with experiments in system-dependent manors (Hummer and Köfinger, 2015; Orioli et al., 2020).

### 3.2.1 *BME Reweighting*

MD simulations provide an ensemble of conformations, while usually only the average of a very large number of molecules can be obtained experimentally. By reweighting a trajectory, the consistency between average properties back-calculated from the trajectory, $\langle F^{calc} \rangle$, and experimental data, $F^{exp}$, can be improved. The initial trajectory usually has a uniform set of weights for each frame i.e. each frame has $w_i = 1/N$ where $N$ is the number of frames in the trajectory. In the reweighting process, the individual weights are altered until the optimal agreement with the experimental data is reached.
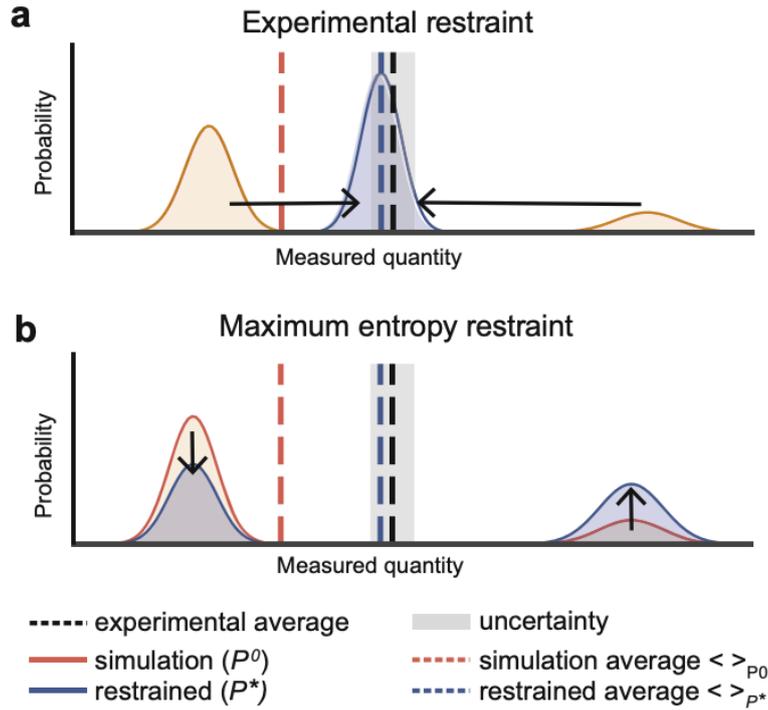
**Figure 3.2:** The benefits of maximum entropy reweighting. (A) Shows a situation where only the experimental data is used. Only conformations which match the experiments are considered, thereby ignoring the two-state distribution predicted by the simulation. (B) Shows the ME approach, where the prior distributions are minimally modified to get the calculated average to match the experimental data. Figure from Bottaro, Bengtsen, and Lindorff-Larsen, 2020 with permission from the publisher, *Springer Nature*.

The caveat is that many sets of weights can lead to the same $\langle F^{calc} \rangle$ and therefore the same level of agreement with the experimental data. Optimising the weights in an unconstrained way can lead to unrealistic underlying distributions of the system.

One way to tackle this is the Bayesian/MaxEnt approach (BME) (Bottaro, Bengtsen, and Lindorff-Larsen, 2020). The maximum entropy (MaxEnt) part of the approach states the optimal solution is the one adding the least amount of information, i.e. the solution with minimal deviation from the initial set of weights, while still being in agreement with the data. The Bayesian part of the approach considers the experimental errors also, with the intention of preventing overfitting. Therefore, during BME the following equation should be minimised:

$$\mathcal{L}(w_1...w_n) = \chi_R^2(w_1...w_n) - \theta S(w_1...w_n) \tag{3.5}$$

where $\chi^2_R$ quantifies the agreement with the data if $n$ is the number of frames in the trajectory, $m$ is the number of experimental data points and $\sigma^{\text{exp}}$ is the uncertainty associated with the experimental data point:

$$\chi^2_R(w_1...w_n) = \frac{1}{m} \sum_i^m \frac{(\sum_j^n w_j F_i - F_i^{\text{exp}})^2}{(\sigma_i^{\text{exp}})^2} \qquad (3.6)$$

and the relative entropy:

$$S = - \sum_j^n w_j \log(\frac{w_j}{w_0}) \qquad (3.7)$$

measures the deviation from the initial weights. The $\theta$ parameter is tuned in accordance with how confident we are in the initial distributions. If $\theta$ is very high, the initial distributions are minimally perturbed and if $\theta$ is very low, the weights become less uniform to minimise the discrepancy with the data. The optimal value for $\theta$ should be systematically chosen on a case-by-case basis.

BME can be used in combination with many different types of data, including RMSDs from solvent-accessible surface area (SASA) data (Xie and Frank, 2021), chemical shift perturbations (Crehuet et al., 2019) and relaxation data (Kummerer et al., 2021) from NMR experiments, and radii of gyration ($R_g$) from small-angle scattering (Ahmed, Crehuet, and Lindorff-Larsen, 2020; Pesce and Lindorff-Larsen, 2021). When it come to SAS, if the data is information rich, it is beneficial to reweight against the entire SAS profile directly (Larsen et al., 2020; Thomasen and Lindorff-Larsen, 2022). Furthermore, a trajectory can be reweighted against multiple sources of data simultaneously to ensure confidence in more than one property of the system, for example SAXS and SANS at different contrast situations (Larsen et al., 2020), SAXS and NMR chemical shifts (Gomes et al., 2020) or $R_g$ and $R_h$ (Choy et al., 2002).

### 3.2.2 *Transition state simulations for protein folding*

A second way to integrate experimental data with MD simulations is to use the data in the simulation directly. By adding a bias to the underlying force field, the simulation can be guided to focus sampling efforts only on the most relevant regions of the conformational landscape. In particular interest for this thesis, restraints based on experimental Φ-values (Fersht and Daggett, 2002) can be used to simulate the transition state ensemble (TSE) of proteins.

The transition state (TS) in protein folding is the structure (or small ensemble of structures) which sit at the top of a free-energy barrier dividing the unfolded and folded states. Reaching the TS is the rate-limiting step in protein folding, but it is only visited briefly and its

structure cannot be probed by experiment directly. The protein engineering approach can give structural information about individual residues in the TS through kinetic and thermodynamic experiments on a series of point mutations (Matouschek et al., 1989). For each mutation, the information is usually presented as a $\Phi$-value between 0 and 1. If $\Phi \approx 0$, the residue is unstructured and has few or no interactions in the TS. If $\Phi \approx 1$, the residue is as structured in the TS as it is in the native state. Fractional $\Phi$-values can be interpreted as partially formed native interactions in a single folding pathway (Davis, Dobson, and Vendruscolo, 2002).

It has been demonstrated many times that MD simulations guided by experimental $\Phi$-values can generate molecular models of the TSE and give unique insight into the process of protein folding (Gsponer et al., 2006; Gsponer and Caflisch, 2002; Paissoni et al., 2021; Sato et al., 2004). Vendruscolo et al., 2001 first utilised the 'native contacts approximation', that $\Phi$-values can be interpreted in terms of the loss of native state contacts:

$$\Phi^{\text{calc}} = \frac{N_{\ddagger}}{N_N} \tag{3.8}$$

where $N_{\ddagger}$ and $N_N$ are the number of native atom-atom contacts in the TS and the native state, respectively. $\Phi^{\text{calc}}$ is calculated at each time step in the simulation, and a psuedo-energy term added to the force field minimises the discrepancy:

$$\rho = \frac{1}{N} \sum \kappa (\Phi_i^{\text{calc}} - \Phi_i^{\text{exp}})^2 \tag{3.9}$$

where $\kappa$ is the force applied to the restraint. The 'native contact approximation' is computationally inexpensive compared to calculating thermodynamics properties of the system. With this procedure, an ensemble of conformations which are compatible with the experimental data and with the chemical information held in the force field are obtained.

## 3.3 TRANSITION STATE SIMULATIONS FOR AMYLOID ELONGATION

The TS of amyloid elongation is the configuration that must be reached by a free monomer attaching to a fibril end, in order to fully attach and adopt a more stable energetic state (Figure 3.3). This dominant free-energy barrier which slows down the reaction bears resemblance to the classical folding process. The experimental $\Phi$ values are calculated with the free-energy ratio:

$$\Phi^{\text{exp}} = \frac{\Delta\Delta G_{\ddagger-U}}{\Delta\Delta G_{A-U}} \tag{3.10}$$
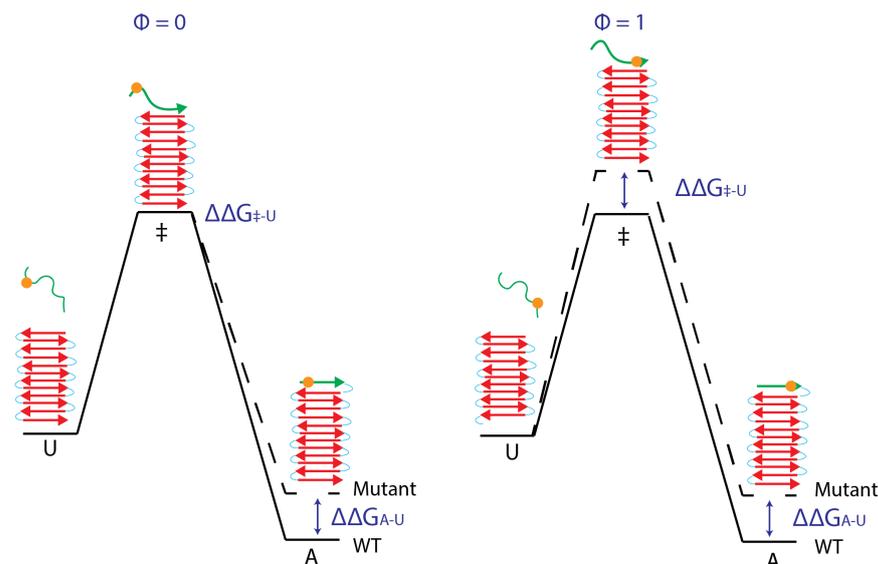
**Figure 3.3:** Principles of Φ-value analysis for amyloid elongation. The transition state lies at the top of the free energy barrier. If the mutated residue is unstructured in the TS, Φ = 0. If the mutated residue is folded in the TS, the TS and amyloid state are destabilised by the same amount and Φ = 0.

where $\Delta\Delta G_{\ddagger-U}$ represents the difference in Gibbs free energy change between the TS and the unfolded state for a mutant and the wild type (WT). $\Delta\Delta G_{A-U}$ represents the difference in Gibbs free energy change between the folded amyloid and the unfolded state for a mutant and the WT.

Again, if $\Phi \approx 0$, the mutation does not destabilise the TS with respect to the unfolded state. This implies that the residue does not make interactions in the TS. If $\Phi \approx 1$, the mutation perturbs the TS and native state by the same amount and the residue is expected to have already formed many of its contacts in the TS.

In Manuscript II, for the first time, experimental Φ-values were calculated for amyloid fibril elongation. The amyloid fibril under investigation is the SH3 domain of PI3-kinase, which is one of the first proteins discovered to form fibrils in the test tube (Guijarro et al., 1998). PI3K-SH3 itself is not associated with a known disease, but the fibrils are cytotoxic to the cell and it is often used as a model system for protein folding and fibril formation. PI3K-SH3 monomers carry a positive net charge at acidic pH conditions. Vettore and Buell, 2019 have previously suggested that in the TS, the incoming monomer has already overcome the electrostatic repulsion to the fibril end and is in a similar state to the final fibril state. This suggests that the rate-limiting step along the elongation pathway is not an intra-molecular structural arrangement of the monomer to become compatible with fibrillation. Rather the rate-limiting step is an inter-molecular event between the incoming monomer and the fibril end.

$$\rho(t) = \sum_i ( \Phi_i^{calc}(t) - \Phi_i^{exp})^2$$
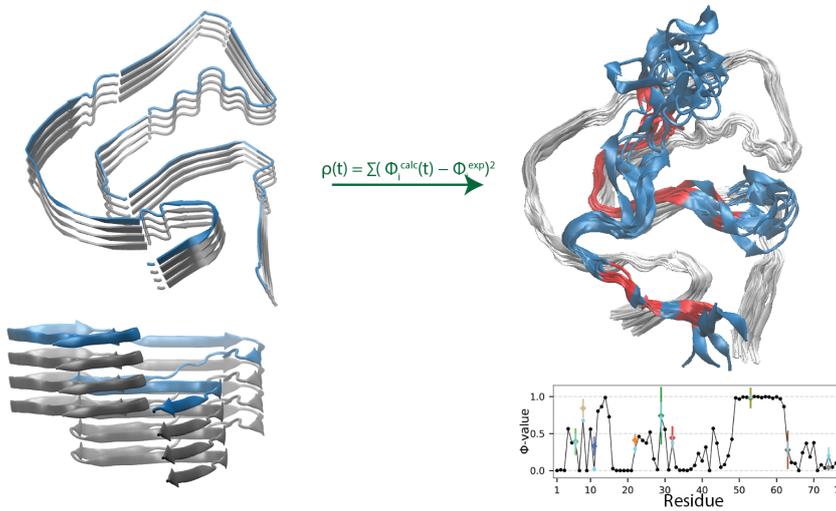
**Figure 3.4:** Left: Starting amyloid structure for the simulations consisting of four staggered subunits. Right: A representative transition state ensemble. Blue indicates regions which are completely dissociated from the fibril end while red indicates regions which are attached. The plot shows the predicted $\Phi$-values for every residue in the monomer, which are calculated from the simulation using Equation 3.8.

In Manuscript II we show that experimental $\Phi$-values can be used to guide simulations of the TS of the PI3K-SH3 fibril, using the same procedure which was described for protein folding. Thereby the most important TS contacts across the whole sequence length can be identified.

The atomic structure of PI3K-SH3 was determined by cryo-EM (Röder et al., 2019). The fibril consists of two intertwined photofilaments with cross $\beta$-sheet structures that are characteristic of amyloids. The monomers are staggered to span four layers of the fibril. Therefore the starting point for the simulations was the structure consisting of four subunits (Figure 3.4). The simulations were carried out using the Amber99sb-*disp* force field, which was designed to provide an accurate description of both disordered and folded proteins (Robustelli, Piana, and Shaw, 2018). The fibril structure with only four stacked monomers cannot be expected to be stable, since the staggered arrangement means each monomer should form stabilising interactions with subunits four layers above and below. Instead of simulating a longer fibril which would increase the system size dramatically, the backbone atoms of the lower three subunits were restrained to their starting positions to maintain the rigid fibril structure. In the set of native contacts, only side-chain atoms were considered, since the experimental $\Phi$-values were obtained mostly from deletion mutations.

The structural characteristics of the TSE are shown in Figure 3.4. With 4 μ*s* of simulation, there is little structural variability in the ensemble, suggesting the conformational space that can satisfy the $\Phi$-value restraints is small, aside from residues 30 to 45 which are mainly disordered and
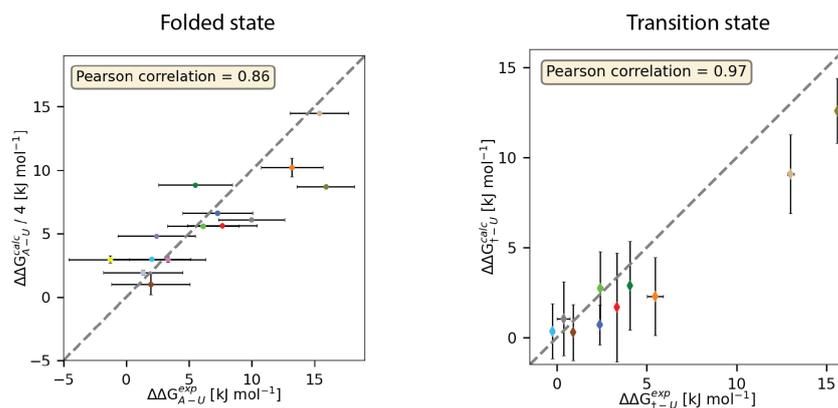
**Figure 3.5:** The good correlation between experimental and FoldX ΔΔGs from ensembles of folded and transition state structures.

detached from the fibril end. The simulation suggests residues at positions 49 to 62, within the hydrophobic core, make all of their amyloid contacts in the TS and form the largest interaction site of the incoming monomer with the fibril-end. Presumably it is critical for these contacts to form as a nucleation site for the rest of the monomer to sequentially fall into place.

Many assumptions are made to perform the simulations, including the assumption that the starting model and force field is reliable, and that modelling Φ-values in terms of native contacts, without taking into account non-native contacts, is a good approximation. Lindorff-Larsen et al., 2003 outlined a procedure to validate the TSE of protein folding. The same procedure was used to validate the TSE of PI3K-SH3, where free energies are estimated directly from the simulated structures using the program FoldX (Schymkowitz et al., 2005). The specific mutations used in the experiments are modeled in a large number of native and TSE conformations and the calculated ΔΔGs are averaged, since the experimental data are also averages over a large ensemble. The high agreement between the experimental and calculated ΔΔGs (Figure 3.5) confirms that the native contact approximation holds and that the guided ensemble provides a good model of the TS of PI3K-SH3 elongation. It also indicates that the energy functions used in FoldX can be used to investigate the energy landscape of amyloid fibrils as well as well-behaved folded proteins.

The outcomes of these simulations provide valuable initial guesses for attempts to inhibit fibril formation. Since more and more high-resolution structures of amyloid fibrils are becoming available (Gremer et al., 2017; Iadanza et al., 2018; Radamaker et al., 2019), it is feasible to explore the transition states of disease-related fibrils using computational instructions developed in Manuscript II.

# 4

## SMALL-ANGLE SCATTERING

Small-angle x-ray scattering (SAXS) and small-angle neutron scattering (SANS) are powerful, complementary experimental techniques that have a broad range of applications in various fields such as biology, polymers, food, metals, powders, and pharmaceuticals. The techniques enable the examination of structures at length scales ranging from approximately 1 to 100 nm. Although SAS data is often considered "low resolution" because it does not provide atomic-level details, this label shouldn't be viewed as discouraging or unattractive; the field of SAS is evolving (Brosey and Tainer, 2019). Advances in neutron and synchrotron technology is moving the field away from simple shape determination towards the characterisation of complicated systems, but the community must continue to push computational methods in order to fully utilise the information contained in the data. In the field of structural biology, SAS opens the doors to study biological samples including multi-component protein complexes, protein-lipid complexes, and flexible proteins. SAS can help to determine the overall arrangement of subunits (Yan et al., 2020), aggregation states (Sauter et al., 2016), exact protein conformation (Tang, Tainer, and Hura, 2017), and distributions of conformational ensembles (Thomasen et al., 2023). Since SAS is usually measured in solution, macromolecular changes can be triggered by temperature (Molodenskiy et al., 2017) or pressure (Marion et al., 2015) jumps, or by microfluidic mixing (Pham et al., 2017).

SAS is even sensitive to detecting *differences* in macromolecular conformations even below 1 nm. For example, Tang, Tainer, and Hura, 2017 demonstrated that it is possible for SAXS to detect a 0.5 nm decrease in maximum dimension of the protein NBS1 when it moves from an unbound to contracted bound state, as long as the data is collected from a very homogeneous sample.

Sebastiani et al., 2021 used SANS to investigate how RNA-carrying-lipid nanoparitcles are internally structured in the presence and absence of apolipoprotein. They were able to calculate the exact composition of lipids, PEG and cholesterol in the core and shell of the particle from their models.

Time-resolved SAXS offers temporal separation of changes in conformation or oligomerisation down to 100 pico seconds (Cho et al., 2010). Cho et al., 2016 investigated the real-time photocylce of a signalling protein from 100 ps to 1 s and observed globular structural changes. Martin et al.,

2021 were able to map the phase separation of A1-LCD on the micro to milli second timescales, observing the transformation from from dimers, to a distribution of small clusters, to very large clusters. Their results suggest the initial aggregation into small clusters is unfavourable and delays phase-separation. These important dynamics would be invisible to static approaches like crystallography and EM.

## 4.1 THEORY OF SAS

SAS theory is laregly unchanged since its inauguration in the mid-1900s (Guinier, 1939; Guinier et al., 1955; Porod, 1948). The theory presented here is mainly taken from works by Feigin, Svergun, et al., 1987 and Svergun and Koch, 2003, and a recent review by Gommes, Jaksch, and Frielinghaus, 2021.

### 4.1.1  *X-rays versus neutrons*

X-rays are electromagnetic radiation. They primarily interact with electric charges and therefore the electrons of atoms in the sample. Neutrons, on the other hand, are neutral nuclear particles which interact with atomic nuclei. Despite their difference in physical nature, x-ray and neutron scattering can be described using the same mathematical formalism due to their dual-characterisation as both particles and waves. X-rays and neutrons share comparable wavelengths which make them both suitable for structural studies on the nanoscale.

The wavelength for electromagnetic radiation can be calculated with

$$\lambda_x = \frac{hc}{E_x} \tag{4.1}$$

where $h$ is Planck's constant, $c$ is the speed of light and $E_x$ is the energy of the photon. SAXS experiments typically have energies around 7 to 20 keV corresponding to wavelengths around 0.6 to 1.8 Å. For example, the bioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble. France has a nominal wavelength of 0.99 Å. For home sources, typically x-ray tubes are used to produce K$\alpha$ photons which have a set wavelength of 0.76 Å.

For neutrons, the wavelength is given by the de Broglie relationship

$$\lambda_n = \frac{h}{\sqrt{2E_n m_n}} \tag{4.2}$$

where $E_n$ is the energy of the neutron and $m_n$ is the mass of a neutron. Typically the neutrons in SANS experiments have energies in the range 0.8 to 20 keV and span a larger range of wavelengths, 0.2 to 10 Å. At the SANS beamline D22 at the Institut Laue-Langevin (ILL) in Grenoble, France, the wavelengths are in the range 0.5 to 4 Å with a band of 10%.

| Atom | H | D | C | N | O | P |
|---|---|---|---|---|---|---|
| Z | 1 | 1 | 6 | 7 | 8 | 15 |
| $b_x$ [$10^{-12}$ cm] | 0.282 | 0.282 | 1.69 | 1.97 | 2.16 | 4.23 |
| $b_n$ [$10^{-12}$ cm] | -0.374 | 0.667 | 0.665 | 0.940 | 0.580 | 0.510 |

**Table 4.1:** X-ray and neutron atomic scattering lengths for some biologically relevant elements (Sears, 1992).

Scattering lengths help quantify the ability of elements or isotopes to scatter radiation. X-rays have coherent atomic scattering lengths proportional to the electron number, $Z$, of the particular atom, $b_x = Zr_0$ where $r_0 = 2.82 \cdot 10^{-13}$ cm is the Thomson radius. This means heavier atoms scatter x-rays more strongly than lighter atoms. Neutrons do not show any apparent trend between atomic number and scattering length, and therefore neutron scattering lengths had to be obtained experimentally (Sears, 1992). The variation between atomic scattering lengths for x-rays and neutrons is show in Figure 4.1.
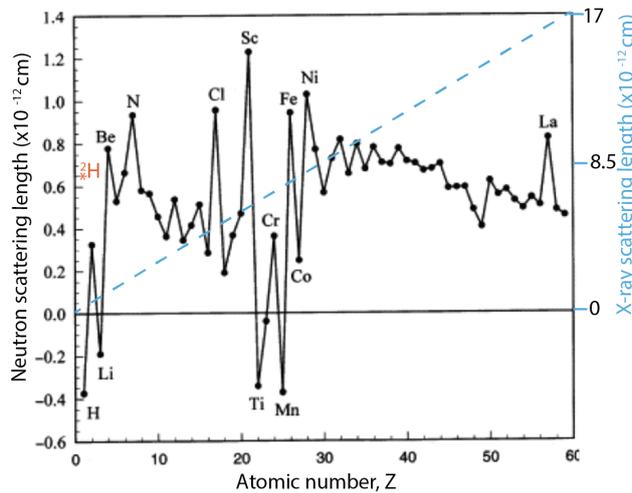


**Figure 4.1:** Neutron and x-ray atomic scattering lengths across the periodic table. Black points: neutrons, blue dashed line: x-rays. Adapted from Long, 2003 with permission from the publisher, *Elsevier Books*

Neutron scattering lengths also vary between isotopes. For biological samples the difference in neutron scattering length between hydrogen and deuterium opens up many exciting opportunities (Table 4.1). Selective deuteration, where H is replaced with D to manipulate the scattering properties of molecules without altering their chemistry, can be used to probe specific components of a system. Furthermore, H has a considerable incoherent scattering cross-section while D does not, and therefore D-based solvent is much more favourable to minimise background noise in scattering experiments.
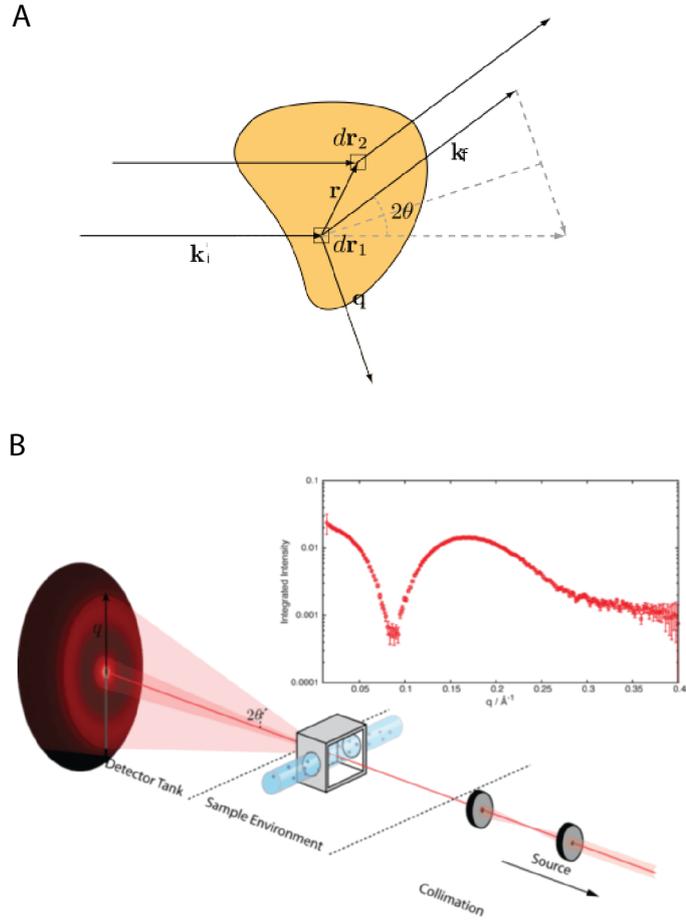
**Figure 4.2:** (A) Schematic diagram of scattering from two volume elements, $dr_1$ and $dr_2$. (B) The basic instrumental set-up for a SAXS or SANS experiment. The source beam is collimated and shone onto the sample. The intensity of the scattered beam is recorded as a function of the scattering angle, $\theta$. The isotropic scattering pattern on the 2D detector can be azimuthally averaged to obtain the 1D profile. Both figures are attributed to Skar-Gislinge, 2014 with permission from the author.

### 4.1.2  *Experimental set-up*

Figure 4.2A shows an object illuminated by two monochromatic plane waves. The incoming wave vector, $\vec{k}_i$, has a corresponding wavelength $\lambda$ and magnitude $|\vec{k}| = \frac{2\pi}{\lambda}$, and interacts with a scatterer (which could be a single atom or a small volume containing many atoms). In small-angle scattering, usually only elastically scattered waves are considered, where the modulus of the scattered wave $|\vec{k}_f| = |\vec{k}_i|$. Therefore the scattering vector $\vec{q} = \vec{k}_f - \vec{k}_i$ can be calculated as

$$|\vec{q}| = 2|\vec{k}|\sin\theta = \frac{4\pi\sin\theta}{\lambda} \tag{4.3}$$

where $2\theta$ is the scattering angle.

In a typical SAS experiment (Figure 4.2B) a beam of either x-rays or neutrons irradiates the sample. The beam passes through a monochromater and a collimation system to remove diverging radiation. Most of the beam passes straight through the sample and hits the beamstop on the detector. A small percentage of the beam interacts with atoms in the sample and are scattered in every direction. The detector measures the intensity of the scattered radiation within a certain range of angles. In order to resolve small angles ($< 10°$), the detector must be placed far away from the sample. The resulting scattering pattern is the Fourier transform of the real-space distribution of scatterers into reciprocal space. SAS provides information about the real-space distances as $d = \frac{2\pi}{q}$.

Scattering from assemblies of scatterers will produce scattering patterns of constructive and destructive interference. The scattering pattern is dependent on the positions (phase) of scatterers and their scattering lengths. Each scattered wave is described as a complex number $b \exp(i\vec{q} \cdot \vec{r})$ where $b$ is the scattering length of the scatterer and $\exp(i\vec{q} \cdot \vec{r})$ is the phase. The total scattering intensity is the linear sum of the contribution from all scatterers in the sample:

$$A(\vec{q}) = b_1 \exp(i\vec{q} \cdot \vec{r}_1) + b_2 \exp(i\vec{q} \cdot \vec{r}_2) + ... \tag{4.4}$$

However, detectors cannot measure the wave amplitude but only the square of the modulus so $I(q) = |A(q)|^2$. SAS data is usually presented as the scattered intensity $I$ as a function of $q$, using Equation 4.3. One advantage of converting $I(\theta)$ to $I(q)$ is that it is independent of wavelength and so experimental data from different sources can be directly compared.

In SAS experiments on biomolecules in solution, the buffer produces the dominating scattering contribution. Therefore, two measurements must be taken: one from the sample and one from the buffer alone. In order to isolate the scattering contribution from the particle of interest, the scattering profile from the buffer is subtracted from the scattering profile from the sample:

$$I(q)_{\text{particle}} = I(q)_{\text{particle+buffer}} - I(q)_{\text{buffer}} \tag{4.5}$$

This also means that the scattering from the buffer does not usually need to be accounted for in models for SAS data.

### 4.1.3  *Central formulae for SAS from biomolecules*

Since SAS does not provide atomic resolution, it is easier to consider the bulk volume of the molecule, or smaller volume elements, rather than individual atoms. The scattering length density, $\rho$, of a scattering volume $V$ containing $N$ atoms is the total scattering length per unit volume

$$\rho = \sum_{j=1}^{N} \frac{b_j}{V_j} \tag{4.6}$$

By taking the solvent subtraction into account (Equation 4.5), the contrast of a volume element to the buffer can be found, also known as the excess scattering length density, $\Delta\rho = \rho - \rho_{\text{buffer}}$. The scattering amplitude from a particle is the Fourier transform of $\Delta\rho$:

$$A(\vec{q}) = \int_V \Delta\rho(\vec{r}) \exp(i\vec{q} \cdot \vec{r}) d\vec{r} \tag{4.7}$$

where the integration is performed over the volume of the particle and $\Delta\rho(\vec{r})$ is zero everywhere outside of the particle.

In dilute solutions the molecules are free to rotate, and accordingly the intensity is isotropic and represents the scattering averaged over space and time. This can be calculated by averaging over all orientations, $I(q) = \langle A(\vec{q}) A^*(\vec{q}) \rangle_\Omega = \langle I(\vec{q}) \rangle_\Omega$ where * denotes the complex conjugate.

With a few steps and exploiting the fact

$$\langle \exp(i\vec{q} \cdot \vec{r}) \rangle_\Omega = \frac{\sin(qr)}{qr} \tag{4.8}$$

The Debye equation can be derived (Debye, 1915)

$$I(q) = 4\pi \int p(r) \frac{\sin(qr)}{qr} dr \tag{4.9}$$

where $p(r) = \gamma(r)r^2$ and $\gamma(r) = \langle \int_V \Delta\rho(\vec{r_1}) \Delta\rho(\vec{r_1} - \vec{r}) \, d\vec{r_1} \rangle_\Omega$ is known as the auto-correlation function. $p(r)$ is known as the pair-distance distribution and can be interpreted as the distribution of distances between each pair of scatterers in a particle. This is revisited in Section 4.2.1.

The total scattering intensity from an ensemble of molecules in solution is simply the sum of their individual scattering intensities, as long as they are randomly distributed and oriented, and their movements and positions are uncorrelated.

If the sample does not contain molecules of identical shape and size, the recorded SAS pattern will be the weighted averaged scattering intensity from every molecule present. When polydispersity or polymorphism is present, interpreting the data becomes much more ambiguous.
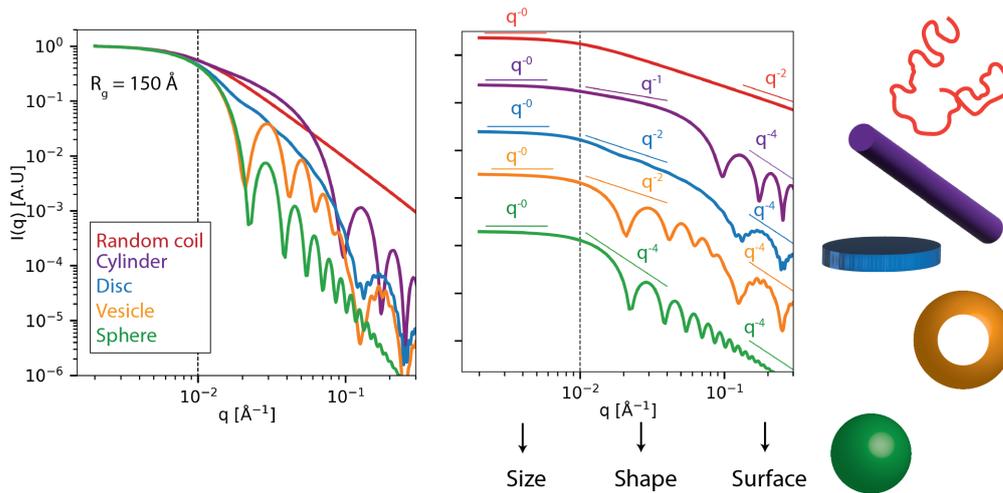
**Figure 4.3:** Different objects produce distinctly different SAS profiles. Real-space structural information can be extracted from different regions of a SAS profile. The form factors of five different shapes with the same radius of gyration are plotted. The Guinier region at very low-$q$ is indicated with the dashed line at $q \approx 1.3/R_g$. The form factors are identical in this region. The initial slope characterises the shape of the object. Oscillating features in the form factor are the result of periodic order in the object. The Porod exponent at high-$q$ informs on the surface of the object.

## 4.2 THE INVERSE SCATTERING PROBLEM

SAS data is often difficult to interpret since the scattering information is recorded in reciprocal space. Due to only being able to measure $I(q)$ and not $A(\vec{q})$, orientational averaging and loss of 3D spatial information, it is not possible to transform a SAS profile back into its real-space coordinates. In reciprocal space, the scattering vector, $q$, is inversely proportional to the characteristic length scale in real-space. Small $q$-values correspond to long interatomic distances, for example the overall size of the particle, while larger angles correspond with smaller distances, like local structural features. Information about the shape of the particle lies in the mid-$q$ region (Figure 4.3).

### 4.2.1 *Direct data analysis*

Still, there are some methods that can be employed to quickly obtain real-space parameters directly from solution SAS data, without relying on prior assumptions. These methods should always be performed as a preliminary step before beginning more complex analysis. These methods can be used as a quality check of the SAS data and provide insights for good starting points for modelling, which always require prior knowledge.

**Radius of gyration**

The initial decay in $I(q)$ in the very low-$q$ region known as the Guinier region obeys a universal law that is only dependent on the average size of the particle (Guinier et al., 1955)

$$P(q) \approx \exp\left(\frac{(-qR_g)^2}{3}\right) \tag{4.10}$$

where $R_g$ is the radius of gyration. By plotting the experimental scattering data as $\ln(I(q))$ against $q^2$, a linear relationship can be fit which will have a slope of $-R_g^2/3$. Guinier's law is usually only valid in the range $\sim q < 1.3/R_g$, which means it is crucial to measure SAS data at very low-$q$ values in order to to record a valid Guinier region and accurately determine the size of larger particles with SAS.

**Power laws**

By investigating the power law decay of the scattering intensity at small angles, particles can be roughly characterised in terms of their general shape. Power laws are in the $I(q) \approx q^{-x}$, which can be easily identified as straight lines when the SAS data is plotted on log-log scale. Initial slopes of -1, -2 or -4 indicate cylindrical, lamaller or spherical structures, respectively (Schnablegger and Singh, 2013).

The high-$q$ region of the SAS profile is known as the Porod region. The Porod exponent, -4, indicates a smooth surface with a sharp contrast to the surrounding solvent (Ciccariello, Goodisman, and Brumberger, 1988). An exponent of -3 indicates rough surfaces or surface fractals. Gaussian random coils have a Porod exponent of -2.

**Pair-distance distributions**

Pair-distance distributions, $p(r)$, are arguably the most powerful representations of SAS data as they contain intuitive (1D) information in real-space. $p(r)$'s are essentially weighted histograms of all possible distances of pairs in the particle.

According to Equation 4.9, in principle, the $p(r)$ could be obtained by the inverse transform of the scattering profile:

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty I(q)qr\sin(qr)dq \tag{4.11}$$

In practice, the integration is not possible since the $q$ regime is limited to a finite number of points. $q \to 0$ cannot be measured due to the presence of the beam stop in and the upper limit is due to the fading of signal into the noise level.

Glatter found an alternative solution to the problem and introduced the Indirect Fourier Transform (IFT) (Glatter, 1977). The IFT method involves an iterative fitting procedure, where a $p(r)$ is approximated
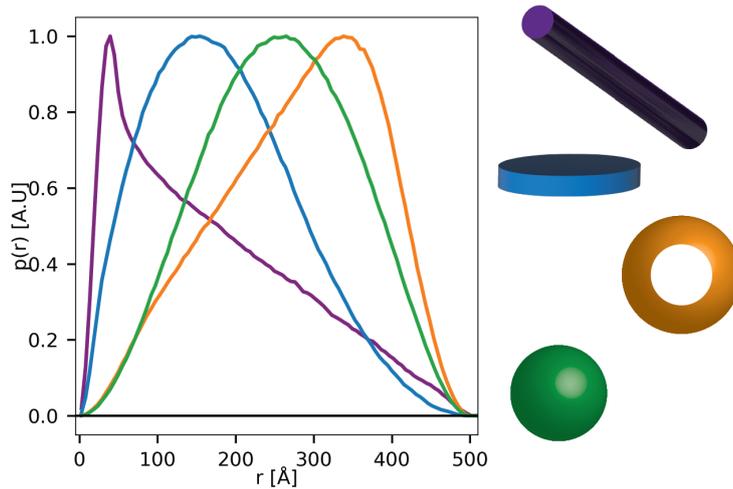
**Figure 4.4:** Different shapes produce distinctly different pair-distance distributions. The distributions show that all structures have a maximum dimension of 500 Å. The distributions were generated using the software *Shape2SAS* (Larsen et al., 2023).

as a linear combination of many polynomial functions. Their coefficients are adjusted by a least-squares method so that when the $p(r)$ is transformed into $I(q)$, the optimal agreement with the experimental scattering intensity is found.

The $p(r)$ contains valuable information about the average shape and size of the particle. The maximum dimension of the particle, $D_{max}$, is found at the value of $r$ where the $p(r) = 0$. Furthermore, various structures are distinctly recognisable from the shape of the $p(r)$ profiles, e.g. spheres, hollow spheres, discs and rods (Figure 4.4).

The $p(r)$ can have negative values depending on $\Delta\rho$ of different components of the particle. This is often seen for lipids where the tailgroups have a lower scattering density than solvent in SAXS.

**Complex samples**

Of course, as the complexity or heterogeneity of samples increases, interpreting the SAS profile is less straightforward than the simple cases illustrated in Figures 4.3 and 4.4, where the data provides a clear representation of their shape and dimensions. When dealing with samples containing multiple components, the scattering pattern becomes a composite of distinct scattering behaviours which must be disentangled to extract meaningful structural information. To address this complexity, analytical models and other computational techniques must be employed.
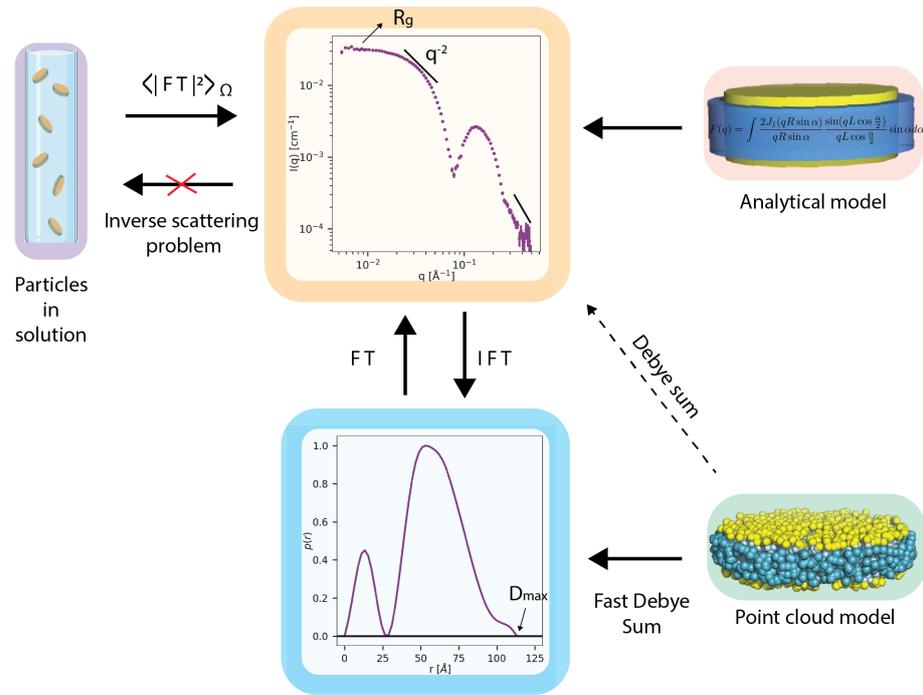
**Figure 4.5:** The inverse scattering problem. The SAS data cannot directly provide the original structure of the nanodiscs. One strategy to extract structural information is to perform model-independent analysis, such as investigating the $R_g$, the initial slope of the data and the $p(r)$ distribution. The other strategy is to construct a geometrical model and compare it to the data. Analytical models make use of mathematical form factors. Point-cloud models are built in real-space. The corresponding scattering profile can be calculating using the Debye sum or the Fast Debye sum, where first the $p(r)$ distribution is generated and then Fourier transformed. The SAXS data is published in Paper I (Barclay et al., 2023) and represents MSP1D1 nanodiscs with DMPC lipids.

### 4.2.2 *Model-based analysis*

Prior knowledge of the system can be used to build geometrical models and calculate theoretical scattering profiles. The parameters of the model are varied in a least-squares fashion to obtain the best agreement with experimental SAS data. The reduced $\chi^2$ is usually used to evaluate the goodness of fit.

$$\chi_R^2 = \frac{1}{N - M} \sum_i^N \left[ \frac{I_i(q) - I_{\text{exp},i}(q)}{\sigma_i(q)} \right]^2 \tag{4.12}$$

Where $N$ is the number of data points, $M$ is the number of fitted parameters, $I(q)$ is the calculated scattering intensity, $I_{\text{exp}}(q)$ is the experimental scattering intensity and $\sigma(q)$ is the uncertainty associated with $I_{\text{exp}}(q)$. $N - M$ is the number of degrees of freedom.

In general $\chi_R^2 \approx 1$ indicates an ideal fit where the optimal set of parameters for the model has been found. However, $\chi_R^2$ is extremely sensitive to experimental noise levels and the degrees of freedom in the model. Often model fits are assessed by eye and intuition. Some work-arounds exist to systematically asses $\chi_R^2$s, such as robust statistical methods (Rambo and Tainer, 2013) and rescaling of error bars (Larsen and Pedersen, 2021), but there is not a general consensus in the community.

**Analytical modelling**

SAS profiles are characteristic of the shape and size of the particles they represent. Analytical modelling involves Form factors, $P(q)$, which are mathematical expressions that describe the geometry of particles in SAS. The form factor for a variety of shapes, e.g. spheres, ellipsoids, cylinders, shells etc. can be determined analytically and are readily available in the literature (Pedersen, 1997). Fitting parameters could be, for example, the radius of a sphere or the length of a cylinder.

For dilute systems without particle-particle effects, the theoretical scattering intensity from monodisperse biomolecular particles can be expressed as:

$$I(q) = n\Delta\rho^2 V^2 P(q) \tag{4.13}$$

where $n$ is the particle number density of the sample. Usually the form factor $P(q) = |F(q)|^2$ where $F(q)$ is the form factor amplitude. The forward scattering intensity, $I(0) = n\Delta\rho^2 V^2$ since $P(0) = 1$. This is a helpful calculation to compare with theoretical values on absolute scale and can be used to calculate the molecular weight.

A great strength of this kind of modelling is its adaptability; by combining simpler shapes a very wide variety of complex models can be constructed. For example, systems with several layers of different scattering properties ($\Delta\rho$) can be represented with multishell models as
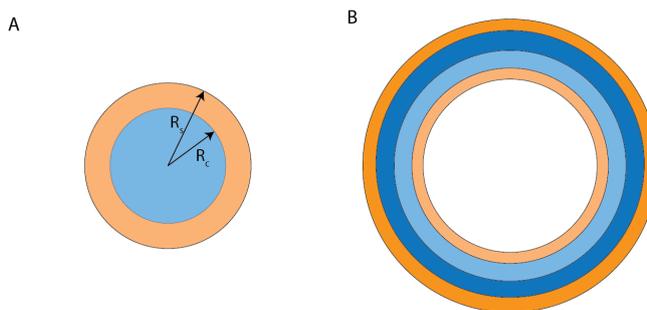
**Figure 4.6:** Examples of how spherical form factor amplitudes can be combined to build multishell models. Each component can be assigned its respective $\Delta\rho$. (A) Core-shell micelle model (B) Four-shell vesicle model.
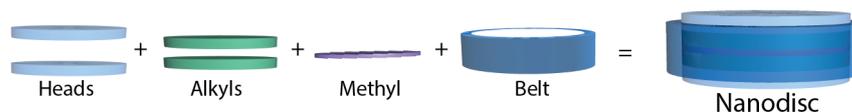


**Figure 4.7:** The nanodisc model. The lipid bilayer is represented as a collection of discs and the MSP belt is represented as a hollow disc. Figure from Paper III (Barclay et al., 2023).

depicted in Figure 4.6. Micelles, like copolymer micelles (Manet et al., 2011) and surfactant micelles (Jensen et al., 2013), can be modelled as a spherical core and a surrounding shell, which takes into account the size and $\Delta\rho$ of both components. Lipid vesicles can be modelled with a spherical core representing the solvent and multiple concentric shells representing lipid headgroups and tailgroups. Additional shells can be included in the model for vesicles with more lamellae (Sakuragi et al., 2011).

Another example is the nanodisc model, illustrated in Figure 4.7 (Skar-Gislinge and Arleth, 2011; Skar-Gislinge et al., 2010). The nanodisc is split into four components with different $\Delta\rho$s: lipid headgroups, alkyl chains, methyls, and the MSP belt. Multiple cylindrical form factor amplitudes are included to represent a stack of discs, with each disc representing a different component of the nanodisc.

$$I(q) = n \cdot \langle |A_{\mathrm{head}}(\vec{q}) + A_{\mathrm{alkyl}}(\vec{q}) + A_{\mathrm{meth}}(\vec{q}) + A_{\mathrm{belt}}(\vec{q})|^2 \rangle_\Omega \qquad (4.14)$$

Where the scattering amplitude $A(\vec{q}) = \Delta\rho V F(\vec{q})$ and $\Delta\rho$ and $V$ are dependent on the particular component.

If a sample contains more than one species, the resulting scattering intensity can be modeled as a linear combination of the scattering intensity from each species. Furthermore, many real systems contain some size distribution of particles, even if their shapes are similar. This

polydispersity can be incorporated into analytical models by integrating over the form factor:

$$I(q) = \int_0^\infty \Delta\rho^2 N(r) V(r)^2 P(q,r) dr \qquad (4.15)$$

where $N(r)$ is the number density of particles with characteristic property $r$.

**Scattering from coordinates**

Theoretical SAS profiles can also be calculated directly from real-space coordinates. This is particularly useful for proteins if their atomic structures are available, either from experiments or through computational methods. If a protein structure has atoms (or beads representing a collection of atoms) positioned at $r_1, r_2, ... r_N$ with atomic form factor amplitudes $f_1, f_2, ... f_N$, the total scattering amplitude can be calculated as:

$$f_{\text{total}}(q) = \sum_{j=1}^{N} f_j(q) \exp(-iq \cdot r_j) \qquad (4.16)$$

The intensity is the absolute square of the intensity, averaged over all orientations, $I(q) = f(q)f^*(q)$.

When $j = k$, the phase factor becomes one and $I(q)$ is the intensity from the atom alone. When $j \neq k$, $I(q)$ represents the interference between two atoms.

By making use of Equation 4.8, we again arrive at the Debye formula, which is split into the 'auto-correlation' and 'cross-correlation' terms:

$$I(q) = \sum_{j=1}^{N} I_j(q) + \sum_{j}^{N} \sum_{k \neq j}^{N} f_j(q) f_k(q) \frac{\sin(q(r_j - r_k))}{q(r_j - r_k)} \qquad (4.17)$$

The double sum of all atoms makes it a computationally heavy calculation with $N^2$ calculations for each value of $q$. This is a substantial obstacle, particularly when calculating SAS profiles from large ensembles of conformations. It is not unusual to want to calculate the SAS profiles from thousands or tens of thousands of structures. Due to this high computational demand, various methods have been formulated to reduce the number of calculations.

One method is to simplify the orientational averaging by using the multipole expansion of $\exp(i\vec{q} \cdot \vec{r})$ and expanding the scattering intensity into a series of spherical harmonics (Svergun, Barberato, and Koch, 1995). This is a common approach and is used in multiple programs for calculating SAXS intensity profiles from x, y, z coordinates: e.g. CRYSOL (Svergun, Barberato, and Koch, 1995) and PEPSI-SAXS (Grudinin, Garkavenko,

and Kazennov, 2017). The computational complexity is reduced from $N^2$ to $N$ for each $q$-value.

**Fast Debye Sums**

Another l way of transforming real-space models into scattering intensities was first introduced by Hansen, 1990. Point-clouds are constructed from dense collection of points that can represent particles of any shape and composition. The $p(r)$ is directly calculated from the point cloud and then Fourier transformed to produce the scattering profile (Figure 4.5). The $p(r)$ can be calculated as:

$$p(r) = \sum_{j=1}^{N} \sum_{k=1}^{j} \delta(r - r_{jk}) \Delta b_j \Delta b_k \tag{4.18}$$

where $N$ is the number of points, $\Delta b$ is the excess scattering length, $r_j k$ is the distance $j$th and $k$th point and $\delta(...)$ is the Dirac function. Contributions from $k > j$ can be discarded and the number of pair-wise terms to be calculated is $\frac{N(N+1)}{2}$. Therefore both the speed and the accuracy of the model depends on $N$. If too few points are used, discrepancies will appear in the high-$q$ region of the scattering profile. Since $p(r)$s are independent $q$, the double sum only needs to be calculated once.

Pedersen et al., 2012 combined this method with least-squares modelling to optimise geometric parameters against SAXS data. Scattering profiles can be rapidly generated from complex shapes where the intensity cannot be found analytically. Starting with a large number of randomly and uniformly distributed points, each component of the structure can be modeled by selecting points according to the geometric parameters. Various components can be assigned different scattering lengths (e.g. lipid headgroups and tailgroups). Pedersen et al., 2012 show the calculation can be sped up by dividing the point-cloud into subsets. The $p(r)$ is calculated for each subset and then simply summed to give $p(r)_{\text{total}}$. E.g. calculating 10 $p(r)$'s with 4000 points each is ten times faster than calculating the $p(r)$ from 40,000 points.

While the Fast Debye Sum (FDS) approach is not the most popular in SAS, parallelisation via GPUs makes it computationally attractive. Pedersen et al., 2022 recently presented FDS as an effective model for calculating the scattering intensity from ensembles of flexible membrane proteins-nanodisc complexes, where the nanodisc parameters can be refined from the SAS data. This will be revisited in detail in Section 5.3.

## 4.3 SAS AT LARGE-SCALE FACILITIES

Progress in x-ray and neutron science has been advancing at a rapid pace since the 1960s (Jaeschke et al., 2020) (Figure 4.8). Synchrotron radiation is very important for SAXS from biological samples due to the much

better signal-to-noise as well as larger $q$-range compared with in-house instruments. Over the last few decades many substantial advancements have been made, particularly in source improvements, instrumentation, detectors and data collection. These advancements facilitate the trend towards more complex and heterogeneous samples, which require high-resolution measurements both spatially and temporally. The higher brilliance and more coherent x-rays at SAXS beamlines means experiments are becoming shorter and smaller sample volumes are required. Furthermore time-resolved experiments down to sub-microsecond resolutions provide exciting opportunities to study dynamic processes of biomolecules in solution, such as self-assembly in early-stage amyloid formation, (Ortore et al., 2011), protein unfolding(Hsu et al., 2021) and protein-protein binding (Rimmerman et al., 2017).

SANS is also benefiting from emerging more powerful neutron sources. In particular, the upcoming European Spallation Source (ESS) in Lund will become the world's brightest neutron source providing a significantly higher neutron flux compared to other facilities. All of the instruments are designed to offer world-leading performances. The ESS will have two SANS instruments which will together cover a wide $q$-range from $1x10^{-4}$ Å$^{-1}$ to 2 Å$^{-1}$, spanning real-space sizes from a few angstrom to a micrometre. The high flux, small beam and wide $q$-range with a single detector setting will make it possible to perform time-resolved structural studies of biological systems with SANS, opening up many exciting prospects (Andersen et al., 2020). Additionally a very-SANS instrumental set-up will be able to reach down to $q = 10^{-5}$ Å$^{-1}$ and a semi-transparent beam stop will allow measurement essentially down to $q = 0$ Å$^{-1}$ limited by pixel size (Jaksch et al., 2021) meaning the overall sizes and structures or biomolecules can be effectively resolved. The ESS will also offer simultaneous collection of SANS and dynamic light scattering (DLS) in order to monitor stability and measure the microscopic dynamics (through DLS) as well structural properties (through SANS) on the exact same sample (Schmid et al., 2021).

Another exciting recent implementation is at the Institut Laue–Langevin (ILL) where a portable SAXS instrument is installed directly at the SANS instrument. Therefore SAXS and SANS data from the same sample volume can be fitted simultaneously using a common structural model, allowing unambiguous interpretation of the scattering data and complimentary information from different contrasts (Metwalli et al., 2021).

Finally, one advancement in instrumentation that has been very popular is the introduction of size-exclusion chromatography (SEC) coupled with SAXS (SEC-SAXS) and SANS (SEC-SANS), where SEC is performed inline with scattering measurements to obtain data from a homogeneous sample. SEC-SAXS was first installed by Mathew, Mirza, and Menhart, 2004 at the BIOCAT beamline at Argonne National Laboratory, and
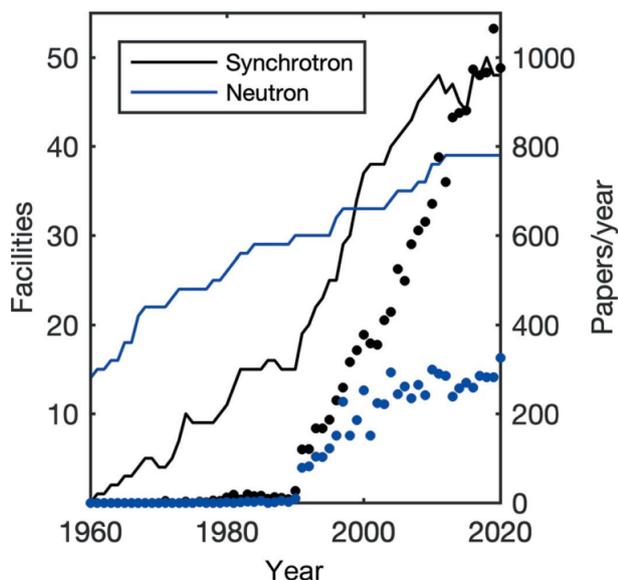
**Figure 4.8:** Solid lines: the number of synchrotrons and neutrons facilities worldwide, points: the number of articles published yearly on SANS and SAXS. Figure from Gommes, Jaksch, and Frielinghaus, 2021 with permission from IUCr Journals.

is now a standard sample environment at most bioSAXS beamlines (Blanchet et al., 2015; David and Pérez, 2009; Pernot et al., 2013; Watanabe and Inoko, 2009). SEC-SANS comes with its own limitations, due to the typically low flux and therefore long exposure times at neutron facilities. The first instrument was installed at the ILL in Grenoble (Jordan et al., 2016; Martel et al., 2023). The high brilliance of the upcoming ESS is very well suited for SEC-SANS experiments.

The principles of SEC-SAS are outlined in Section 4.4 and a novel method for analysing data from well-resolved SEC-SAXS experiments is presented in Paper I.

## 4.4   SEC-SAS

To obtain high-quality data from SAS experiments, highly homogeneous, non-aggregated samples are essential. In SAS where the intensity is roughly proportional to the square of the molecular volume, even a small amount of impurity can disrupt the information content in the scattering curve. Furthermore, biomolecules are prone to aggregation, which can occur shortly after purification and is evident as an upturn at very low-$q$ data points. To address this issue and optimize data quality, many bioSAS beamlines introduced SEC-SAS to reduce the number of aggregated samples reaching the beam.
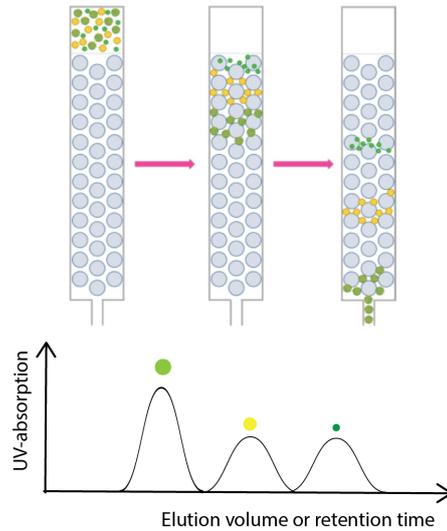
**Figure 4.9:** The main principle of size exclusion chromatography, where particles are separated according to their size. Smaller molecules become trapped in the porous beads while larger molecules flow more rapidly. Therefore, larger molecules will elute earlier than smaller ones. The plot shows the elution UV trace. Figure adapted from Majeed, Sekhosana, and Tuhl, 2020 with permission from the publisher *Elsevier*.

SEC, also known as gel-filtration, separates particles based on their Stoke's radius (the radius a hard sphere with the same hydrodynamic properties would have)(La Verde, Dominici, and Astegno, 2017). The main principle of SEC is illustrated in Figure 4.9. The sample is passed through a column filled with porous beads, where small particles become trapped in the matrix while larger particles flow more freely. As a result, the largest molecules, typically aggregates, elute from the column first, followed by smaller and smaller molecules (Nagy and Vékey, 2008). This selective separation allows for the removal of aggregates and the isolation of the target protein or molecule of interest.

UV-absorption is measured as molecules elute from the SEC column. In proteins, aromatic amino acids e.g. tryptophan, tyrosine, and phenylalanine, absorb UV light with a wavelength of 280 nm. SEC elution profiles are plotted as absorption versus elution volume, with each peak representing particles of a certain size distribution. UV-absorption ($A$) is directly related to the protein concentration via the Beer-Lambert law

$$A = \epsilon l c \tag{4.19}$$

where $l$ is the path length through the sample, $c$ is the molar concentration and $\epsilon$ is the molar extinction coefficient which indicates a molecule's
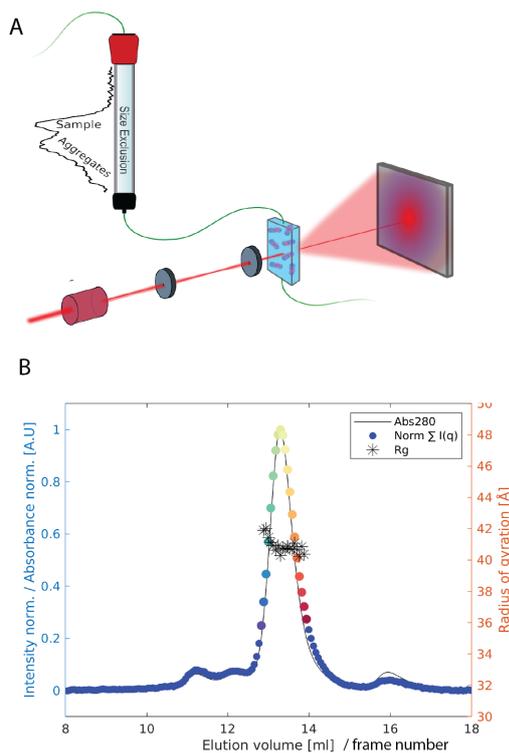
**Figure 4.10:** (A)Diagram of a typical SEC-SAS set-up. Figure from Johansen et al., 2018 with permission from IUCr Journals. (B) Example scattergram. The black line represents the UV-absorption. Each point represents a SAS curve. The intensity of the point is determined by calculating the total intensity of that frame's scattering curve. The colourful points indicate SAS frames corresponding to the target peak of the elution profile. The $R_g$ is calculated for these SAS frames and plotted as stars.

ability to absorb light and can be estimated based on the number of aromatic residues in the protein sequence.

In SEC-SAS setups, a SEC column is positioned so that as the sample elutes from the column, it passes directly through tubing and into the capillary where it is illuminated by the x-ray or neutron beam, as shown in Figure 4.10(A). Scattering patterns are continuously recorded. Scattergrams are plots of the SEC elution profile along with the total scattering intensity for each SAS frame. Scattergrams, like the one shown in Figure 4.10, provide a clear overview of the data set can be used to select the SAS profiles that correspond to the pure target molecule. The alignment between peaks in the SEC elution profile and peaks in the total scattering intensity per SAS frame is generally good since both x-ray/neutron scattering and UV-absorbance depend on the sample concentration. However, the relative height of peaks may differ.

On SEC-SAXS instruments, UV-absorption is measured in the thin tubing between the column and the capillary. However, due to dispersion and capillary effects, the estimated concentration may not accurately

represent the sample being illuminated by x-rays. To address this, the SEC-SANS instrument at ILL measures UV absorption at the capillary, perpendicular to the neutron beam, which provides a more reliable estimation of concentration.

### 4.4.1 *Current procedures for SEC-SAXS data analysis*

Cutting-edge software is constantly being developed to extract structural information from SEC-SAXS experiments. A comprehensive overview of these advancements is given in a recent report by Pérez *et al* (Pérez, Thureau, and Vachette, 2022).

In an ideal case, the species of interest is successfully separated from larger oligomers or aggregates and elutes as a distinct peak. By examining the evolution of the radius of gyration $R_g$ across the elution peak, it is possible to assess if the successive scattering frames represent a monomeric species. Identical SAS profiles can be averaged and used for further analysis. Easy-to-use softwares are available to perform this procedure including CHROMIXS from the ATSAS package (Panjkovich and Svergun, 2018) and BioXTAS RAW (Hopkins, Gillilan, and Skou, 2017).

One common but usually unproblematic situation sees a small unresolved shoulder on the left-hand side of the main elution peak. Consequently, a systematic decrease in the $R_g$ is observed along the left-hand side, while stable $R_g$ values are typically obtained in the centre and on the right. These frames with stable $R_g$ can be averaged and further analysed.

A much more difficult problem arises when multiple species are poorly separated under the resolution of the SEC column. To extract uncontaminated SAXS profiles, singular value decomposition (SVD) analysis can in some cases be employed on chromatograms to determine the number of species present, and then a scattering profile for each species can be extracted (Brookes et al., 2013; Hopkins, Gillilan, and Skou, 2017).

# 5

## MODELLING PROTEIN-LIPID COMPLEXES FOR SAS

SAS modelling is an ill-posed problem with a large solution space. There are many structures which could give rise to the same scattering profiles. This is especially problematic for experimental data with large error bars, where wide ranges of models could potentially fit the data equally well and lead to great uncertainty in the parameters of the fitted model.

Prior knowledge about the sample *must* be included in the analysis to increase the likelihood of finding a solution that is physically meaningful.

### 5.1 MOLECULAR CONSTRAINTS AND THE NANODISC MODEL

A number of strategies can be used to reduce the solution-space of analytical form factors, including the employment of molecular constraints (Hayter and Penfold, 1981), Bayesian regularisation (Larsen, Arleth, and Hansen, 2018) and joint refinement against multiple scattering profiles e.g. SAXS and SANS data. To prevent overfitting, which is a common problem in SAS, simple models with few fitting parameters relative to the information content should be selected. It's also important to keep in mind that fitting parameters in SAS are often highly correlated (e.g. increasing concentration, volume or scattering length density will increase the forward scattering intensity (Equation 4.13)).

Molecular constraints were introduced by Hayter and Penfold, 1981 where additional information such as sample concentration and chemical composition were systematically built into their model for SDS micelles, in order to reduce the number of fitting parameters. Our models continue with this philosophy. For example, in nanodisc samples the MSP concentration can be well-determined via UV-absorption and converted to particle number density. It is crucial to fit analytical models on absolute scale, since introducing a free scaling parameter also introduces a lot of ambiguity in the model.

The analytical model for nanodiscs (Skar-Gislinge and Arleth, 2011; Skar-Gislinge et al., 2010) is built as a collection of discs (Equation 4.14 and figure 4.7) using the form factor amplitude for cylinders (Pedersen, 1997). Scattering lengths for each component of the system are calculated from their molecular compositions and this should be almost always fixed during refinement. Partial specific volumes of lipids, $V_{\mathrm{lip}}$ and MSPs are usually also well-known, as is the mean area per lipid headgroup $A_{\mathrm{lip}}$

for many lipids. Then, for example, the disc model can be parameterised so that the height of the lipid layer is equal to $V_{\mathrm{lip}} / A_{\mathrm{lip}}$.

However, if there are too few fitting parameters the fit can be over-constrained, leading to poor fits and no new information on the fixed parameters. There is a balance in choosing model constraints. The ideal case would be where well-known model parameters are taken as fitting parameters, but the refined values become very close to their expected values as this adds an extra layer of confidence in the model results. The contrast situation in SAXS means the model is extremely sensitive to minor deviations of the partial specific molecular volumes and so these are usually fitting parameters in SAXS but not SANS.

The nanodisc model (Figure 4.7) is a geometrical models but it can be parameterised in different ways into more meaningful molecular parameters. In Paper 1 and 2/3 the nanodisc model has the following fitting parameters:

- Number of lipids per nanodisc

- Area per lipid headgroup

- Axis ratio which describes the ellipticity of the nanodisc

- Partial molecular volume of a single lipid

- Partial molecular volume of the MSP

- Height of the MSP belt

- Interface roughness, which smoothens the sharp interfaces between components of the model with different contrasts (Als-Nielsen and McMorrow, 2011)

- Background

In Paper I the goal was to maximise the amount of structural information which can be extracted from SEC-SAXS data from nanodiscs, by modelling multiple SAXS frames from the SEC-SAXS same data set. Fitting the nanodisc model to each SAXS frame individually could be tempting as the $\chi_R^2$ will probably be smaller, but this runs the risk for overfitting and misidentifying trends across the data set. It also ignores some of our *a priori* knowledge that the geometry of lipids and MSP is independent of elution volume. Therefore a preventative measure for overfitting larger data sets was explored. This is done by fitting the nanodisc model to multiple frames simultaneously and achieve good fits to the entire data set with relatively few parameters. From seven frames from the same data set, one value was refined for the partial specific volumes of the lipids, MSP, and for the mean area per lipid headgroup, respectively. This provides robust estimates which are not prone to overfitting. Additional the underlying polydispersity in the population,

which is inherent to self-assembly systems, could be quantified by allowing some parameters to vary frame-to-frame. These parameters were the number of lipids per nanodisc, and the ellipticity of the disc. These values are not fit to each frame individually, as this could again lead to overfitting and they might compensate for the constrained parameters. Instead a linear function across all frames in enforced, to ensure the fitted parameters are consistent frame-to-frame. This approach promotes a consistent, overarching understanding of the SEC-SAXS data set rather than focusing too closely on individual frames.

## 5.2 MODELLING MEMBRANE PROTEINS EMBEDDED IN NANODISCS

The next level of nanodisc modelling is calculating the theoretical scattering from membrane proteins embedded in nanodiscs. The major difficulty is determining the shape of the nanodisc alongside the membrane protein (MP). If the structure of the MP is unknown, the analytical nanodisc models can be combined with *ab initio* methods to reconstruct a low-resolution model of the MP (Orioli, Henning Hansen, and Arleth, 2021). *Ab initio* methods can also potentially be used to reconstruct the whole MP-ND complex (Childers et al., 2022; Morrison et al., 2022), though perhaps there are too many degrees of freedom to predict the structure with certainty (for now). If the structure of the MP is available, the analytical model for nanodiscs can be combined with a bead-based model to calculated the scattering from the MP (Kynde et al., 2014; Skar-Gislinge et al., 2015). This is relatively simple if the MP has a rigid structure, but becomes more complicated if the MP is flexible or adopts more than one state, since SAS profiles are the averaged scattering pattern from every conformation present.

The growth hormone receptor (GHR) embedded in a nanodisc is an example of a complex system which is pushing the boundaries of integrative structural biology (Paper II) and modelling for SAS data (Paper III). The GHR is a transmembrane protein with over 50% disorder. It has a folded extracellular (ECD) - and transmembrane domain (TMD), and a 350-residue-long intracellular domain (ICD) which is intrinsically disordered. An additional challenge to the 'usual' modelling strategies, is the green fluorescent protein (GFP) fused to the end of the ICD, which has neither a fixed position or orientation with respect to the rest of the system.

### 5.2.1 *Semi-analytical model for the GHR*

In Paper II, SAXS data on the GHR-GFP fusion embedded in nanodisc was collected. The first goal was to extract the nanodisc parameters from SAXS data from the complex. The semi-analytical (SA) model is split
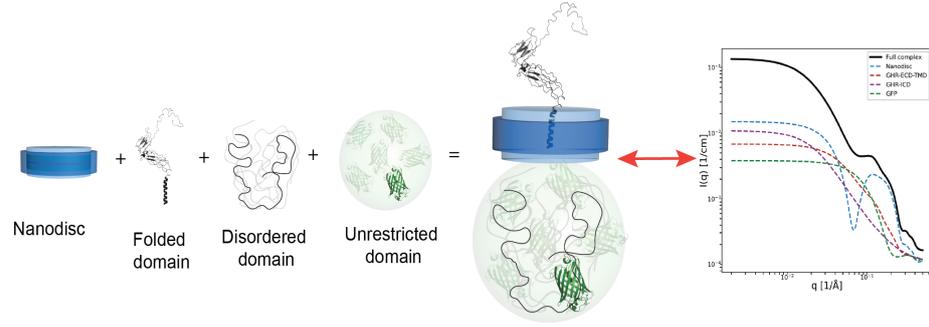
**Figure 5.1:** The architecture of the semi-analytical model for modelling the GHR-GFP embedded in a nanodisc. The nanodisc is described with the usual analytical model. The folded ECD-TMD is modelled from its atomic cooridnates. The disordered ICD is described as a Gaussian random coil. The GFP, which is fused to the end of the ICD, is evenly distributed within a spherical volume in order to capture its movement. The scattering amplitude from each component is calculated individually and then combined to calculate the total scattering intensity of the system. Figure is adapted from Paper III (Barclay et al., 2023).

into four components: the ND, the GHR-ECD-TMD, the GHR-ICD and the GFP. The model (Figure 5.1) has the architecture:

$$I(q) = n \cdot \langle |A_{\mathrm{ND}}(\vec{q}) + A_{\mathrm{ECD\text{-}TMD}}(\vec{q}) + A_{\mathrm{ICD}}(\vec{q}) + A_{\mathrm{GFP}}(\vec{q})|^2 \rangle_\Omega \quad (5.1)$$

Where $\langle ... \rangle_\Omega$ denotes orientational averaging, and each amplitude term contains the usual terms, $A(\vec{q}) = V \Delta \rho F(\vec{q}) \exp(i\vec{q} \cdot \vec{r})$ where $\exp(i\vec{q} \cdot \vec{r})$ and $\vec{r}$ is the position relative to the origin of the system. The scattering amplitude from each component is calculated individually and then combined to calculate the total scattering intensity of the system. The absolute square of the complex expression must be taken, resulting in four auto-correlation terms and 12 cross-correlation terms.

The ND is modelled with the usual analytical nanodisc model. For the GHR-ECD-TMD, the spherical harmonics method is used to calculate the scattering from an atomic structure, since it is a well-folded domain. Following the procedure by Kynde et al., 2014, the structure is positioned in the nanodisc so that the contrast of the beads inside the nanodisc are calculated with respect the the lipid environment instead of the buffer. For the ICD we use a Gaussian random coil (GRC) model. For the GFP we again use an atomic structure, and developed the concept of a 'distribution volume' to help capture its movement in solution. The fitting parameters are the nanodisc parameters listed above, and the average $R_g$ of the GHR-ICD. The model is described in great detail in Paper III and so only the most significant aspects will be discussed here.

**Gaussian random coils**

The disordered component is modeled with the form factor for Gaussian random coils, also known as the Debye model (Debye, 1947; Pedersen, 1997). It is a primitive model compared to the alternative, which would have to be an ensemble approach. The Debye model has a single structural free parameter, $R_g$. In polymer statistics, the $R_g$ follows a power-law relationship:

$$R_g = R_0 N^v \tag{5.2}$$

Where $N$ is the number of monomers (residues), $v$ is the scaling exponent and $R_0$ is a constant dependent on persistence length, which is the characteristic length over which a chain is rigid. $R_0$ is estimated to be close to 1.927 Å for denatured proteins (Kohn et al., 2004). The Debye equation implicitly assumes $v = 0.5$ for not self-avoiding chains. Flory estimated $v \approx 0.6$ for excluded-volume, nonzero thickness chains (Flory, 1953), which was corroborated experimentally for denatured proteins where $v = 0.598$ was found (Kohn et al., 2004). Still, the Debye equation generally works well for modelling SAS data from IDPs, as long as the IDP does not deviate too far from ideal random coil behaviour.

**Distribution volume**

The distribution volume is our analytical solution for modelling the co-existing positions of the GFP in solution. The GFP moves in all directions within some volume, constrained by the length of the GHR-ICD. It cannot simply be placed in one position in the model. The confusion volume is is an extension of work by Pedersen and Gerstenberg, 1996 and Pedersen, 2000, and manifests itself as a special 'phase factor'.

Pedersen and Gerstenberg, 1996 modeled micelles as a spherical core surrounded by polymer chains. They showed analytically that Gaussian random coils could be uniformly distributed on the outside of the sphere by convoluting the scattering density of the chains with an infinitely thin spherical shell. i.e the cross-correlation term between the sphere and the chains is given as:

$$S_{\text{sph.GRC}}(q) = F_{\text{sph}}(q, R) F_{\text{GRC}}(q, R_g) \mathbf{F_{shell}}(\mathbf{q}, \mathbf{R} + \mathbf{R_g}) \tag{5.3}$$

where $F_{\text{sph}}(q)$ is the form factor amplitude for a sphere with radius $R$, $F_{\text{GRC}}(q)$ is the form factor amplitude for Gaussian random coils (Hammouda, 1992) and $F_{\text{shell}}(q)$ is the form factor amplitude for infinitely thin spherical shells. $F_{\text{shell}}(q)$ is highlighted since it acts as the 'phase factor' which distributes the random coils properly. The shell is given a radius of $R + R_g$ to mimic non-penetration of the chains into the core.

In the GHR model, the same philosophy is followed to uniformly distribute the GFP within another volume. The scattering length density of the GFP is convoluted with the form factor amplitude of the shape of the
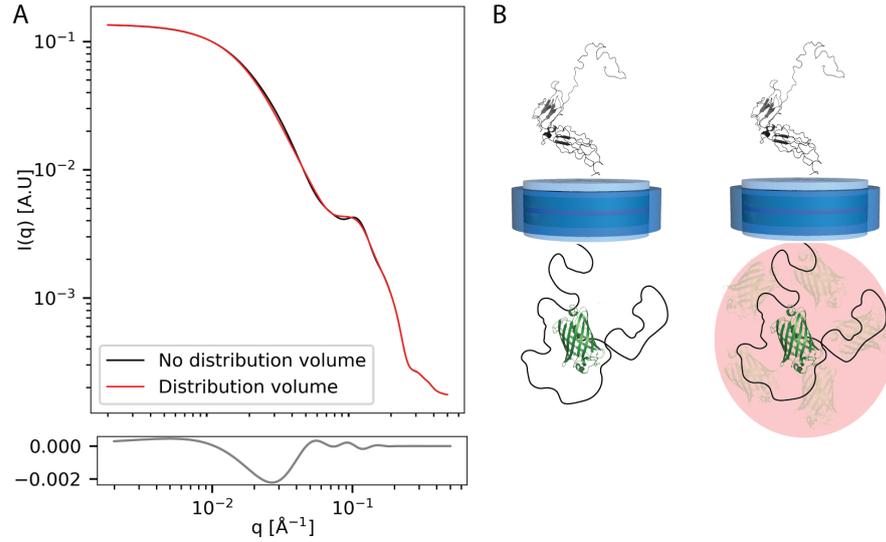
**Figure 5.2:** The need for the distribution volume. (A) Plots of the theoretical scattering profile from GHR-GFP-embedded in a nanodisc with the GFP placed in one position (black) and the GFP distributed within a distribution volume (red). The residual plot is shown below and shows variation in the mid-$q$ region. Modelling the distribution volume produces a smoother scattering profile. (B) Visual representation of the two models.

distribution volume, e.g. a sphere. The cross-correlation term between the GFP and any of the other three components (X = ND, ECD-TMD or ICD) is given by:

$$S_{\text{GFP.X}}(q) = F_{\text{X}}(\vec{q}, x) F_{\text{GFP}}(\vec{q}) \mathbf{F_{sph}}(\vec{\mathbf{q}}, \mathbf{R_g}^{\mathbf{ICD}}) \exp(i\vec{q} \cdot \vec{r}) \tag{5.4}$$

where the size of the spherical distribution volume depends on the $R_g$ of the GHR-ICD and $\vec{r}$ is the displacement of the centre of the sphere to the origin.

Figure 5.2 illustrates the distribution volume in this model. The effect is mostly visible in the mid-$q$ region of the scattering profile. Modelling the distribution volume smoothens the scattering profile, as would be expected for increased polydispersity in a sample.

Taken together, the Gaussian random coil model and the distribution volume are tools which can be used to analytically model flexible or dynamic systems. This approach is useful for extracting an overall description of flexible MP-ND complexes from SAS data, or as a precursor for higher-detailed or ensemble modelling. We also used the model as a tool to make quick theoretical calculations of how the complex would scatter under different contrast situations. We find that in standard SANS experiments, the excess scattering density from lipids is much greater than from protein and such a data set would hold almost no information about a membrane protein. In SAXS the scattering contribution from
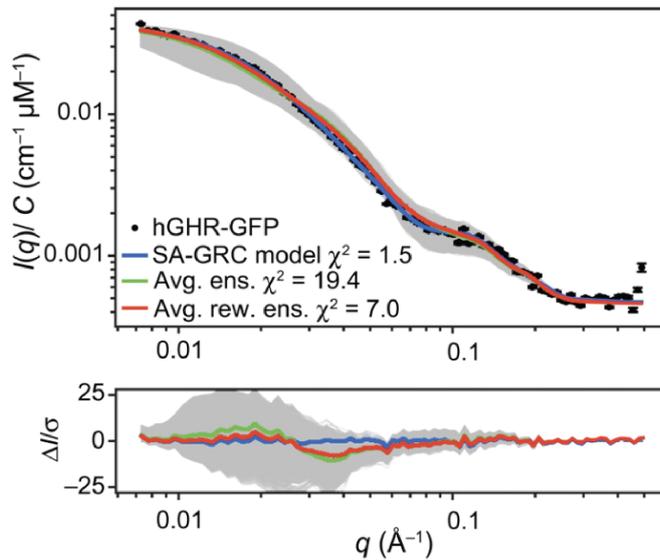
**Figure 5.3:** Results from the SAXS modelling of GHR-GFP-embedded nanodiscs in Paper II. Black: experimental data points. Blue: fit of the semi-analytical model. Green: averaged scattering profiles from the simulated ensemble of GHR-GFP in a nanodisc. Red: averaged scattering profile from the BME reweighted ensemble of GHR-GFP in a nanodisc. Grey: theoretical scattering profiles from various structures in the ensemble. Figure is taken from Paper II (Kassem et al., 2021)

lipids and protein is more equal, allowing the complex to be properly studied, including the disordered domains. The best-case scenario is making use of stealth nanodiscs for SANS (Maric et al., 2014), where the nanodisc is matched-out in 100% $D_2O$ and the scattering profile exclusively represents the MP. These experiments come at the cost of more expensive and laborious sample preparation, but come with the massive benefit of simpler and less-ambiguous analysis. Stealth nanodiscs have already been used successfully to study the low-resolution structures of MPs (Johansen et al., 2022; Josts et al., 2018) and are invaluable for SANS moving towards studies of more structurally complex and dynamic MPs. Although it may seem intuitive, the importance of conducting preliminary model calculations before setting off on a beam time or any beginning computationally-heavy analysis is stressed, in order to guide studies that are time- and resource-efficient. The semi-analytical model could naturally be applied to other multi-domain systems where one domain is free to move with respect to the other.

### 5.2.2 *Integrating simulated trajectories*

In Paper II, course-grained molecular dynamics simulations of the full-length GHR-GFP in a POPC lipid bilayer were performed. The simulations integrated experimental data from different sources along the way,

and resulted in a broad structural ensemble of the protein. The final step was to compare the simulated ensemble of 6000 GHR-GFP frames with the SAXS data from the GHR-GFP in a nanodisc.

The 6000 GHR-GFP conformations were extracted and transferred to the nanodisc model one-by-one. The nanodisc parameters were kept fixed at the values refined from the semi-analytical model described above (Section 5.2.1). The averaged scattering intensity from the entire ensemble was calculated with:

$$I(q) = \frac{1}{N} \sum_{i=1}^{N} \langle |A_{\mathrm{ND}}(\vec{q}) + A^i_{\mathrm{GHR\text{-}GFP}}(\vec{q})|^2 \rangle_\Omega \qquad (5.5)$$

Where $i$ is the frame number and $N$ is the total number of frames. Even though this was not a refinement, the comparison to the SAXS data showed good agreement. BME reweighting was employed to improve the agreement further, and by discarding 25% of the simulated conformations, the $\chi^2_R$ decreased from 7 to 19. The final ensemble therefore contains some information about the conformational preferences of the GHR-GFP complex which was encoded in the SAXS data.

After reweighting the agreement with the SAXS data is good, but the $\chi^2_R$ from the semi-analytical model is much better. Although the semi-analytical model was designed to closely mimic the GHR-GFP ensemble and create the same scattering profiles, there are differences between the two. For one, the semi-analytical model accounts for the ECD-TMD with a single structure while the simulation shows a broad ensemble of structures due to a flexible linker between the ECD and TMD. Another discrepancy could be the Gaussian random coil and distribution volume approximations. The refinement of the semi-analytical model landed on $R_g^{\mathrm{ICD}} = 73 \pm 13$ Å, while the averaged $R_g^{\mathrm{ICD}}$ from the simulated conformations was $62 \pm 4$ Å. Regardless, in conclusion, the semi-analytical model successfully captured the ordered and disordered regions of the GHR-GFP-embedded in a nanodisc. The nanodisc parameters were extracted and used to validate the simulated conformations in order to obtain a highly detailed description of the system.

## 5.3    AN ALTERNATIVE METHOD FOR MODELLING FLEXIBLE MEMBRANE PROTEINS IN NANODISCS

In the case of the GHR, perhaps an even more representative description of the system could have been obtained if the nanodisc parameters were optimised for the simulated ensemble directly. A method like this could only be feasible if the computational calculation is quick, since the scattering intensity must be computed for every protein conformation in the ensemble, for every iteration of the refinement. This could easily spiral into weeks of refinement to obtain a set of nanodisc parameters.

```
Parameters["Axis ratio of bilayer"]       = 1.40
Parameters["Minor radius"]                = 45.0
Parameters["Height of belt"]              = 25.78
Parameters["Hydrophobic height"]          = 29.4
Parameters["Volume of one tail"]          = 943.0
Parameters["Volume of one headgroup"]     = 324.0
Parameters["Volume of one belt protein"]  = 35000.8
Parameters["Volume of one solvent molecule"] = 30.0
Parameters["Rotation of membrane protein"]  = 0.0

Parameters["Scattering length of one tail"]            = 721.0
Parameters["Scattering length of one headgroup"]       = 462.0
Parameters["Scattering length of one belt protein"]    = 37290.0
Parameters["Scattering length of one solvent molecule"] = 28.2
```

**Figure 5.4:** Building point-cloud nanodiscs. The system starts with a cube of randomly distributed points. Using the nanodisc parameters, the nanodisc shape is cut-out. The points are assigned a scattering length according to whether they are in a lipid headgroup, tailgroup or MSP belt region. If a protein structure is provided a hole is made in the nanodisc for the protein to sit it.

In this section a method based on Fast Debye Sums (FDS) (Section 4.2.2) will be described. While the FDS method is not very popular in the SAS comunity, it was recently re-visited by Pedersen et al., 2022, as outlined on bioRxiv alongside some relevant software. Pedersen et al., 2022 suggests that this method is computationally attractive for handling large ensembles of protein conformations, in order to refine scattering profiles from flexible membrane proteins in nanodiscs.

The model system presented here is the growth hormone secretagogue receptor (GHSR) (Bockaert and Pin, 1999; Guillien et al., 2022) embedded in nanodiscs with POPC and POPG lipids and lsMSP1E3D1. The GHSR has 20 disordered residues at both the C- and N-terminus. The SAXS data was kindly provided by Nathalie Sibillie (*Centre de Biologie Structurale, CNRS*).

### 5.3.1  *Point-cloud nanodisc models*

The nanodisc is represented as a dense cloud of points where each point is assigned some scattering properties according to where it is in the structure (MSP, lipid headgroups, lipid CH2 chains, or CH3groups). The nanodisc point-clouds are created with geometric shapes which are specified by the fitting parameters of the model. The fitting parameters are identical to those listed in Section 5.1 for the analytical nanodisc model.
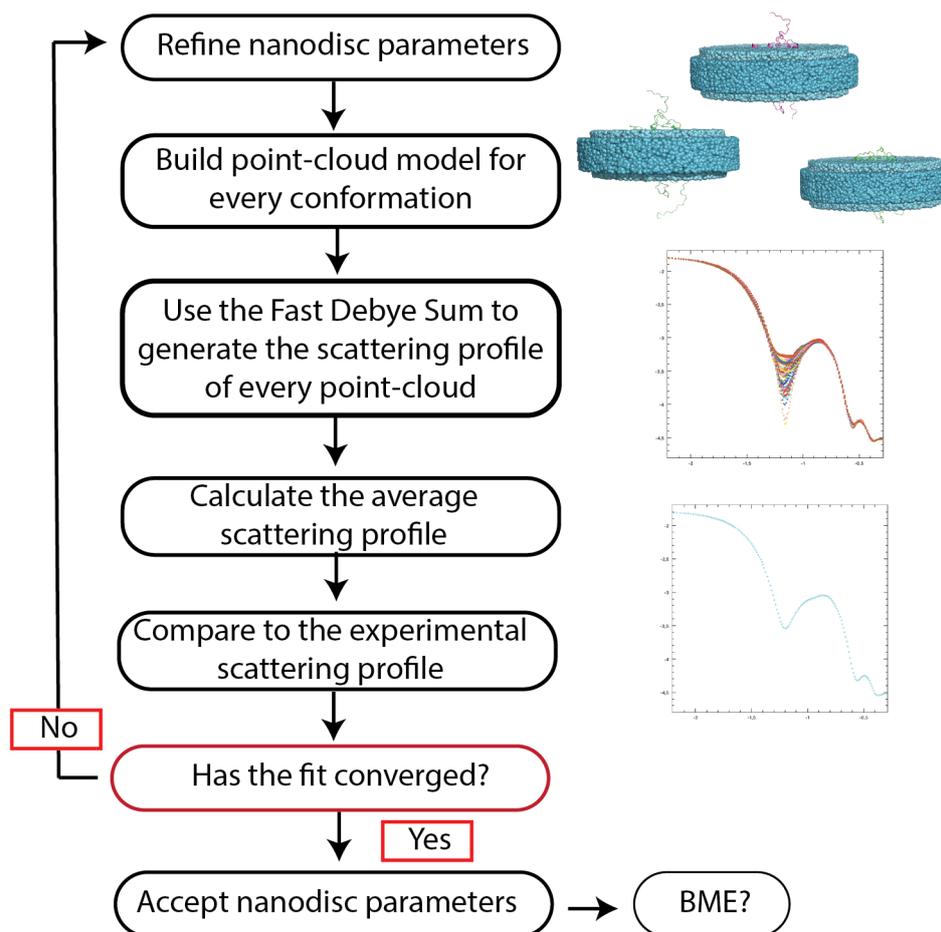
**Figure 5.5:** Refinement strategy for optimising one set of nanodisc parameters for an entire ensemble of protein structures.

Furthermore, protein structures can be inserted into the nanodisc representation. The atoms of the protein structure displace points in the nanodisc to make a hole to sit in. The recipe is shown in Figure 5.4. The MP structure is then included in the calculation of the $p(r)$ distribution. The scattering intensity which would be produced from such a MP-ND point-cloud can then be calculated using FDS (Figure 4.5).

Pedersen et al., 2022 show that point-clouds made up of 25,000 points are sufficient for accurately calculating the scattering intensity profiles in the $q$-range of a typical SAS experiment. They report it takes $\approx 0.5$ seconds to compute the scattering from one point-cloud made up of 25,000 points on a laptop with an NVidia Quadro 290 P2000 GPU. The speed of this model means optimising nanodisc parameters from SAS data for an entire ensemble of membrane protein structures is realistic. For 10,000 structures the refinement could take approximately one week. The refinement strategy is shown in Figure 5.5 .
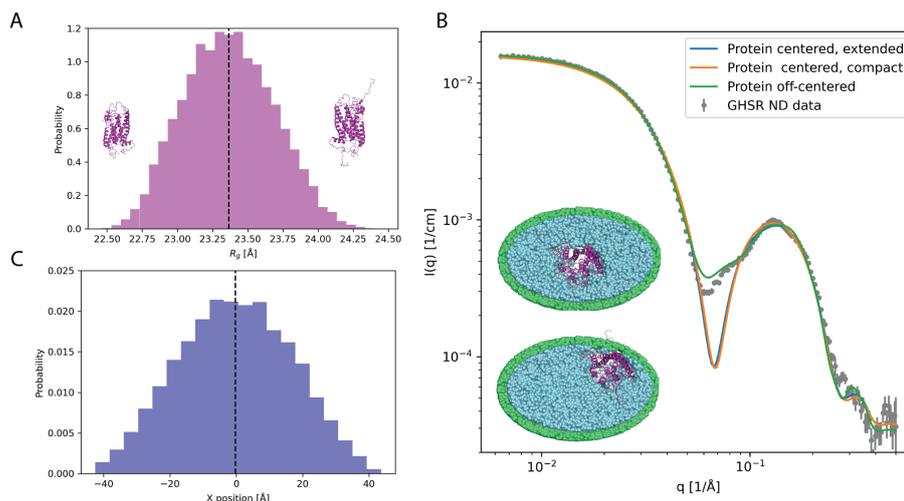
**Figure 5.6:** (A) $R_g$ distribution of the GHSR ensemble. (B) Grey: experimental SAXS data. Blue: Theoretical scattering profile from an extended protein conformation positioned in the centre of the disc. Orange: Theoretical scattering profile from a contracted protein conformation positioned in the centre of the disc. Green: Theoretical scattering profile where the protein is positioned close to the rim of the disc. The insets show the point-cloud models used to generate the blue and green scattering profiles. (C) Distribution of positions around the centre of the nanodisc for the ensemble of structures.

### 5.3.2  *Analysis of GHSR-embedded nanodiscs*

A structure for the folded domain of GHSR (residues 20 to 356) was already available, meaning 19 residues were missing from both ends. Flexible-meccano (Ozenne et al., 2012) was used to generate conformations of the intrinsically disordered N- and C-terminus tails resulting in 10,000 structures in total. Side chains were reconstructed with PULCHRA (Rotkiewicz and Skolnick, 2008). The $R_g$ of the GHSR structures in the ensemble appear normally distributed around 23.4 Å, with minimum 22.5 Å and maximum 24.5 Å, as shown in Figure 5.6 A.

Quickly into the analysis it was discovered that the experimental SAXS data from the GHSR-embedded nanodiscs cannot be modelled accurately with the GHSR centred in the nanodisc. Rather, there is a better fit to the shoulder in the mid-$q$ bump when the GHSR is placed closer to the rim. Furthermore, extrapolating from Figure 5.6 B, it is hypothesised that the best solution to capture the mid-$q$ bump and depth could be an (weighted) average of profiles with the protein in the centre and off-centre. We therefore normally distributed the 10,000 structures around the centre of the nanodisc prior to the model refinement (Figure 5.6 C). The conformation of the flexible tails, whether extended or compact, did not have a noticeable effect on the SAXS profiles (5.6 B).

The refinement strategy outlined in Figure 5.5 was used to fit the data, thereby optimising the nanodisc parameters for all 10,000 structure
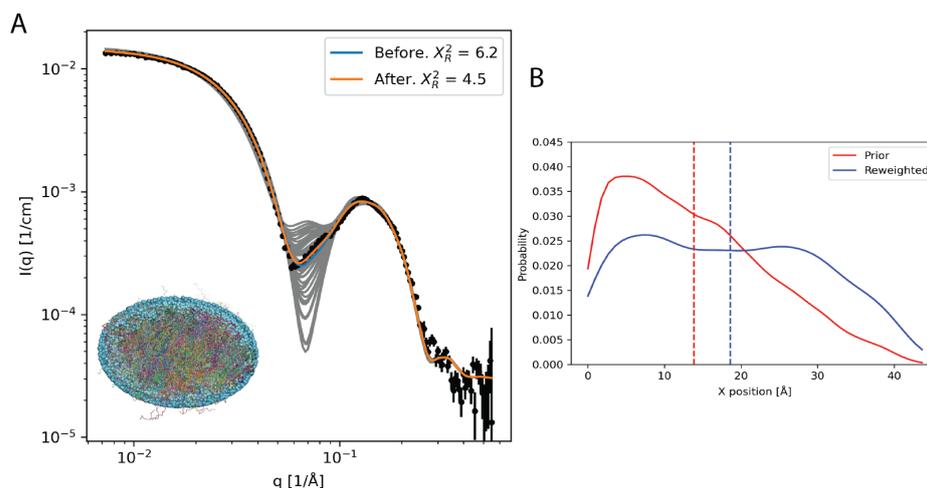
**Figure 5.7:** (A) Model fits to the experimental SAXS data. Blue: fit of the GHSR ensemble nanodisc model. Orange: averaged scattering profile after BME reweighting. Grey: scattering profiles from single GHSR-ND conformations. (B) Distribution of the distances of the protein from the centre, prior (red) and after (blue) reweighting. Dashed lines indicate the mean value.

simultaneously. The model results are shown in Figure 5.7 and Table 5.1. The overall model provides a good fit to the data and maintains reasonable fitting parameters for the GHSR-NDs. However, the double bump in the mid-$q$ minimum is not captured.

| Parameter | Value | Expected |
|---|---|---|
| Axis ratio | $1.03 \pm 0.52$ | |
| Area per headgroup [Å$^2$] | $59.0 \pm 1.9$ | 62.7* |
| Number of lipids | $441 \pm 54$ | 400† |
| Volume of one MSP [Å$^3$] | $40400 \pm 7000$ | 36900‡ |
| Volume of one lipid [Å$^3$] | $1200 \pm 20$ | 1240* |
| Roughness [Å] | $4.11 \pm 0.70$ | 3 - 5 |
| Background [cm$^{-1}$] | $3.00 \cdot 10^{-5} \pm 1.33 \cdot 10^{-5}$ | |
| $\chi_R^2$ | 6.2 | |

**Table 5.1:** Structural nanodisc parameters refined from the data in Figure 5.7.*(Kučerka, Nieh, and Katsaras, 2011). †Reconstitution ratio. ‡Calculated using an average mass density of proteins 1.35 cm$^2$ g$^{-1}$(Mylonas and Svergun, 2007). #Data are not on absolute scale so a free scale parameter is required.

This method lends itself well to BME, since a SAXS profile is calulated for every structure in the ensemble. The individual SAXS profiles were reweighted against the SAXS data to try to pinpoint preferred positions of the GHSR in the ND. By discarding 66% of the pool of conformations, the $\chi_R^2$ improves slightly but the should is still not properly captured. Looking at the reweighted histogram of the distance of GHSR from the centre (Figure 5.7 B) , conformations with GHSR closer to the rim are upweighted and the average distance increases from 14 Å to 19 Å. Since

this trend has been identified, it could be studied in more depth with another round of SAXS modelling, or if the SAXS analysis was combined with MD or cryo-EM.

Displacement of the membrane protein in a nanodisc towards the rim has been documented before (Baas, Denisov, and Sligar, 2004; Kynde et al., 2014; Orioli, Henning Hansen, and Arleth, 2021) where one average position was refined from SAXS data and from MD simulations where BAX in a nanodisc was shown to interact with the MSP (López et al., 2019). Very recently, Kim, Koh, and Roh, 2023 investigated many EM data sets of MPs in nanodiscs and found a distinct preference of MPs to be close the ND edge.

To conclude, the goal of this section was to further explore the method presented by Pedersen et al., 2012 for determining and analysing ensembles of flexible membrane proteins in nanodiscs with SAS data. The advantage of point-cloud nanodiscs over form factor models is the reduced computational cost. The method easily facilitates the integration of MD simulations of membrane proteins and SAS data. It is a more efficient and robust way to refine models of MP-NDs directly against SAS data without additional steps.

## 5.4    MODELS FOR $\alpha$-SYN-LIPID CO-STRUCTURES

Manuscript I uses SAXS and contrast variation SANS to investigate large-scale morphological changes to negatively charged lipid structures (DLPS or DMPS) induced by binding with $\alpha$-Syn. It is apparent from the data that a large-scale reorganisation of vesicles (or large planar structures), into small discs or rod-like structures occurs, driven by the amphiphilic properties of the system (Figure 5.8). These transformations were already clear from visual investigation of the data, particularly from the corresponding $p(r)$s and the changes to the initial slope of the SAS profiles. These observations were taken as the starting point for direct-modelling of the data.

The models use core-shell form factors to represent the different scattering properties of lipid headgroups and tailgroups. The models used to describe the $\alpha$-Syn-lipid co-structures closely follow the models presented by Pedersen and Gerstenberg, 1996 and Arleth et al., 2005. Essentially the nanodisc model described above is combined with the hairy micelle model (Arleth et al., 2005), where we use the form factor for a disc-shaped bilayer with (multiple) Gaussian random coils attached to its surface. The $\alpha$-Syn is represented by the protein belt and the Gaussian random coils represent the hydrophilic tail of $\alpha$-Syn which is not expected to interact with the lipids. The form factor contains four different terms: the auto-correlation term of the disc, the auto-correlation terms of the chains, the cross-correlation term between disc and the chains, and the cross-correlation terms between the chains.
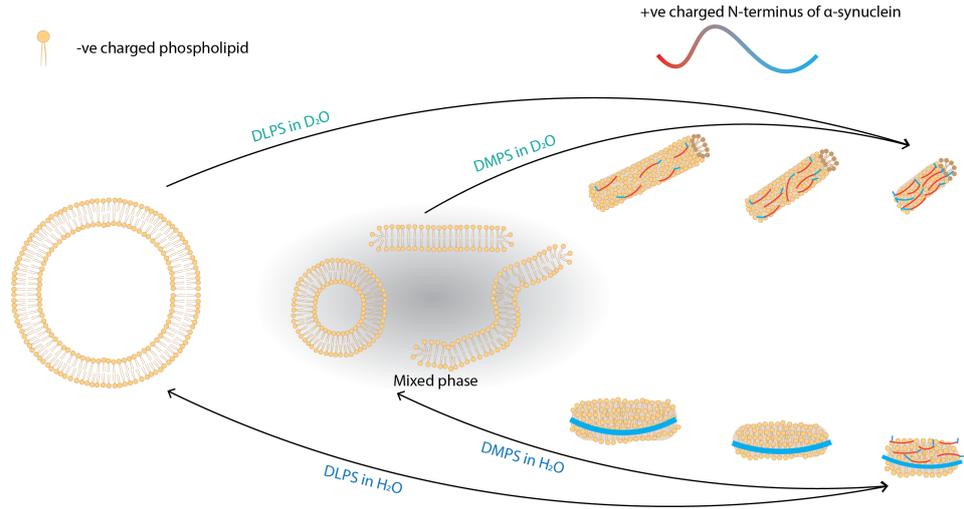
**Figure 5.8:** Overview of the results from the SAXS and SANS experiments in Manuscript I. The SAXS experiments are conducted in $H_2O$ with protonated $\alpha$-Syn. The SANS experiments are conducted in $D_2O$ with deuterated $\alpha$-Syn. Pure DLPS forms rigid vesicles, while pure DMPS appears to form planar structures or mixed micelles. However, upon binding with $\alpha$-Syn, both DLPS and DMPS structures break down into smaller particles with the same shape. In the SANS experiments rod-like co-structures are formed. In the SAXS experiments disc-like co-structures are formed. The SAXS experiments show the reorganisation is completely reversible to the original lipid structures upon removal of $\alpha$-Syn. The figure is from Manuscript I.

$$
\begin{aligned}
P(q) =& P_{\text{disc}}(q) + N_{\text{GRC}}\Delta b_{\text{GRC}}^2 P_{\text{GRC}}(q) + 2N_{\text{GRC}}\Delta b_{\text{GRC}} S_{\text{disc.GRC}}(q) \\
& + N_{\text{GRC}}(N_{\text{GRC}} - 1)b_{\text{GRC}}^2 S_{\text{GRC.GRC}}(q)
\end{aligned}
\tag{5.6}
$$

where $P_{\text{disc}}(q)$ is the form factor for the nanodisc already including different contrasts for lipid headgroups and tailgroups, $N_{\text{GRC}}$ is the number of chains attached to the disc, $\Delta b_{\text{GRC}}$ is the excess scattering length of $\alpha$-Syn chains, $P_{\text{GRC}}$ is the form factor for Gaussian random coils, $S_{\text{disc.GRC}}$ is the cross-correaltion term between the nanodiscs and chains, and $S_{\text{GRC.GRC}}$ is the cross-correlation term between the many chains attached to the disc.

### 5.4.1    *Reorganisation of $\alpha$-Syn-lipid co-structures*

The disruption and remodelling of DLPS vesicles into discoids, tubules, and ribbons upon binding with $\alpha$-Syn was recently demonstrated by Hoover et al., 2021 with transmission-EM images. Small-angle scattering is able to provide valuable insights into the initial aggregation of $\alpha$-Syn with the lipids, and the underling mechanism which leads to fibril formation. The co-structure can be studied in solution without further sample manipulation. Furthermore, SAXS was able to capture the reorganisation on a ms time scale, investigate temperature dependence and reversibility

of the system, and monitor individual components through contrast variation. This information is not easily accessible by other techniques.

A distinct isotope effect between SAXS ($H_2O$ solvent and protonated $\alpha$-Syn) and SANS ($D_2O$ solvent and deuterated $\alpha$-Syn) was observed. In the SAXS experiments we observe a reversible transformation of large planar structures into small discs. In the SANS experiments, a transformation into small rod-like structures is observed. We believe these two states are thermodynamically similar, and that a small difference in the sample can push the system from one state to the other.

The transformation of lipid vesicles into discs or rods/elongated structure can be triggered by sample conditions such as environmental factors and the relative concentrations of the components. Lipid bilayers are balanced by the hydrophobic effect of lipid headgroups and tailgroups, and it is often energetically favourable for bilayers to form vesicles. When introducing $\alpha$-Syn it is drawn to the curved surface of the vesicle and may disrupt the surface to find a new balance where the protein shields hydrophobic edges of the planar lipid bilayer from the aqueous environment. The phase behaviour is reminiscent of an extensive study of the mixing of lipids and detergents (Singh, 2009) where a discoidal phase was observed at low temperatures and an elongated, non circular phase was observed at higher temperatures (above the melting temperature) for the same system. However, in the system presented in Manuscript I it is not clear exactlywhat factor determines the resulting structure.

## 5.5  SUMMARY

In biological SAS, models are not one-size-fits all and they must be tuned to the unique characteristics of each biolmolecular system in question. Direct data analysis, particularly Guinier fits and $p(r)$ distributions, can be used to obtain a general overview of the shape and size of the system and make hypotheses. For example, in Paper I, the $R_g$ calculated from the nanodisc SAXS data was found to decrease moving across the SEC peak, which prompted deeper structural analysis into the underlying polydispersity of the sample. Paper II contains an example of poorly-resolved SEC-SAXS and SEC-SANS elution profiles. Inspection of the $R_g$ across the SEC peaks indicated the data contained an ordered distribution of dimeric protein on the left-hand side to monomers on the right-hand side. A basic strategy was adopted where the SAXS data to the very right of the peak was assumed to contain uncontaminated monomeric sample and these frames were selected for structural modeling. It would have also been interesting to try one of the specially designed deconvolution softwares for this problem. Finally, in Manuscript I, inspection of the $p(r)$ distributions made it obvious that the data sets did not contain pure lipid: $\alpha$-Syn vesicles as first expected, but that there was a change of structure upon binding into either discs or rods.

This was then used as the starting point for various analytical models to attempt to fit the data.

These 'model-free' information become less insightful when dealing with complex or heterogeneous samples. Furthermore, the solution space in SAS is large as the same scattering profile can arise from many different structures. To address this, prior knowledge must be incorporated into direct modelling in order to select the appropriate form factors or structural parameters to accurately describe the system. All of the analytical models and the point-cloud model presented in this chapter enforce molecular constraints, including exploiting sample concentrations and including the chemical composition and volumes of all lipids, MSP and $\alpha$-Syn.

Although SAS models are usually system specific, the underlying principles presented in the models here can be transferred to aid future studies of protein and lipid complexes. The procedure laid out in Paper I could be applied to any polydisperse system where an analytical model is available. The architecture of the semi-analytical model (Equation 5.1) is built in such a way that components can be inserted or removed to best describe other systems with similar topology. The model enables SAS to be combined with high-resolution structures from x-ray crystallography, cryo-EM and AlphaFold, by accounting for IDRs or other rigid domains which are free to move in solution. The FDS model for point-clouds can be employed for SAS modelling of any membrane protein with conformational diversity, for example flexible domains or varying positions in the nanodisc.

6

CONCLUDING REMARKS

The overall aim of this PhD thesis was to develop methods or tools that optimise the amount of information which can be extracted from experimental data, while simultaneously investigating structural aspects of selected biomolecular systems. The methods developed here will hopefully facilitate the movement in structural biology towards uncovering the structure and mechanisms of challenging systems of biomolecules and macromolecular assemblies.

The majority of the thesis focused on how the limits of small-angle scattering can be expanded, starting with instrumentation and a procedure to fully utilise the information content in SEC-SAXS data sets. Paper I presents an advanced fitting platform, where analytical models are refined against multiple frames from the same SEC-SAXS data set simultaneously. Global and frame-specific parameters can be employed to investigate trends across the data set. This is particularly relevant for samples of single species which elute in a well-resolved SEC peak, but where there is an underlying size distribution. The approach achieves a detailed and robust overview of the entire data set without overfitting, which is often a problem in small-angle scattering. Paper I investigates the size and shape polydispersity for three next-generation circularised and supercharged nanodiscs populations, but the method is well-suited to other self-assembly systems which naturally contain some degree of polydispersity.

Paper III and Manuscript I focus on interpreting SAS data from protein: lipid complexes and building advanced analytical models. For these multiplexed systems it is not possible to isolate the scattering signal from each component, and models with molecular constraints are especially important for uncovering the architecture of the system and refining structural parameters from the data.

The semi-analytical model for the growth hormone receptor in a nanodisc employs the well-established analytical model for nanodiscs in combination with atomic structures of the folded protein domains. The real challenge behind modelling this system was the large intrinsically disordered region and the GFP fused to the end which does not have a fixed position. We showed that the disordered region could be approximated as a Gaussian random coil attached to the nanodisc. Furthermore, we developed the concept of a 'distribution volume', where the scattering intensity arising from the GFP is evenly distributed within some volume.

This means the average scattering intensity over many conformations of the GFP can be quickly calculated. This is a simplified model compared to an ensemble approach, but it is sufficient for comparison with SAS data. The semi-analytical model is an inexpensive calculation. Its adaptability means it can be used in other contexts to make predictions of flexible, multi-domain complexes for refinement against SAS data or as a check ahead of experiments. The model can be used to account for scattering arising from disordered regions alongside scattering from high-resolution structures of folded domains determined by cryo-EM or predicted by AlphaFold2. Therefore, a complete description of the system can be refined from SAS data.

In the modelling strategy presented in Paper II, the nanodisc parameters from the semi-analytical model were combined with the simulated ensemble of GHR structures. The the theoretical scattering from the simulated structures inserted in the nanodisc model agree well with the SAXS data. BME reweighting was then performed to reduce computational inaccuracies and generate an ensemble which was even more representative of the data.

In chapter 5, an alternative method for refining models for flexible proteins in nanodiscs against SAS data was explored. Using the Fast Debye Sum (FDS) method, the nanodisc parameters can be optimised for every conformation in an ensemble simultaneously in order to find the best set of parameters to describe the entire ensemble. This is the most direct integration of SAS data from membrane proteins in nanodiscs and simulations. The use of FDS makes the method computationally feasible. In the example in chapter 5, the computational efficiency of FDS was used to investigate the position of the GHSR protein in a nanodisc. The protein was found to have a preference towards the rim of the nanodisc. Hopefully in the future this method can be explored further in combination with MD. This method is a very attractive choice for future analysis for SAS data from flexible membrane proteins in nanodiscs.

Paper II presents the structure for the GHR and demonstrates the power of integrative structural biology, particularly the combination of SAXS and MD. Through a 'divide and conquer' approach, data were first collected on the individual domains of the GHR. E.g. SAXS, MD and BME facilitated building the model of the extraceullular domain, chemical shifts from NMR were used to build the structure of the $\alpha$-helical transmembrane domain (TMD), and x-ray diffraction found the tilt of the TMD with respect to the membrane. The information were integrated into one MD simulation, and the simulation was verified against SAXS data from the full-length GHR. The results showed a broad ensemble of conformations, emphasising the critical importance of taking flexibility into account when studying proteins in any capacity. This is an exemplary method which can be applied to other membrane proteins

with structural disorder, particularly signaling, binding and regulating proteins which often contain large disordered chains.

Where possible, cryo-EM is an additional complementary technique to studies of MPs, which can provide high-resolution structures of biomolecules in their native state and give some insight into their dynamic conformations. Cryo-EM structures can provide the starting point for MD, which can then incorporate dynamics and be refined against SAS. Together, MD and SAS have the potential to fill in the disordered gaps that cannot be captured with cryo-EM, all contributing to a more comprehensive understanding of the system.

Continuing these kinds of investigations with SAS is advantageous since data can be recorded from the MP in a nanodisc under different conditions, e.g. different types of lipids or with binding partners or ions. Particularly exciting is the development of stealth nanodiscs (Maric et al., 2014) or invisible detergents (Midtgaard et al., 2018) which are matched-out in 100% $D_2O$ when investigated by SANS. This means the resulting scattering profile only contains signal from the membrane protein, which significantly eases structure determination and the subsequent data analysis, making the technique more accessible to nonspecialists.

In Manuscript I the structure of $\alpha$-Synuclein with negatively charged DMPS or DLPS was investigated. The presence of certain lipids have been shown to accelerate the formation of fibrils of $\alpha$-Syn which are known to play a central role in neurological diseases. The goal of the SAXS and SANS experiments was to investigate the initial interaction between the protein and lipids. The SAXS and SANS data show an obvious transformation from vesicles / large bilayers into other kinds of smaller structures. Firstly the data were investigated with model-free methods, where $p(r)$ distribution functions proved essential for grasping an understanding of the system. Then analytical models were refined against the data to confirm our hypotheses, which were somewhat surprising. The analytical models incorporate molecular constraints and are calculated on absolute scale to increase their reliability. In SAXS (100 % $H_2O$ buffer and hydrogenated $\alpha$-Syn) a transformation into predominantly disc-like structures is observed and in SANS (100 % $D_2O$ buffer and deuterated $\alpha$-Syn) predominately elongated rod-like structures is observed. This kind of phase transition has been observed for mixtures of lipids and detergents (Singh, 2009), but it is not clear what causes the isotope effect in our study. This sensitivity of the system should be investigated further. Apart from the large-scale structural reorganisation of the lipids, the SAXS data indicates that the transformation is fully reversible by removing the $\alpha$-Syn from the lipids, and that the reorganisation occurs on fast timescales of < 1 ms. Finally, contrast-matched SANS experiments investigate fibril formation over many hours and demonstrate that DLPS lipid do in fact become incorporated into the fibril structures.

The $\alpha$-Syn study is an example of a system which would have benefited significantly from some of the recent advancements at the ILL and the upcoming ESS. In particular, data points collected at very low $q$ values is necessary to determine the large size of these co-structures and allow a proper quantification of the particle break-down. Furthermore, simultaneous SAXS and SANS measurements would make the investigation less ambiguous. Improvements large-scale facilities can definitely lead to exciting insights of biological processes which were not possible before, and ultimately drive advancements in the field of structural biology, particularly when integrated with other kinds of information.

Finally, in Manuscript II, an integrative method which was commonly used to simulate the transition state ensemble (TSE) of protein folding in the early 2000s was applied to study the TSE of PI3K-SH3 amyloid elongation. Since the TSE lies at the top of a free energy barrier, it cannot be sufficiently sampled with standard MD. Experimental data in the form of $\Phi$-values can be integrated to bias the simulations towards the transition state. The all-atom simulations revealed the full structure of the TSE of a monomer and the fibril end, including key interaction sites which facilitate binding. The structures in the TSE were cross-validated with experimental $\Delta\Delta G$ data. Although PI3K-SH3 fibrils are not associated with a physiological disease, this same protocol could be applied to other amyloid fibrils which play a role in neurodegenerative diseases, where understanding the structure and interactions could help to identify molecules to inhibit the formation or disrupt the stability of the fibrils.

In summary, this thesis contributes to the overarching theme of pushing the boundaries of structural biology using innovative computational approaches. By synergizing techniques like SAS and MD with other integrative methods, we can continue to tackle complex biological systems and advance our understanding of protein structures and functions.

# BIBLIOGRAPHY

Ahmed, Mustapha Carab, Ramon Crehuet, and Kresten Lindorff-Larsen (2020). "Computing, analyzing, and comparing the radius of gyration and hydrodynamic radius in conformational ensembles of intrinsically disordered proteins." In: *Intrinsically disordered proteins: methods and protocols*, pp. 429–445.

Akey, Christopher W, Digvijay Singh, Christna Ouch, Ignacia Echeverria, Ilona Nudelman, Joseph M Varberg, Zulin Yu, Fei Fang, Yi Shi, Junjie Wang, et al. (2022). "Comprehensive structure and functional adaptations of the yeast nuclear pore complex." In: *Cell* 185.2, pp. 361–378.

Alber, Frank, Svetlana Dokudovskaya, Liesbeth M Veenhoff, Wenzhu Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprapto, Orit Karni-Schmidt, Rosemary Williams, Brian T Chait, et al. (2007). "Determining the architectures of macromolecular assemblies." In: *Nature* 450.7170, pp. 683–694.

Allison, Jane R, Robert C Rivers, John C Christodoulou, Michele Vendruscolo, and Christopher M Dobson (2014). "A relationship between the transient structure in the monomeric state and the aggregation propensities of $\alpha$-synuclein and $\beta$-synuclein." In: *Biochemistry* 53.46, pp. 7170–7183.

Als-Nielsen, Jens and Des McMorrow (2011). *Elements of modern X-ray physics*. John Wiley & Sons.

Andersen, Ken H, Dimitri N Argyriou, Andrew J Jackson, J Houston, Paul F Henry, Pascale P Deen, Rasmus Toft-Petersen, P Beran, M Strobl, T Arnold, et al. (2020). "The instrument suite of the European Spallation Source." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 957, p. 163402.

Andreas, Loren B, Marcel Reese, Matthew T Eddy, Vladimir Gelev, Qing Zhe Ni, Eric A Miller, Lyndon Emsley, Guido Pintacuda, James J Chou, and Robert G Griffin (2015). "Structure and mechanism of the influenza A M218–60 dimer of dimers." In: *Journal of the American Chemical Society* 137.47, pp. 14877–14886.

Anfinsen, Christian B (1973). "Principles that govern the folding of protein chains." In: *Science* 181.4096, pp. 223–230.

Araya-Secchi, Raul, Katrine Bugge, Pernille Seiffert, Amalie Petry, Gitte W Haxholm, Kresten Lindorff-Larsen, Stine Falsig Pedersen, Lise Arleth, and Birthe B Kragelund (2023). "The prolactin receptor scaffolds Janus kinase 2 via co-structure formation with phosphoinositide-4,5-bisphosphate." In: *Elife* 12, e84645.

Arleth, Lise, Beena Ashok, Hayat Onyuksel, Pappannan Thiyagarajan, Jaby Jacob, and Rex P Hjelm (2005). "Detailed structure of hairy mixed micelles formed by phosphatidylcholine and PEGylated phospholipids in aqueous media." In: *Langmuir* 21.8, pp. 3279–3290.

Auluck, Pavan K, Gabriela Caraveo, and Susan Lindquist (2010). "α-Synuclein: membrane interactions and toxicity in Parkinson's disease." In: *Annual review of cell and developmental biology* 26, pp. 211–233.

Baas, Bradley J, Ilia G Denisov, and Stephen G Sligar (2004). "Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment." In: *Archives of biochemistry and biophysics* 430.2, pp. 218–228.

Bansal, Akanksha, Matthias Schmidt, Matthies Rennegarbe, Christian Haupt, Falk Liberta, Sabrina Stecher, Ioana Puscalau-Girtu, Alexander Biedermann, and Marcus Fändrich (2021). "AA amyloid fibrils from diseased tissue are structurally different from in vitro formed SAA fibrils." In: *Nature Communications* 12.1, p. 1013.

Barclay, Abigail, Birthe B Kragelund, Lise Arleth, and Martin Cramer Pedersen (2023). "Modeling of flexible membrane-bound biomolecular complexes for solution small-angle scattering." In: *Journal of Colloid and Interface Science* 635, pp. 611–621.

Bayburt, Timothy H, Yelena V Grinkova, and Stephen G Sligar (2002). "Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins." In: *Nano letters* 2.8, pp. 853–856.

Berg, Jeremy M., John L. Tymoczko, and Lubert Stryer (2002). "Biochemistry." In: W.H. Freeman. Chap. 1 - The Molecular Design of Life.

Bernado, Pau and Dmitri I Svergun (2012). "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering." In: *Molecular biosystems* 8.1, pp. 151–167.

Blanchet, Clement E, Alessandro Spilotros, Frank Schwemmer, Melissa A Graewert, Alexey Kikhney, Cy M Jeffries, Daniel Franke, Daniel Mark, Roland Zengerle, Florent Cipriani, et al. (2015). "Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY)." In: *Journal of applied crystallography* 48.2, pp. 431–443.

Bockaert, Joel and Jean Philippe Pin (1999). "Molecular tinkering of G protein-coupled receptors: an evolutionary success." In: *The EMBO journal* 18.7, pp. 1723–1729.

Bottaro, Sandro, Tone Bengtsen, and Kresten Lindorff-Larsen (2020). "Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach." In: *Structural bioinformatics: methods and protocols*, pp. 219–240.

Brookes, Emre, Javier Pérez, Barbara Cardinali, Aldo Profumo, Patrice Vachette, and Mattia Rocco (2013). "Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler

(US-SOMO) enhanced SAS module." In: *Journal of applied crystallography* 46.6, pp. 1823–1833.

Brosey, Chris A and John A Tainer (2019). "Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology." In: *Current opinion in structural biology* 58, pp. 197–213.

Bryant, Patrick, Gabriele Pozzati, Wensi Zhu, Aditi Shenoy, Petras Kundrotas, and Arne Elofsson (2022). "Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search." In: *Nature communications* 13.1, p. 6028.

Butterfield, Sara M and Hilal A Lashuel (2010). "Amyloidogenic protein–membrane interactions: mechanistic insight from model systems." In: *Angewandte Chemie International Edition* 49.33, pp. 5628–5654.

Chatani, Eri and Naoki Yamamoto (2018). "Recent progress on understanding the mechanisms of amyloid nucleation." In: *Biophysical reviews* 10.2, pp. 527–534.

Childers, Kenneth C, Shaun C Peters, Pete Lollar, Harold Trent Spencer, Christopher B Doering, and Paul C Spiegel Jr (2022). "SAXS analysis of the intrinsic tenase complex bound to a lipid nanodisc highlights intermolecular contacts between factors VIIIa/IXa." In: *Blood Advances* 6.11, pp. 3240–3254.

Cho, Hyun Sun, Naranbaatar Dashdorj, Friedrich Schotte, Timothy Graber, Robert Henning, and Philip Anfinrud (2010). "Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering." In: *Proceedings of the National Academy of Sciences* 107.16, pp. 7281–7286.

Cho, Hyun Sun, Friedrich Schotte, Naranbaatar Dashdorj, John Kyndt, Robert Henning, and Philip A Anfinrud (2016). "Picosecond photobiology: watching a signaling protein function in real time via time-resolved small-and wide-angle X-ray scattering." In: *Journal of the American Chemical Society* 138.28, pp. 8815–8823.

Cho, Min-Kyu, Anindita Gayen, James R Banigan, Maureen Leninger, and Nathaniel J Traaseth (2014). "Intrinsic conformational plasticity of native EmrE provides a pathway for multidrug resistance." In: *Journal of the American Chemical Society* 136.22, pp. 8072–8080.

Cholak, Ersoy, Saskia Bucciarelli, Katrine Bugge, Nicolai Tidemand Johansen, Bente Vestergaard, Lise Arleth, Birthe B Kragelund, and Annette E Langkilde (2019). "Distinct $\alpha$-synuclein: lipid co-structure complexes affect amyloid nucleation through fibril mimetic behavior." In: *Biochemistry* 58.50, pp. 5052–5065.

Choy, Wing-Yiu, Frans AA Mulder, Karin A Crowhurst, DR Muhandiram, Ian S Millett, Sebastian Doniach, Julie D Forman-Kay, and Lewis E Kay (2002). "Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques." In: *Journal of molecular biology* 316.1, pp. 101–112.

Ciccariello, S, J Goodisman, and H Brumberger (1988). "On the Porod law." In: *Journal of applied crystallography* 21.2, pp. 117–128.

Claridge, Jolyon K, Chloe Martens, Brajabandhu Pradhan, Frank Sobott, Mike Sleutel, and Han Remaut (2023). "The folding-limited nucleation of curli hints at an evolved safety mechanism for functional amyloid production." In: *bioRxiv*, pp. 2023–05.

Crehuet, Ramon, Pedro J Buigues, Xavier Salvatella, and Kresten Lindorff-Larsen (2019). "Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts." In: *Entropy* 21.9, p. 898.

Danmaliki, Gaddafi I and Peter M Hwang (2020). "Solution NMR spectroscopy of membrane proteins." In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1862.9, p. 183356.

Dauter, Zbigniew and Alexander Wlodawer (2016). "Progress in protein crystallography." In: *Protein and peptide letters* 23.3, pp. 201–210.

David, G and J Pérez (2009). "Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline." In: *Journal of applied crystallography* 42.5, pp. 892–900.

Davidson, W Sean, Ana Jonas, David F Clayton, and Julia M George (1998). "Stabilization of $\alpha$-synuclein secondary structure upon binding to synthetic membranes." In: *Journal of Biological Chemistry* 273.16, pp. 9443–9449.

Davis, Rayhan, Christopher M Dobson, and Michele Vendruscolo (2002). "Determination of the structures of distinct transition state ensembles for a $\beta$-sheet peptide with parallel folding pathways." In: *The Journal of chemical physics* 117.20, pp. 9510–9517.

Debye, P (1915). "Scattering from non-crystalline substances." In: *Ann. Physik* 46, pp. 809–823.

– (1947). "Molecular-weight determination by light scattering." In: *The Journal of Physical Chemistry* 51.1, pp. 18–32.

Dedmon, Matthew M, Kresten Lindorff-Larsen, John Christodoulou, Michele Vendruscolo, and Christopher M Dobson (2005). "Mapping long-range interactions in $\alpha$-synuclein using spin-label NMR and ensemble molecular dynamics simulations." In: *Journal of the American Chemical Society* 127.2, pp. 476–477.

Denisov, Ilia G and Stephen G Sligar (2016). "Nanodiscs for structural and functional studies of membrane proteins." In: *Nature structural & molecular biology* 23.6, pp. 481–486.

Dobson, Christopher M (2003). "Protein folding and misfolding." In: *Nature* 426.6968, pp. 884–890.

Erskine, Elliot, Cait E MacPhee, and Nicola R Stanley-Wall (2018). "Functional amyloid and other protein fibers in the biofilm matrix." In: *Journal of Molecular Biology* 430.20, pp. 3642–3656.

Evans, Rachel, Sravani Ramisetty, Prakash Kulkarni, and Keith Weninger (2023). "Illuminating Intrinsically Disordered Proteins with Integrative Structural Biology." In: *Biomolecules* 13.1, p. 124.

Farrell, Daniel P, Ivan Anishchenko, Shabih Shakeel, Anna Lauko, Lori A Passmore, David Baker, and Frank DiMaio (2020). "Deep learning enables the atomic structure determination of the Fanconi Anemia core complex from cryoEM." In: *IUCrJ* 7.5, pp. 881–892.

Feigin, LA, Dmitri Iwanowitsch Svergun, et al. (1987). *Structure analysis by small-angle X-ray and neutron scattering.* Vol. 1. Springer.

Fersht, Alan R and Valerie Daggett (2002). "Protein folding and unfolding at atomic resolution." In: *Cell* 108.4, pp. 573–582.

Fields, Carroll Rutherford, Nora Bengoa-Vergniory, and Richard Wade-Martins (2019). "Targeting alpha-synuclein as a therapy for Parkinson's disease." In: *Frontiers in molecular neuroscience* 12, p. 299.

Fink, Anthony L (2006). "The aggregation and fibrillation of $\alpha$-synuclein." In: *Accounts of chemical research* 39.9, pp. 628–634.

Flory, Paul J (1953). *Principles of polymer chemistry.* Cornell university press.

Galvagnion, Céline, James WP Brown, Myriam M Ouberai, Patrick Flagmeier, Michele Vendruscolo, Alexander K Buell, Emma Sparr, and Christopher M Dobson (2016). "Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of $\alpha$-synuclein." In: *Proceedings of the National Academy of Sciences* 113.26, pp. 7065–7070.

Galvagnion, Céline, Alexander K Buell, Georg Meisl, Thomas CT Michaels, Michele Vendruscolo, Tuomas PJ Knowles, and Christopher M Dobson (2015). "Lipid vesicles trigger $\alpha$-synuclein aggregation by stimulating primary nucleation." In: *Nature chemical biology* 11.3, pp. 229–234.

Ghiglieri, Veronica, Valeria Calabrese, and Paolo Calabresi (2018). "Alpha-synuclein: From early synaptic dysfunction to neurodegeneration." In: *Frontiers in Neurology* 9, p. 295.

Glatter, Otto (1977). "A new method for the evaluation of small-angle scattering data." In: *Journal of Applied Crystallography* 10.5, pp. 415–421.

Gomes, Gregory-Neal W, Mickaël Krzeminski, Ashley Namini, Erik W Martin, Tanja Mittag, Teresa Head-Gordon, Julie D Forman-Kay, and Claudiu C Gradinaru (2020). "Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET." In: *Journal of the American Chemical Society* 142.37, pp. 15697–15710.

Gommes, Cedric J, Sebastian Jaksch, and Henrich Frielinghaus (2021). "Small-angle scattering for beginners." In: *Journal of applied crystallography* 54.6, pp. 1832–1843.

Goossens, Kenneth and Hans De Winter (2018). "Molecular dynamics simulations of membrane proteins: An overview." In: *Journal of chemical information and modeling* 58.11, pp. 2193–2202.

Gosal, Walraj S, Isobel J Morten, Eric W Hewitt, D Alastair Smith, Neil H Thomson, and Sheena E Radford (2005). "Competing pathways determine fibril morphology in the self-assembly of $\beta$2-microglobulin into amyloid." In: *Journal of molecular biology* 351.4, pp. 850–864.

Gremer, Lothar, Daniel Schölzel, Carla Schenk, Elke Reinartz, Jörg Labahn, Raimond BG Ravelli, Markus Tusche, Carmen Lopez-Iglesias, Wolfgang Hoyer, Henrike Heise, et al. (2017). "Fibril structure of amyloid-$\beta$ (1–42) by cryo–electron microscopy." In: *Science* 358.6359, pp. 116–119.

Grudinin, Sergei, Maria Garkavenko, and Andrei Kazennov (2017). "Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles." In: *Acta Crystallographica Section D: Structural Biology* 73.5, pp. 449–464.

Gsponer, Joerg, Harri Hopearuoho, Sara B-M Whittaker, Graham R Spence, Geoffrey R Moore, Emanuele Paci, Sheena E Radford, and Michele Vendruscolo (2006). "Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7." In: *Proceedings of the National Academy of Sciences* 103.1, pp. 99–104.

Gsponer, Jörg and Amedeo Caflisch (2002). "Molecular dynamics simulations of protein folding from the transition state." In: *Proceedings of the National Academy of Sciences* 99.10, pp. 6719–6724.

Guijarro, J Iñaki, Margaret Sunde, Jonathan A Jones, Iain D Campbell, and Christopher M Dobson (1998). "Amyloid fibril formation by an SH3 domain." In: *Proceedings of the National Academy of Sciences* 95.8, pp. 4224–4228.

Guillien, Myriam, Assia Mouhand, Aurelie Fournet, Amandine Gontier, Aleix Martí Navia, Tiago N Cordeiro, Frédéric Allemand, Aurélien Thureau, Jean-Louis Banères, Pau Bernadó, et al. (2022). "Structural insights into the intrinsically disordered GPCR c-terminal region, major actor in arrestin-GPCR interaction." In: *Biomolecules* 12.5, p. 617.

Guinier, André (1939). "La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques." In: *Annales de physique*. Vol. 11. 12. EDP Sciences, pp. 161–237.

Guinier, André, Gérard Fournet, Christopher B Walker, and Kenneth L Yudowitch (1955). *Small-angle Scattering of X-rays*. Wiley New York.

Hammouda, Boualem (1992). "Structure factor for starburst dendrimers." In: *Journal of Polymer Science Part B: Polymer Physics* 30.12, pp. 1387–1390.

Hansen, S (1990). "Calculation of small-angle scattering profiles using Monte Carlo simulation." In: *Journal of applied crystallography* 23.4, pp. 344–346.

Hayter, John B and Jeff Penfold (1981). "Self-consistent structural and dynamic study of concentrated micelle solutions." In: *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* 77.8, pp. 1851–1863.

Hegedűs, Tamás, Markus Geisler, Gergely László Lukács, and Bianka Farkas (2022). "Ins and outs of AlphaFold2 transmembrane protein structure predictions." In: *Cellular and Molecular Life Sciences* 79.1, p. 73.

Heller, Gabriella T, Francesco A Aprile, Massimiliano Bonomi, Carlo Camilloni, Alfonso De Simone, and Michele Vendruscolo (2017). "Sequence specificity in the entropy-driven binding of a small molecule and a disordered peptide." In: *Journal of molecular biology* 429.18, pp. 2772–2779.

Henriques, Joao, Carolina Cragnell, and Marie Skepo (2015). "Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment." In: *Journal of chemical theory and computation* 11.7, pp. 3420–3431.

Hollingsworth, Scott A and Ron O Dror (2018). "Molecular dynamics simulation for all." In: *Neuron* 99.6, pp. 1129–1143.

Hoover, Brandon M, Zhizhang Shen, Curran G Gahan, David M Lynn, Reid C Van Lehn, and Regina M Murphy (2021). "Membrane remodeling and stimulation of aggregation following $\alpha$-synuclein adsorption to phosphotidylserine vesicles." In: *The Journal of Physical Chemistry B* 125.6, pp. 1582–1594.

Hopkins, Jesse Bennett, Richard E Gillilan, and Soren Skou (2017). "BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis." In: *Journal of applied crystallography* 50.5, pp. 1545–1553.

Hsu, Darren J, Denis Leshchev, Irina Kosheleva, Kevin L Kohlstedt, and Lin X Chen (2021). "Unfolding bovine $\alpha$-lactalbumin with T-jump: Characterizing disordered intermediates via time-resolved x-ray solution scattering and molecular dynamics simulations." In: *The Journal of Chemical Physics* 154.10.

Hummer, Gerhard and Jürgen Köfinger (2015). "Bayesian ensemble refinement by replica simulations and reweighting." In: *The Journal of chemical physics* 143.24.

Iadanza, Matthew G, Robert Silvers, Joshua Boardman, Hugh I Smith, Theodoros K Karamanos, Galia T Debelouchina, Yongchao Su, Robert G Griffin, Neil A Ranson, and Sheena E Radford (2018). "The structure of a $\beta$2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism." In: *Nature Communications* 9.1, p. 4517.

Ingólfsson, Helgi I, Cesar A Lopez, Jaakko J Uusitalo, Djurre H de Jong, Srinivasa M Gopal, Xavier Periole, and Siewert J Marrink (2014). "The power of coarse graining in biomolecular simulations." In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.3, pp. 225–248.

Ivankov, Oleksandr, Tatiana N Murugova, Elena V Ermakova, Tomáš Kondela, Dina R Badreeva, Pavol Hrubovčák, Dmitry Soloviov, Alexey Tsarenko, Andrey Rogachev, Alexander I Kuklin, et al. (2021). "Amyloid-beta peptide (25–35) triggers a reorganization of lipid membranes driven by temperature changes." In: *Scientific Reports* 11.1, p. 21990.

Jaeschke, Eberhard J, Shaukat Khan, Jochen R Schneider, and Jerome B Hastings (2020). *Synchrotron light sources and free-electron lasers: accelerator physics, instrumentation and science applications.*

Jaksch, Sebastian, Alexis Chennevière, Sylvain Désert, Tadeusz Kozielewski, Herbert Feilbach, Pascal Lavie, Romuald Hanslik, Achim Gussen, Stephan Butterweck, Ralf Engels, et al. (2021). "Technical specification of the small-angle neutron scattering instrument SKADI at the European Spallation Source." In: *Applied Sciences* 11.8, p. 3620.

Jensen, Grethe Vestergaard, Reidar Lund, Jeremie Gummel, Michael Monkenbusch, Theyencheri Narayanan, and Jan Skov Pedersen (2013). "Direct observation of the formation of surfactant micelles under nonisothermal conditions by synchrotron SAXS." In: *Journal of the American Chemical Society* 135.19, pp. 7214–7222.

Jensen, Malene Ringkjøbing, Rob WH Ruigrok, and Martin Blackledge (2013). "Describing intrinsically disordered proteins at atomic resolution by NMR." In: *Current opinion in structural biology* 23.3, pp. 426–435.

Jiang, Zhiping, Jessica D Flynn, Walter E Teague Jr, Klaus Gawrisch, and Jennifer C Lee (2018). "Stimulation of $\alpha$-synuclein amyloid formation by phosphatidylglycerol micellar tubules." In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1860.9, pp. 1840–1847.

Johansen, Nicolai Tidemand, Marta Bonaccorsi, Tone Bengtsen, Andreas Haahr Larsen, Frederik Grønbæk Tidemand, Martin Cramer Pedersen, Pie Huda, Jens Berndtsson, Tamim Darwish, Nageshewar Rao Yepuri, et al. (2022). "Mg2+-dependent conformational equilibria in CorA and an integrated view on transport regulation." In: *Elife* 11, e71887.

Johansen, Nicolai Tidemand, Martin Cramer Pedersen, Lionel Porcar, Anne Martel, and Lise Arleth (2018). "Introducing SEC–SANS for studies of complex self-organized biological systems." In: *Acta Crystallographica Section D: Structural Biology* 74.12, pp. 1178–1191.

Johansen, Nicolai Tidemand, Frederik Grønbæk Tidemand, Tam TTN Nguyen, Kasper Dyrberg Rand, Martin Cramer Pedersen, and Lise Arleth (2019). "Circularized and solubility-enhanced MSP s facilitate simple and high-yield production of stable nanodiscs for studies of membrane proteins in solution." In: *The FEBS journal* 286.9, pp. 1734–1751.

Johansen, Nicolai Tidemand, Frederik Grønbæk Tidemand, Martin Cramer Pedersen, and Lise Arleth (2023). "Travel light: Essential packing for membrane proteins with an active lifestyle." In: *Biochimie* 205, pp. 3–26.

Jordan, Ashley, Mark Jacques, Catherine Merrick, Juliette Devos, V Trevor Forsyth, Lionel Porcar, and Anne Martel (2016). "SEC-SANS: size exclusion chromatography combined in situ with small-angle neutron scattering." In: *Journal of applied crystallography* 49.6, pp. 2015–2020.

Josts, Inokentijs, Julius Nitsche, Selma Maric, Haydyn D Mertens, Martine Moulin, Michael Haertlein, Sylvain Prevost, Dmitri I Svergun, Sebastian Busch, V Trevor Forsyth, et al. (2018). "Conformational states of ABC transporter MsbA in a lipid environment investigated by small-angle scattering using stealth carrier nanodiscs." In: *Structure* 26.8, pp. 1072–1079.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. (2021). "Highly accurate protein structure prediction with AlphaFold." In: *Nature* 596.7873, pp. 583–589.

Kassem, Noah, Raul Araya-Secchi, Katrine Bugge, Abigail Barclay, Helena Steinocher, Adree Khondker, Yong Wang, Aneta J Lenard, Jochen Bürck, Cagla Sahin, et al. (2021). "Order and disorder—An integrative structure of the full-length human growth hormone receptor." In: *Science Advances* 7.27, eabh3805.

Kikhney, Alexey G and Dmitri I Svergun (2015). "A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins." In: *FEBS letters* 589.19, pp. 2570–2577.

Kim, Seung Joong, Javier Fernandez-Martinez, Ilona Nudelman, Yi Shi, Wenzhu Zhang, Barak Raveh, Thurston Herricks, Brian D Slaughter, Joanna A Hogan, Paula Upla, et al. (2018). "Integrative structure and functional anatomy of a nuclear pore complex." In: *Nature* 555.7697, pp. 475–482.

Kim, So-Jung, Young Hoon Koh, and Soung-Hun Roh (2023). "Unraveling membrane protein localization and stabilization in nanodiscs." In: *bioRxiv*, pp. 2023–07.

Kohn, Jonathan E, Ian S Millett, Jaby Jacob, Bojan Zagrovic, Thomas M Dillon, Nikolina Cingel, Robin S Dothager, Soenke Seifert, P Thiyagarajan, Tobin R Sosnick, et al. (2004). "Random-coil behavior and the dimensions of chemically unfolded proteins." In: *Proceedings of the National Academy of Sciences* 101.34, pp. 12491–12496.

Kučerka, Norbert, Mu-Ping Nieh, and John Katsaras (2011). "Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature." In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808.11, pp. 2761–2771.

Kummerer, Felix, Simone Orioli, David Harding-Larsen, Falk Hoffmann, Yulian Gavrilov, Kaare Teilum, and Kresten Lindorff-Larsen (2021). "Fitting side-chain NMR relaxation data using molecular simulations." In: *Journal of Chemical Theory and Computation* 17.8, pp. 5262–5275.

Kynde, Søren AR, Nicholas Skar-Gislinge, Martin Cramer Pedersen, Søren Roi Midgaard, Jens Baek Simonsen, Ralf Schweins, Kell Mortensen, and Lise Arleth (2014). "Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment." In: *Acta Crystallographica Section D: Biological Crystallography* 70.2, pp. 371–383.

La Verde, Valentina, Paola Dominici, and Alessandra Astegno (2017). "Determination of hydrodynamic radius of proteins by size exclusion chromatography." In: *Bio-protocol* 7.8, e2230–e2230.

Larsen, Andreas Haahr, Lise Arleth, and Steen Hansen (2018). "Analysis of small-angle scattering data using model fitting and Bayesian regularization." In: *Journal of Applied Crystallography* 51.4, pp. 1151–1161.

Larsen, Andreas Haahr, Emre Brookes, Martin Cramer Pedersen, and Jacob Judas Kain Kirkensgaard (2023). "Shape2SAS–a web application to simulate small-angle scattering data and pair distance distributions from user-defined shapes." In: *arXiv preprint arXiv:2301.04976*.

Larsen, Andreas Haahr and Martin Cramer Pedersen (2021). "Experimental noise in small-angle scattering can be assessed using the Bayesian indirect Fourier transformation." In: *Journal of Applied Crystallography* 54.5, pp. 1281–1289.

Larsen, Andreas Haahr, Yong Wang, Sandro Bottaro, Sergei Grudinin, Lise Arleth, and Kresten Lindorff-Larsen (2020). "Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution." In: *PLoS computational biology* 16.4, e1007870.

Leach, Andrew R. (2001). "Molecular modelling Principles and Applications." In: Pearson Education Limited. Chap. 4 - Empirical Force Field Models: Molecular Mechanics.

Lerner, Eitan, Anders Barth, Jelle Hendrix, Benjamin Ambrose, Victoria Birkedal, Scott C Blanchard, Richard Börner, Hoi Sung Chung, Thorben Cordes, Timothy D Craggs, et al. (2021). "FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices." In: *Elife* 10, e60416.

Linderstrøm-Lang, Kaj Ulrik (1952). *Lane medical lectures: proteins and enzymes*. Vol. 6. Stanford University Press.

Lindorff-Larsen, Kresten, Emanuele Paci, Luis Serrano, Christopher M Dobson, and Michele Vendruscolo (2003). "Calculation of mutational free energy changes in transition states for protein folding." In: *Biophysical journal* 85.2, pp. 1207–1214.

Long, G. J (2003). "Comprehensive Coordination Chemistry II." In: Pergamon. Chap. 2.6 Neutron Diffraction.

López, Cesar A, Mark F Swift, Xiao-Ping Xu, Dorit Hanein, Niels Volkmann, and S Gnanakaran (2019). "Biophysical characterization of a nanodisc with and without BAX: an integrative study using molecular dynamics simulations and cryo-EM." In: *Structure* 27.6, pp. 988–999.

Majeed, Shereen A, Kutloano Edward Sekhosana, and Ahmad Tuhl (2020). "Progress on phthalocyanine-conjugated Ag and Au nanoparticles: Synthesis, characterization, and photo-physicochemical properties." In: *Arabian Journal of Chemistry* 13.12, pp. 8848–8887.

Manet, Sabine, Amélie Lecchi, Marianne Impéror-Clerc, Vladimir Zholobenko, Dominique Durand, Cristiano LP Oliveira, Jan Skov Pedersen, Isabelle

Grillo, Florian Meneau, and Cyrille Rochas (2011). "Structure of micelles of a nonionic block copolymer determined by SANS and SAXS." In: *The Journal of Physical Chemistry B* 115.39, pp. 11318–11329.

Maric, Selma, Nicholas Skar-Gislinge, Søren Midtgaard, Mikkel B Thygesen, Jurgen Schiller, Henrich Frielinghaus, Martine Moulin, Michael Haertlein, V Trevor Forsyth, Thomas Gunther Pomorski, et al. (2014). "Stealth carriers for low-resolution structure determination of membrane proteins in solution." In: *Acta Crystallographica Section D: Biological Crystallography* 70.2, pp. 317–328.

Marion, Jérémie, Marie Trovaslet, Nicolas Martinez, Patrick Masson, R Schweins, Florian Nachon, M Trapp, and Judith Peters (2015). "Pressure-induced molten globule state of human acetylcholinesterase: structural and dynamical changes monitored by neutron scattering." In: *Physical Chemistry Chemical Physics* 17.5, pp. 3157–3163.

Martel, Anne, Cristina Cocho, Francesca Caporaletti, Mark Jacques, Abdelali El Aazzouzi, Franck Lapeyre, and Lionel Porcar (2023). "Upgraded D22 SEC–SANS setup dedicated to the biology community." In: *Journal of Applied Crystallography* 56.4.

Martin, Erik W, Tyler S Harmon, Jesse B Hopkins, Srinivas Chakravarthy, J Jeremías Incicco, Peter Schuck, Andrea Soranno, and Tanja Mittag (2021). "A multi-step nucleation process determines the kinetics of prion-like domain phase separation." In: *Nature communications* 12.1, p. 4513.

Mathew, Elizabeth, Ahmed Mirza, and Nick Menhart (2004). "Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins." In: *Journal of synchrotron radiation* 11.4, pp. 314–318.

Matouschek, Andreas, James T Kellis Jr, Luis Serrano, and Alan R Fersht (1989). "Mapping the transition state and pathway of protein folding by protein engineering." In: *Nature* 340.6229, pp. 122–126.

McCammon, J Andrew, Bruce R Gelin, and Martin Karplus (1977). "Dynamics of folded proteins." In: *nature* 267.5612, pp. 585–590.

Mertens, Haydyn DT and Dmitri I Svergun (2017). "Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics." In: *Archives of biochemistry and biophysics* 628, pp. 33–41.

Metwalli, Ezzeldin, Klaus Götz, Tobias Zech, Christian Bär, Isabel Schuldes, Anne Martel, Lionel Porcar, and Tobias Unruh (2021). "Simultaneous SAXS/SANS method at D22 of ILL: Instrument upgrade." In: *Applied Sciences* 11.13, p. 5925.

Middleton, Elizabeth R and Elizabeth Rhoades (2010). "Effects of curvature and composition on $\alpha$-synuclein binding to lipid vesicles." In: *Biophysical journal* 99.7, pp. 2279–2288.

Midtgaard, Søren Roi, Tamim A Darwish, Martin Cramer Pedersen, Pie Huda, Andreas Haahr Larsen, Grethe Vestergaard Jensen, Søren Andreas Røssell Kynde, Nicholas Skar-Gislinge, Agnieszka Janina Zygadlo Nielsen, Claus Olesen, et al. (2018). "Invisible detergents for

structure determination of membrane proteins by small-angle neutron scattering." In: *The FEBS journal* 285.2, pp. 357–371.

Mizuno, Naoko, Jobin Varkey, Natalie C Kegulian, Balachandra G Hegde, Naiqian Cheng, Ralf Langen, and Alasdair C Steven (2012). "Remodeling of lipid vesicles into cylindrical micelles by α-synuclein in an extended α-helical conformation." In: *Journal of Biological Chemistry* 287.35, pp. 29301–29311.

Molodenskiy, Dmitry, Evgeny Shirshin, Tatiana Tikhonova, Andrey Gruzinov, Georgy Peters, and Francesco Spinozzi (2017). "Thermally induced conformational changes and protein–protein interactions of bovine serum albumin in aqueous solution under different pH and ionic strengths as revealed by SAXS measurements." In: *Physical Chemistry Chemical Physics* 19.26, pp. 17143–17155.

Montepietra, Daniele, Giulio Tesei, João M Martins, Micha BA Kunze, Robert Best, and Kresten Lindorff-Larsen (2023). "FRETpredict: A Python package for FRET efficiency predictions using rotamer libraries." In: *bioRxiv*, pp. 2023–01.

Morrison, Kerrie A, Aswin Doekhie, George M Neville, Gareth J Price, Paul Whitley, James Doutch, and Karen J Edler (2022). "Ab initio reconstruction of small angle scattering data for membrane proteins in copolymer nanodiscs." In: *BBA advances* 2, p. 100033.

Mosalaganti, Shyamal, Agnieszka Obarska-Kosinska, Marc Siggel, Reiya Taniguchi, Beata Turoňová, Christian E Zimmerli, Katarzyna Buczak, Florian H Schmidt, Erica Margiotta, Marie-Therese Mackmull, et al. (2022). "AI-based structure prediction empowers integrative structural analysis of human nuclear pores." In: *Science* 376.6598, eabm9506.

Mylonas, Efstratios and Dmitri I Svergun (2007). "Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering." In: *Journal of Applied Crystallography* 40.SUPPL. 1, pp. 245–249. DOI: 10.1107/S002188980700252X.

Nagy, Kornél and Károly Vékey (2008). "Separation methods." In: *Medical Applications of Mass Spectrometry*. Elsevier, pp. 61–92.

Nasr, Mahmoud L, Diego Baptista, Mike Strauss, Zhen-Yu J Sun, Simina Grigoriu, Sonja Huser, Andreas Plückthun, Franz Hagn, Thomas Walz, James M Hogle, et al. (2017). "Covalently circularized nanodiscs for studying membrane proteins and viral entry." In: *Nature methods* 14.1, pp. 49–52.

Neumann, Manuela, Veronika Müller, Hans A Kretzschmar, Christian Haass, and Philipp J Kahle (2004). "Regional distribution of proteinase K-resistant α-synuclein correlates with Lewy body disease stage." In: *Journal of Neuropathology & Experimental Neurology* 63.12, pp. 1225–1235.

Opie, Eugene L (1901). "On the relation of chronic interstitial pancreatitis to the islands of Langerhans and to diabetes melutus." In: *The Journal of experimental medicine* 5.4, p. 397.

Orioli, Simone, Carl G Henning Hansen, and Lise Arleth (2021). "Ab initio determination of the shape of membrane proteins in a nanodisc." In: *Acta Crystallographica Section D: Structural Biology* 77.2, pp. 176–193.

Orioli, Simone, Andreas Haahr Larsen, Sandro Bottaro, and Kresten Lindorff-Larsen (2020). "How to learn from inconsistencies: Integrating molecular simulations with experimental data." In: *Progress in Molecular Biology and Translational Science* 170, pp. 123–176.

Ortore, Maria Grazia, Francesco Spinozzi, Silvia Vilasi, Ivana Sirangelo, Gaetano Irace, Anuji Shukla, Theyencheri Narayanan, Raffaele Sinibaldi, and Paolo Mariani (2011). "Time-resolved small-angle X-ray scattering study of the early formation of amyloid protofibrils on a apomyoglobin mutant." In: *Biophysical Journal* 100.3, 532a.

Ozenne, Valéry, Frédéric Bauer, Loïc Salmon, Jie Rong Huang, Malene Ringkjøbing Jensen, Stéphane Segard, Pau Bernadó, Céline Charavay, and Martin Blackledge (2012). "Flexible-meccano: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables." In: *Bioinformatics* 28.11, pp. 1463–1470. DOI: 10.1093/bioinformatics/bts172.

Paissoni, Cristina, Sarita Puri, Iren Wang, Szu-Yu Chen, Carlo Camilloni, and Shang-Te Danny Hsu (2021). "Converging experimental and computational views of the knotting mechanism of a small knotted protein." In: *Biophysical Journal* 120.11, pp. 2276–2286.

Panjkovich, Alejandro and Dmitri I Svergun (2018). "CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data." In: *Bioinformatics* 34.11, pp. 1944–1946.

Pedersen, Jan Skov (1997). "Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting." In: *Advances in colloid and interface science* 70, pp. 171–210.

– (2000). "Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores." In: *Journal of Applied Crystallography* 33.3, pp. 637–640.

Pedersen, Jan Skov and Michael C Gerstenberg (1996). "Scattering form factor of block copolymer micelles." In: *Macromolecules* 29.4, pp. 1363–1365.

Pedersen, Jan Skov, Cristiano LP Oliveira, Henriette Baun Hübschmann, Lise Arleth, Søren Manniche, Nicolai Kirkby, and Hanne Mørck Nielsen (2012). "Structure of immune stimulating complex matrices and immune stimulating complexes in suspension determined by small-angle X-ray scattering." In: *Biophysical Journal* 102.10, pp. 2372–2380.

Pedersen, Martin Cramer, Nicolai Tidemand Johansen, Jennifer Roche, Michael Järvä, Susanna Törnroth-Horsefield, and Lise Arleth (2022). "Refining structural models of membrane proteins with disordered domains in phospholipid nanodiscs." In: *bioRxiv*. DOI: 10.1101/2022.10.28.512841.

Peelaerts, Wouter, Luc Bousset, Anke Van der Perren, Anastacia Moskalyuk, Rocco Pulizzi, Michele Giugliano, Chris Van den Haute, Ronald Melki, and Veerle Baekelandt (2015). "$\alpha$-Synuclein strains cause distinct synucleinopathies after local and systemic administration." In: *Nature* 522.7556, pp. 340–344.

Pérez, Javier, Aurélien Thureau, and Patrice Vachette (2022). "SEC-SAXS: Experimental set-up and software developments build up a powerful tool." In: *Methods in Enzymology*. Vol. 677. Elsevier, pp. 221–249.

Perilla, Juan R, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten (2015). "Molecular dynamics simulations of large macromolecular complexes." In: *Current opinion in structural biology* 31, pp. 64–74.

Pernot, Petra, Adam Round, Ray Barrett, Alejandro De Maria Antolinos, Alexandre Gobbo, Elspeth Gordon, Julien Huet, Jerôme Kieffer, Mario Lentini, Muriel Mattenet, et al. (2013). "Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution." In: *Journal of synchrotron radiation* 20.4, pp. 660–664.

Pesce, Francesco and Kresten Lindorff-Larsen (2021). "Refining conformational ensembles of flexible proteins against small-angle x-ray scattering data." In: *Biophysical journal* 120.22, pp. 5124–5135.

Petrovic, Stefan, Dipanjan Samanta, Thibaud Perriches, Christopher J Bley, Karsten Thierbach, Bonnie Brown, Si Nie, George W Mobbs, Taylor A Stevens, Xiaoyu Liu, et al. (2022). "Architecture of the linker-scaffold in the nuclear pore." In: *Science* 376.6598, eabm9798.

Pham, Nhat, Dimitri Radajewski, Adam Round, Martha Brennich, Petra Pernot, Béatrice Biscans, Françoise Bonneté, and Sébastien Teychené (2017). "Coupling high throughput microfluidics and small-angle X-ray scattering to study protein crystallization from solution." In: *Analytical chemistry* 89.4, pp. 2282–2287.

Porod, G (1948). "Die abhangigkeitder rontgen-kleinwinkelstreuung von form und gro$\beta$e der kolloiden teilchen in verdunnten systemen, iv." In: *Acta phys. austriaca* 2, pp. 255–292.

Prestel, Andreas, Katrine Bugge, Lasse Staby, Ruth Hendus-Altenburger, and Birthe B Kragelund (2018). "Characterization of dynamic IDP complexes by NMR spectroscopy." In: *Methods in enzymology*. Vol. 611. Elsevier, pp. 193–226.

Qi, Yifei, Jumin Lee, Jeffery B Klauda, and Wonpil Im (2019). "CHARMM-GUI nanodisc builder for modeling and simulation of various nanodisc systems." In: *Journal of Computational Chemistry* 40.7, pp. 893–899.

Radamaker, Lynn, Julian Baur, Stefanie Huhn, Christian Haupt, Ute Hegenbart, Stefan Schönland, Akanksha Bansal, Matthias Schmidt, and Marcus Fändrich (2021). "Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis." In: *Nature Communications* 12.1, p. 875.

Radamaker, Lynn, Yin-Hsi Lin, Karthikeyan Annamalai, Stefanie Huhn, Ute Hegenbart, Stefan O Schönland, Günter Fritz, Matthias Schmidt,

and Marcus Fändrich (2019). "Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis." In: *Nature communications* 10.1, p. 1103.

Rambo, Robert P and John A Tainer (2013). "Accurate assessment of mass, models and resolution by small-angle scattering." In: *Nature* 496.7446, pp. 477–481.

Ray, Soumik, Thomas O Mason, Lars Boyens-Thiele, Azad Farzadfard, Jacob Aunstrup Larsen, Rasmus K Norrild, Nadin Jahnke, and Alexander K Buell (2023). "Mass photometric detection and quantification of nanoscale α-synuclein phase separation." In: *Nature Chemistry*, pp. 1–11.

Rimmerman, Dolev, Denis Leshchev, Darren J Hsu, Jiyun Hong, Irina Kosheleva, and Lin X Chen (2017). "Direct observation of insulin association dynamics with time-resolved X-ray scattering." In: *The journal of physical chemistry letters* 8.18, pp. 4413–4418.

Robustelli, Paul, Stefano Piana, and David E Shaw (2018). "Developing a molecular dynamics force field for both folded and disordered protein states." In: *Proceedings of the National Academy of Sciences* 115.21, E4758–E4766.

Roccatano, Danilo, Andre Barthel, and Martin Zacharias (2007). "Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation." In: *Biopolymers: Original Research on Biomolecules* 85.5-6, pp. 407–421.

Röder, Christine, Nicola Vettore, Lena N Mangels, Lothar Gremer, Raimond BG Ravelli, Dieter Willbold, Wolfgang Hoyer, Alexander K Buell, and Gunnar F Schröder (2019). "Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy." In: *Nature Communications* 10.1, p. 3754.

Rotkiewicz, P and J Skolnick (2008). "Fast procedure for reconstruction of full-atom protein models from reduced representations." In: 29.9, pp. 1460–1465.

Ruff, Kiersten M and Rohit V Pappu (2021). "AlphaFold and implications for intrinsically disordered proteins." In: *Journal of Molecular Biology* 433.20, p. 167208.

Sakuragi, Mina, Kazunori Koiwai, Kouji Nakamura, Hiroyasu Masunaga, Hiroki Ogawa, and Kazuo Sakurai (2011). "Transformation from multilamellar to unilamellar vesicles by addition of a cationic lipid to pegylated liposomes explored with synchrotron small angle X-ray scattering." In: *Journal of physics: conference series*. Vol. 272. 1. IOP Publishing, p. 012011.

Sato, Satoshi, Tomasz L Religa, Valerie Daggett, and Alan R Fersht (2004). "Testing protein-folding simulations by experiment: B domain of protein A." In: *Proceedings of the National Academy of Sciences* 101.18, pp. 6952–6956.

Sauter, Andrea, Fajun Zhang, Noemi K Szekely, Vitaliy Pipich, Michael Sztucki, and Frank Schreiber (2016). "Structural evolution of metastable

protein aggregates in the presence of trivalent salt studied by (V) SANS and SAXS." In: *The Journal of Physical Chemistry B* 120.24, pp. 5564–5571.

Schmid, Andreas Josef, Lars Wiehemeier, Sebastian Jaksch, Harald Schneider, Arno Hiess, Torsten Bögershausen, Tobias Widmann, Julija Reitenbach, Lucas P Kreuzer, Matthias Kühnhammer, et al. (2021). "Flexible sample environments for the investigation of soft matter at the European Spallation Source: Part I—The in situ SANS/DLS setup." In: *Applied Sciences* 11.9, p. 4089.

Schnablegger, H and Y Singh (2013). *The SAXS Guide: Getting acquainted with the principles*. Vol. 3. Anton Paar GmbH.

Schuller, Marion, Galen J Correy, Stefan Gahbauer, Daren Fearon, Taiasean Wu, Roberto Efraín Díaz, Iris D Young, Luan Carvalho Martins, Dominique H Smith, Ursula Schulze-Gahmen, et al. (2021). "Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking." In: *Science advances* 7.16, eabf8711.

Schymkowitz, Joost, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano (2005). "The FoldX web server: an online force field." In: *Nucleic acids research* 33.suppl_2, W382–W388.

Sears, Varley F (1992). "Neutron scattering lengths and cross sections." In: *Neutron news* 3.3, pp. 26–37.

Sebastiani, Federica, Marianna Yanez Arteta, Michael Lerche, Lionel Porcar, Christian Lang, Ryan A Bragg, Charles S Elmore, Venkata R Krishnamurthy, Robert A Russell, Tamim Darwish, et al. (2021). "Apolipoprotein E binding drives structural and compositional rearrangement of mRNA-containing lipid nanoparticles." In: *ACS nano* 15.4, pp. 6709–6722.

Singh, Divya (2009). *Small angle scatting studies of self assembly in lipid mixtures*. The Johns Hopkins University.

Skar-Gislinge, N. (2014). "Developing Nanodiscs as a Tool for Low Resolution Studies of Membrane Proteins." PhD thesis. University of Copenhagen.

Skar-Gislinge, Nicholas and Lise Arleth (2011). "Small-angle scattering from phospholipid nanodiscs: derivation and refinement of a molecular constrained analytical model form factor." In: *Physical Chemistry Chemical Physics* 13.8, pp. 3161–3170.

Skar-Gislinge, Nicholas, Søren AR Kynde, Ilia G Denisov, Xin Ye, Ivan Lenov, Stephen G Sligar, and Lise Arleth (2015). "Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment." In: *Acta Crystallographica Section D: Biological Crystallography* 71.12, pp. 2412–2421.

Skar-Gislinge, Nicholas, Jens Bæk Simonsen, Kell Mortensen, Robert Feidenhans'l, Stephen G Sligar, Birger Lindberg Møller, Thomas Bjørnholm, and Lise Arleth (2010). "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, pp. 13713–13722.

Sligar, Stephen G and Ilia G Denisov (2021). "Nanodiscs: A toolkit for membrane protein science." In: *Protein Science* 30.2, pp. 297–315.

Soto, Claudio and Lisbell D Estrada (2008). "Protein misfolding and neurodegeneration." In: *Archives of neurology* 65.2, pp. 184–189.

Spillantini, Maria Grazia and Michel Goedert (2000). "The α-synucleinopathies: Parkinson's disease, dementia with Lewy bodies, and multiple system atrophy." In: *Annals of the New York Academy of Sciences* 920.1, pp. 16–27.

Stella, Stefano, Pablo Mesa, Johannes Thomsen, Bijoya Paul, Pablo Alcón, Simon B Jensen, Bhargav Saligram, Matias E Moses, Nikos S Hatzakis, and Guillermo Montoya (2018). "Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity." In: *Cell* 175.7, pp. 1856–1871.

Svergun, DIBC, Claudio Barberato, and Michel HJ Koch (1995). "CRYSOL– a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates." In: *Journal of applied crystallography* 28.6, pp. 768–773.

Svergun, Dmitri I. and Michel H. J. Koch (2003). "Small-angle scattering studies of biological macromolecules in solution." In: *Reports on Progress in Physics* 66, pp. 1735–1782.

Sweeney, Patrick, Hyunsun Park, Marc Baumann, John Dunlop, Judith Frydman, Ron Kopito, Alexander McCampbell, Gabrielle Leblanc, Anjli Venkateswaran, Antti Nurmi, et al. (2017). "Protein misfolding in neurodegenerative diseases: implications and strategies." In: *Translational neurodegeneration* 6, pp. 1–13.

Tang, Henry YH, John A Tainer, and Greg L Hura (2017). "High resolution distance distributions determined by X-ray and neutron scattering." In: *Biological Small Angle Scattering: Techniques, Strategies and Tips*, pp. 167–181.

Thomasen, F Emil, Matthew J Cuneo, Tanja Mittag, and Kresten Lindorff-Larsen (2023). "Conformational and oligomeric states of SPOP from small-angle X-ray scattering and molecular dynamics simulations." In: *Elife* 12, e84147.

Thomasen, F Emil and Kresten Lindorff-Larsen (2022). "Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins." In: *Biochemical Society Transactions* 50.1, pp. 541–554.

Tunyasuvunakool, Kathryn, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, et al. (2021). "Highly accurate protein structure prediction for the human proteome." In: *Nature* 596.7873, pp. 590–596.

Uversky, Vladimir N (2019). "Intrinsically disordered proteins and their "mysterious"(meta) physics." In: *Frontiers in Physics* 7, p. 10.

Uversky, Vladimir N, Joel R Gillespie, and Anthony L Fink (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" In: *Proteins: structure, function, and bioinformatics* 41.3, pp. 415–427.

Valentini, Erica, Alexey G Kikhney, Gianpietro Previtali, Cy M Jeffries, and Dmitri I Svergun (2015). "SASBDB, a repository for biological small-angle scattering data." In: *Nucleic acids research* 43.D1, pp. D357–D363.

Van Gerven, Nani, Sander E Van der Verren, Dirk M Reiter, and Han Remaut (2018). "The role of functional amyloids in bacterial virulence." In: *Journal of molecular biology* 430.20, pp. 3657–3684.

Varkey, Jobin, Jose Mario Isas, Naoko Mizuno, Martin Borch Jensen, Vikram Kjøller Bhatia, Christine C Jao, Jitka Petrlova, John C Voss, Dimitrios G Stamou, Alasdair C Steven, et al. (2010). "Membrane curvature induction and tubulation are common features of synucleins and apolipoproteins." In: *Journal of Biological Chemistry* 285.42, pp. 32486–32493.

Varkey, Jobin, Naoko Mizuno, Balachandra G Hegde, Naiqian Cheng, Alasdair C Steven, and Ralf Langen (2013). "$\alpha$-Synuclein oligomers with broken helical conformation form lipoprotein nanoparticles." In: *Journal of Biological Chemistry* 288.24, pp. 17620–17630.

Vendruscolo, Michele, Emanuele Paci, Christopher M Dobson, and Martin Karplus (2001). "Three key residues form a critical contact network in a protein folding transition state." In: *Nature* 409.6820, pp. 641–645.

Vettore, Nicola and Alexander K Buell (2019). "Thermodynamics of amyloid fibril formation from chemical depolymerization." In: *Physical Chemistry Chemical Physics* 21.47, pp. 26184–26194.

Wallace, Bonnie A (2010). "Protein characterisation by synchrotron radiation circular dichroism spectroscopy." In: *Quarterly reviews of biophysics* 42.4, pp. 317–370.

Wang, Yong, Katrine Bugge, Birthe B Kragelund, and Kresten Lindorff-Larsen (2018). "Role of protein dynamics in transmembrane receptor signalling." In: *Current opinion in structural biology* 48, pp. 74–82.

Ward, Andrew B, Andrej Sali, and Ian A Wilson (2013). "Integrative structural biology." In: *Science* 339.6122, pp. 913–915.

Watanabe, Yasushi and Yoji Inoko (2009). "Size-exclusion chromatography combined with small-angle X-ray scattering optics." In: *Journal of Chromatography A* 1216.44, pp. 7461–7465.

Weber, Benedikt, Manuel Hora, Pamina Kazman, Christoph Göbl, Carlo Camilloni, Bernd Reif, and Johannes Buchner (2018). "The antibody light-chain linker regulates domain orientation and amyloidogenicity." In: *Journal of molecular biology* 430.24, pp. 4925–4940.

Westermark, Per, Arne Andersson, and Gunilla T Westermark (2011). "Islet amyloid polypeptide, islet amyloid, and diabetes mellitus." In: *Physiological reviews* 91.3, pp. 795–826.

Whitford, Paul C, Aqeel Ahmed, Yanan Yu, Scott P Hennelly, Florence Tama, Christian MT Spahn, José N Onuchic, and Karissa Y Sanbonmatsu (2011). "Excited states of ribosome translocation revealed through integrative molecular modeling." In: *Proceedings of the National Academy of Sciences* 108.47, pp. 18943–18948.

Wright, Peter E and H Jane Dyson (2015). "Intrinsically disordered proteins in cellular signalling and regulation." In: *Nature reviews Molecular cell biology* 16.1, pp. 18–29.

Xie, Jingru and Aaron T Frank (2021). "RNA Ensembles from Solvent Accessibility Data: Application to the SAM-I Riboswitch Aptamer Domain." In: *The Journal of Physical Chemistry B* 125.14, pp. 3486–3493.

Yan, Yumeng, Huanyu Tao, Jiahua He, and Sheng-You Huang (2020). "The HDOCK server for integrated protein–protein docking." In: *Nature protocols* 15.5, pp. 1829–1852.

Zhou, Yingli, Baoshan Wang, and Fang Yuan (2022). "The role of transmembrane proteins in plant growth, development, and stress responses." In: *International Journal of Molecular Sciences* 23.21, p. 13627.

Zwier, Matthew C and Lillian T Chong (2010). "Reaching biological timescales with all-atom molecular dynamics simulations." In: *Current opinion in pharmacology* 10.6, pp. 745–752.

Part II

RESEARCH ARTICLES AND MANUSCRIPTS

# 7

# GLOBAL FITTING OF MULITPLE FRAMES FROM SEC-SAXS TO INVESTIGATE THE STRUCTURE OF NEXT-GENERATION NANODISCS

CONTRIBUTIONS

The research in this article was driven by two primary interests. Firstly, out of interest in the structural parameters of nanodiscs with supercharged and solubility-enhanced MSPs, which were newly developed in the Structural Biophysics group at NBI. Secondly, we aimed to explore the potential of SEC-SAXS as a tool to investigate the inherent polydispersity within a sample.

I performed the data analysis and model fits. I utilised the analytical model for nanodiscs and extended the pre-existing software from the group, *WillItFit*, to build the simultaneous fitting platform. All of the authors wrote the article together. Nicolai Tidemand Johansen and Frederik Grønbaek Tidemand prepared the nanodisc samples and collected the SAXS data. Lise Arleth and Martin Cramer Pedersen supervised the project.

# Global fitting of multiple data frames from SEC–SAXS to investigate the structure of next-generation nanodiscs

**Abigail Barclay,**[a] **Nicolai Tidemand Johansen,**[a,b] **Frederik Grønbæk Tidemand,**[a] **Lise Arleth**[a] **and Martin Cramer Pedersen**[a]*

[a]Niels Bohr Institute, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen E, Denmark, and [b]Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg C, Denmark. *Correspondence e-mail: mcpe@nbi.ku.dk

The combination of online size-exclusion chromatography and small-angle X-ray scattering (SEC–SAXS) is rapidly becoming a key technique for structural investigations of elaborate biophysical samples in solution. Here, a novel model-refinement strategy centred around the technique is outlined and its utility is demonstrated by analysing data series from several SEC–SAXS experiments on phospholipid bilayer nanodiscs. Using this method, a single model was globally refined against many frames from the same data series, thereby capturing the frame-to-frame tendencies of the irradiated sample. These are compared with models refined in the traditional manner, in which refinement is based on the average profile of a set of consecutive frames from the same data series without an in-depth comparison of individual frames. This is considered to be an attractive model-refinement scheme as it considerably lowers the total number of parameters refined from the data series, produces tendencies that are automatically consistent between frames, and utilizes a considerably larger portion of the recorded data than is often performed in such experiments. Additionally, a method is outlined for correcting a measured UV absorption signal by accounting for potential peak broadening by the experimental setup.

## 1. Introduction

Small-angle scattering (SAS) is a well established and widely applied method that is used to investigate a broad range of soluble samples, ranging from particles of biomolecular origin, including proteins and nucleotide-based structures, to self-assembled systems such as micelles, vesicles and various lipid–protein complexes, including nanodiscs. The use of small-angle scattering for investigating biomolecular structures has triggered large improvements on both the instrument and the sample-environment sides. These improvements have been driven by the frequent scarcity of sample and the relatively small signal over the background, as well as the propensity of many biomolecular samples to aggregate.

The combination of size-exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS) into an integrated SEC–SAXS setup and, more recently, of SEC and small-angle neutron scattering (SANS) into SEC–SANS, are great examples of such improvements (David & Pérez, 2009; Mathew *et al.*, 2004; Watanabe & Inoko, 2009; Jordan *et al.*, 2016; Johansen *et al.*, 2018). Despite the fact that SEC–SAS leads to a dilution of the sample and hence a decreased signal over the background, this is in most cases counterbalanced as the remaining part of the signal comes from a single species or a narrow distribution of species, making the data interpretation less ambiguous.
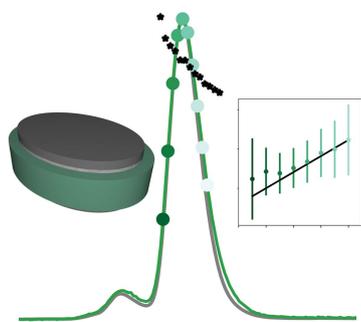
# research papers

With the introduction of SEC–SAXS and SEC–SANS, size-exclusion-based segregation splits the sample into size-sorted fractions from which data are then continuously recorded by SAXS or SANS. Using this setup on a polydisperse sample, the investigator will obtain much more information than if the SAS analysis is performed on the nonfractionated sample. For example, for pure protein samples which are prone to oligomerization this setup may be used to separate and collect information on the different oligomeric states of the protein (Pedersen *et al.*, 2021). Usually, SEC–SAXS and SEC–SANS are used with the goal of overcoming protein-aggregation issues since the sample is irradiated immediately after SEC purification (Johansen *et al.*, 2018; Jeffries *et al.*, 2016; Ryan *et al.*, 2018). In these cases there is a narrow focus on a single species.

There are circumstances in which SEC fails to fully separate molecules with differing structures. Initial SEC–SAS data processing often involves checking for monodispersity within the relevant peak in the chromatogram by calculating radii of gyration ($R_g$) or molecular weight (MW) per frame. The use of a program such as *CHROMIXS* (Panjkovich & Svergun, 2018), for example, makes this process very simple. Using this information, typically the average of a small set of consecutive frames are selected for further analysis. Usually the rest of the SEC–SAS data series is not analysed in depth, despite possibly also containing relevant information about the species. Furthermore, in cases where two or more discrete populations are merged into a single chromatographic peak there are advanced mathematical techniques available, such as state-of-the-art evolving factor analysis (EFA) software (Hopkins *et al.*, 2017; Konarev *et al.*, 2022; Tully *et al.*, 2021), to devolve the overlapping peaks and isolate SAXS profiles corresponding with each population. This is less applicable, however, to the naturally occurring polydispersity around a single species in self-assembled systems.

Nanodiscs are disc-shaped particles consisting of a central lipid bilayer encircled by two amphipathic membrane-scaffolding proteins (MSPs), as depicted in Fig. 1(*a*) (Bayburt
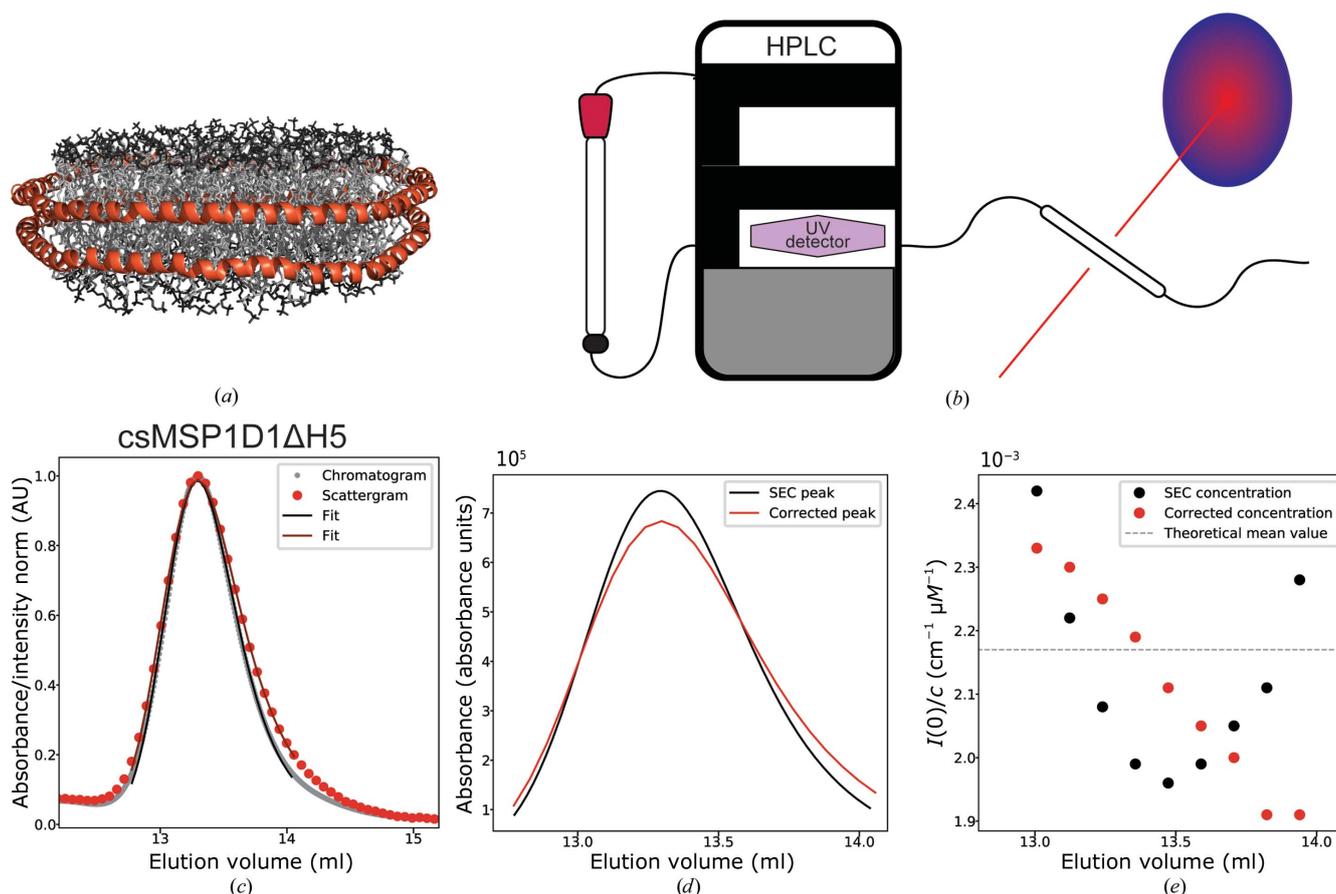


**Figure 1**
The experimental setup and broadening of the peak during SEC–SAXS. (*a*) Molecular visualization of a DMPC-loaded csMSP1D1ΔH5 nanodisc built with CHARMM-GUI *NanodiscBuilder* (Jo *et al.*, 2008; Qi *et al.*, 2019). (*b*) Schematic of the SEC–SAXS setup to reiterate the distance between the HPLC UV280 absorbance detector and the capillary where SAXS is recorded. (*c*) Normalized chromatogram and scattergram for csMSP1D1ΔH5 nanodiscs. The grey points indicate UV absorbance and the red points indicate the total intensity per frame. Solid lines are exponentially modified Gaussian (EMG) fits to the data. The centres of the two peaks are aligned. (*d*) The black profile is the EMG fit to the chromatogram in absorbance units. The red profile is the corrected version substituting in parameters from the fit to the scattergram while keeping the area under the curve constant. (*e*) $I(0)/c$ as a function of the elution volume. Black points are calculated from the original SEC profile. Red points are calculated from the corrected profile. The dashed line is the theoretical value estimated for 120 DMPC per nanodisc.

*et al.*, 2002; Denisov *et al.*, 2004). Nanodiscs are formed by a self-assembly process involving detergent-solubilized lipids and MSPs. The self-assembly is initiated by removal of the detergent, making the lipids and MSPs form particles in a process that is highly dependent on the MSP and lipids of choice. In addition, membrane proteins can be included in the self-assembly, resulting in membrane protein-loaded nano-discs. Due to the presence of lipids, nanodiscs are commonly used as a platform to study the structure and function of membrane proteins in a native-like environment (Denisov & Sligar, 2017).

In this article, we investigate and discuss how the large amount of information obtained in a SEC–SAXS experiment can be brought into play through global analysis of the data. We use dimyristoylphosphatidylcholine (DMPC)-loaded nanodiscs of three various sizes, facilitated through three next-generation circularized (Nasr *et al.*, 2017) and supercharged (Johansen *et al.*, 2019) membrane-scaffold proteins (csMSPs). Circularization refers to the covalent linkage of the MSP N- and C-termini in order to improve size homogeneity, while increasing the number of negatively charged residues enhances the solubility of the nanodisc. The smallest nanodisc that we investigate, csMSP1D1ΔH5, is approximately 8 nm in diameter (Hagn *et al.*, 2013), followed by csMSP1D1, which is approximately 10 nm in diameter (Hagn *et al.*, 2013), and finally csMSP1E3D1, which is 13 nm in diameter (Johansen *et al.*, 2019). The solution structures of these three nanodiscs have been studied previously by offline SEC purification and standard robot SAXS measurements (Johansen *et al.*, 2019, 2021), however, without a focus on the underlying size and shape distributions within the populations. In this study, we demonstrate that this kind of information is easily accessible via SEC–SAXS. To the obtained data we fit a simple geometrical model for the nanodiscs that we have used several times before (Skar-Gislinge & Arleth, 2011; Skar-Gislinge *et al.*, 2010, 2018).

Global fitting of multiple data sets is already used to investigate simultaneously acquired SAXS and SANS data through the fitting of a common model which is then calculated in the relevant contrast. This has been widely exploited and several examples are available in the literature for various types of systems, *i.e.* microemulsions (Arleth & Pedersen, 2001), nanodiscs (Skar-Gislinge *et al.*, 2010), the self-assembly of polymers into toroids (Hollamby *et al.*, 2016) and micelles (Mineart *et al.*, 2019), and in the case of specifically deuterated proteins in solution (Whitten *et al.*, 2007; Heller *et al.*, 2003).

A global fitting approach can also be used to analyse a series of data on the same sample where a subset of the model parameters are conserved throughout the series and others vary. For such shared parameters, a single value is refined for all data sets. For parameters which are not shared, a distinct value is refined for each data set. Such approaches have been applied to diverse cases of analysis of SAXS data from time-dependent fibrillating samples (Herranz-Trillo *et al.*, 2017; Ortore *et al.*, 2011), the variation of monomer–dimer equilibria with concentration (Blobel *et al.*, 2009), temperature-induced aggregation (Mariani *et al.*, 2010; Gonnelli *et al.*,

2020), a SANS analysis of the growth behaviour of SDS micelles (Arleth *et al.*, 2002) and even the analysis of both a series of SAXS data and a series of SANS data simultaneously (Sinibaldi *et al.*, 2008).

The global approach to model fitting has strength in that it ensures a more self-consistent analysis across data sets and with fewer parameters. Additionally, a larger amount of the acquired data are used to evaluate the proposed model and to determine the model parameters. The weakness lies in the added complexity of the modelling setup.

Overall, we show how the global fitting approach provides a more robust analysis of the obtained SEC–SAXS data for nanodiscs. As a part of this, we are able to rationalize the degree of lipid loading in the nanodiscs over the SEC peak. For the small csMSP1D1ΔH5 discs we find that there is very minimal size separation over the peak, but for the slightly larger csMSP1D1 discs as well as the even larger csMSP1E3D1 discs we observe how the SEC splits the sample up into discs with progressively higher to lower lipid-to-MSP stoichiometries. The geometric parameters of the nanodiscs over the SEC peak can be described with a linear frame-to-frame relationship in order to reduce the number of free parameters while still providing a detailed structural overview of the nanodisc populations and without compromising the integrity of the fit to the data sets. The global model provides excellent fits to the whole series of eight SAXS data sets from the same SEC peak simultaneously for each of our three samples. Using our global model we are able to reduce the number of free parameters to 16, compared with 56 free parameters if we were to refine the nanodisc model against eight SAXS frames independently.

As a side note, we introduce a novel approach for quantifying the broadening of the peak during a SEC–SAXS experiment, with the aim of calculating more accurate concentration estimates, which are essential for modelling on an absolute scale.

## 2. Materials and methods

### 2.1. Sample preparation

MSP-based nanodiscs were prepared as described elsewhere (Johansen *et al.*, 2021), excluding the final size-exclusion chromatography (SEC) purification. Briefly, DMPC was solubilized to 50 m$M$ with reconstitution buffer (20 m$M$ Tris–HCl pH 7.5, 150 m$M$ NaCl) containing 100 m$M$ sodium cholate. The solubilized DMPC was mixed with MSP in molar ratios of 55:1 (csMSP1D1ΔH5), 80:1 (csMSP1D1) and 130:1 (csMSP1E3D1) and was diluted with reconstitution buffer to a final DMPC concentration of 10 m$M$. The samples were incubated at 28°C with 15%($w/v$) detergent-absorbing beads (Amberlite XAD-2, Merck) for three hours. The samples were separated from the beads, stored on ice and transported to the SAXS facility.

### 2.2. Data acquisition

SAXS data were collected on BM29 at the European Synchrotron Radiation Facility (ESRF) using the online SEC–

# research papers

SAXS setup (Pernot *et al.*, 2013), where the temperature of the SAXS capillary was kept at 10°C. 200 µl samples were loaded onto a Superdex 200 Increase 10/300 GL column (GE) equilibrated in phosphate buffer. For csMSP1D1ΔH5 and csMSP1E3D1 nanodiscs the buffer was 20 m$M$ sodium phosphate pH 7.0, 150 m$M$ NaCl, while for csMSP1D1 nanodiscs the buffer was phosphate-buffered saline (Sigma) with 1 m$M$ DTT. We note that nanodiscs were initially reconstituted in Tris-based buffer according to standardized procedures; however, the p$K_a$ of Tris is quite temperature-sensitive, and to keep the pH stable in our measurements we opted for buffer exchange into phosphate buffer, which is rather insensitive to temperature. 1 s SAXS frames were continuously measured during sample elution. The intensity was measured as a function of $q$, with $q = 4\pi\sin\theta/\lambda$, where $\theta$ is half the scattering angle and $\lambda$ is the wavelength (here 0.9919 Å), and calibrated to units of cm$^{-1}$ using $H_2O$ as a calibration standard (Orthaber *et al.*, 2000). The absorbance at 280 nm was converted to a concentration using protein extinction coefficients calculated with *ProtParam* (Gasteiger *et al.*, 2005): 18 450 $M^{-1}$ cm$^{-1}$ for csMSP1D1ΔH5 and csMSP1D1 and 26 930 $M^{-1}$ cm$^{-1}$ for csMSP1E3D1. Note that DMPC does not absorb light at this wavelength. The loading nanodisc concentrations were 0.06 m$M$ for csMSP1D1 and csMSP1E3D1 nanodiscs and 0.16 m$M$ for csMSP1D1Δh5 nanodiscs.

## 2.3. Data processing

To reduce the size of the data series, the 1 s SAXS frames were averaged over 10 s. 50 frames collected prior to the elution peak, corresponding to buffer, were then averaged and used for background subtraction. The baseline intensity remains stable before and after the peak, indicating that the chosen buffer frames are suitable for the entire data series (see, for example, Supplementary Fig. S1). SAXS data were rebinned to lie evenly on a logarithmic $q$-scale. Pair-distance [$p(r)$] distributions were obtained by the indirect Fourier transform (IFT) method using the online program *BayesApp* available at https://genapp.rocks/ (Savelyev & Brookes, 2019; Hansen, 2000). Radii of gyration ($R_g$) and the forward scattering [$I(0)$] were calculated using *AUTORG* from *ATSAS* (Petoukhov *et al.*, 2007). Scattergrams were generated by calculating the total intensity in the $q$-range 0.008–0.3 Å$^{-1}$ per SAXS frame and plotting it as a function of elution volume, where we use the HPLC flow rate to convert SAXS time stamps to elution volumes so that the scattergrams and chromatograms can be aligned. The nanodisc model is implemented in *WillItFit* (Pedersen *et al.*, 2013).

## 2.4. Small-angle scattering and principles of the modelling

**2.4.1. Modelling of nanodiscs.** With our SAS data, our main objective is to refine structural models of our nanodiscs from the SEC–SAXS data presented in Fig. 2. The model of choice in this study is the well established nanodisc model (Skar-Gislinge *et al.*, 2010; Skar-Gislinge & Arleth, 2011), in which the geometric structure of the nanodisc is described by a series of form-factor amplitudes, each of which accounts for the

scattering from a distinct part of the nanodisc. The nanodisc model is sketched in Fig. 3(*c*). These form-factor amplitudes have been mathematically described in the literature (Pedersen, 1997). The model is calculated on an absolute scale by utilizing the sample concentration, as well as the molecular composition of the MSP and DMPC, to calculate the scattering length applicable for each part of the nanodisc, as listed in Supplementary Table S7.

Overall, the nanodisc model is described by the following quantities: (i) the axis ratio of the patch of lipid bilayer, $\varepsilon$, (ii) the average area per phospholipid headgroup in the bilayer, $A_L$, (iii) the number of lipids in a nanodisc, $N_L$, (iv) the partial specific molecular volume of a phospholipid, $v_L$, (v) the partial specific molecular volume of an MSP, $v_P$, and (vi) the height of the cylinder describing the protein belt. In this study, we fix this height at 25.8 Å throughout our refinement, in line with previous studies (Bibow *et al.*, 2017). The model is sketched in Fig. 3. Additionally, we refine a constant background contribution, $b$, and a term accounting for the interface roughness in our model, $R$ (Als-Nielsen & McMorrow, 2011). We denote this set of parameters as $\boldsymbol{\theta}$.

Such models are usually refined by minimizing the (reduced) $\chi_r^2$, which estimates the overlap between the data and a specified model function, $I_{Mod}(q, \boldsymbol{\theta})$. This quantity is defined as

$$\chi_r^2 = \frac{1}{N_{DoF}} \sum_{j=1}^{N} \frac{[I_j - I_{Mod}(q_j, \boldsymbol{\theta})]^2}{\sigma_j^2}, \qquad (1)$$

where $q_j$, $I_j$ and $\sigma_j$ constitute the $j$th data point in a data set consisting of $N$ data points. $N_{DoF}$ is the number of degrees of freedom, which we compute as the number of data points minus the number of parameters in the model.

**2.4.2. Global fitting of multiple frames.** In this study, we refined our structural models from several data sets simultaneously and found the best fit for the whole series. As our data sets were collected across a peak in the same SEC experiment, we split our list of parameters into two categories: parameters that we assumed to vary across the irradiated SEC fractions and parameters that we assumed not to vary. All nanodiscs within the same sample comprise the same lipids and MSPs, and hence there should be minimal variation in the volumes of lipids and MSPs. Although there is evidence to suggest that the dynamics and packing of lipids embedded in nanodiscs vary depending on the distance of the lipid from the rim (Bengtsen *et al.*, 2020; Martinez *et al.*, 2017), on average the area per headgroup should remain stable under identical experimental conditions. Rather, depending on sample preparation, there may be a distribution of fully loaded circular discs and underloaded elliptical discs (Skar-Gislinge *et al.*, 2018). Thus for the $k$th data set we refine individual values of $N_L$, $\varepsilon$ and $b$ (which we denote by $\boldsymbol{\theta}_k$). The parameters $v_L$, $v_P$, $A_L$ and $R$ are refined to a single value used in all of the models; we label these parameters $\boldsymbol{\Theta}$.

In order to accommodate for this categorization of our parameters, we redefine our figure of merit, $\chi_r^2$, from equation (1) to
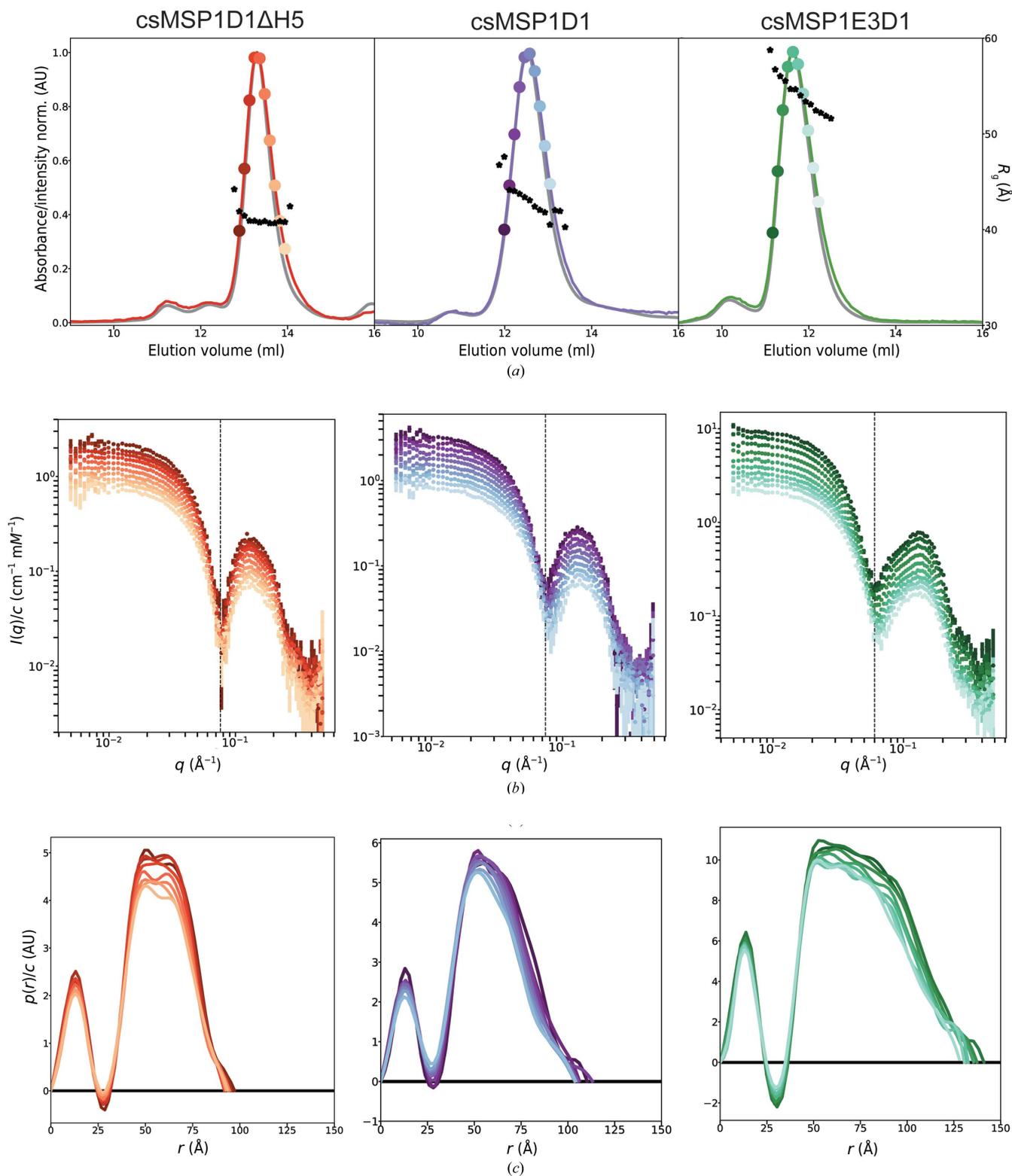
**Figure 2**
SEC–SAXS data indicating structural changes across the size-separated nanodisc samples. (*a*) Normalized SEC elution profiles scaled with SAXS scattergrams. The grey lines indicate UV absorbance at 280 nm. Solid coloured lines indicate the total intensity per frame. The black stars indicate the $R_g$ per frame. (*b*) Series of scattering profiles from various positions in the SEC peak, normalized by concentration, where colours correspond to the highlighted frames in (*a*). The topmost data sets are on an absolute scale, while those below are scaled by $1.1^{-n}$, where $n$ is the frame number. The black dashed line indicates the position of the first minimum of the top scattering profile. (*c*) $p(r)$ distributions corresponding to the highlighted frames, normalized by concentration.

# research papers

$$\chi_{\mathrm{r}}^2 = \frac{1}{N_{\mathrm{DoF}}} \sum_{k=1}^{M} \sum_{j=1}^{N_k} \frac{[I_{k,j} - I_{\mathrm{Mod}}(q_{k,j}, \boldsymbol{\theta}_k, \boldsymbol{\Theta})]^2}{\sigma_{k,j}^2}, \qquad (2)$$

where $N_k$ is the number of data points in the $k$th data set, of which there are $M$, which now prompts us to denote the $j$th data point in the $k$th data set by $(q_{k,j}, I_{k,j}, \sigma_{k,j})$. Note that the model function now depends on not only the parameters specific to the $k$th data set, $\boldsymbol{\theta}_k$, but also the 'global' parameters that are identical across all of the data sets, $\boldsymbol{\Theta}$. This is an adaptation of a similar scheme to analyze temperature series of SAXS data (Johansen *et al.*, 2021).

Additionally, rather than allowing the individual parameters in $\boldsymbol{\theta}_k$ to vary irrespective of the other data sets, this scheme allows us to assume and enforce, for example, linear

trends between the various frames to lower the total number of parameters refined in the scheme: *i.e.* rather than refining $M$ individual values of $N_{\mathrm{L}}$, we assume a linear trend across the SEC fractions, $N_{\mathrm{L}} = an + b$, where $n$ is the frame number in the data series and $a$ and $b$ are parameters to be refined. Hence, we reduce the number of parameters in the refinement scheme by $M - 2$. By employing the same idea for $\varepsilon$, we reduce the number of refined parameters by an additional $M - 2$. In a sense, this notion is a natural extension of the idea of the 'global' parameters in $\boldsymbol{\Theta}$ which are simply kept constant across the frames, and hence their frame-to-frame relationship is described by a single parameter using a zeroth-order polynomial rather than two parameters in a first-order polynomial. We remark that a linear function is sufficient for our
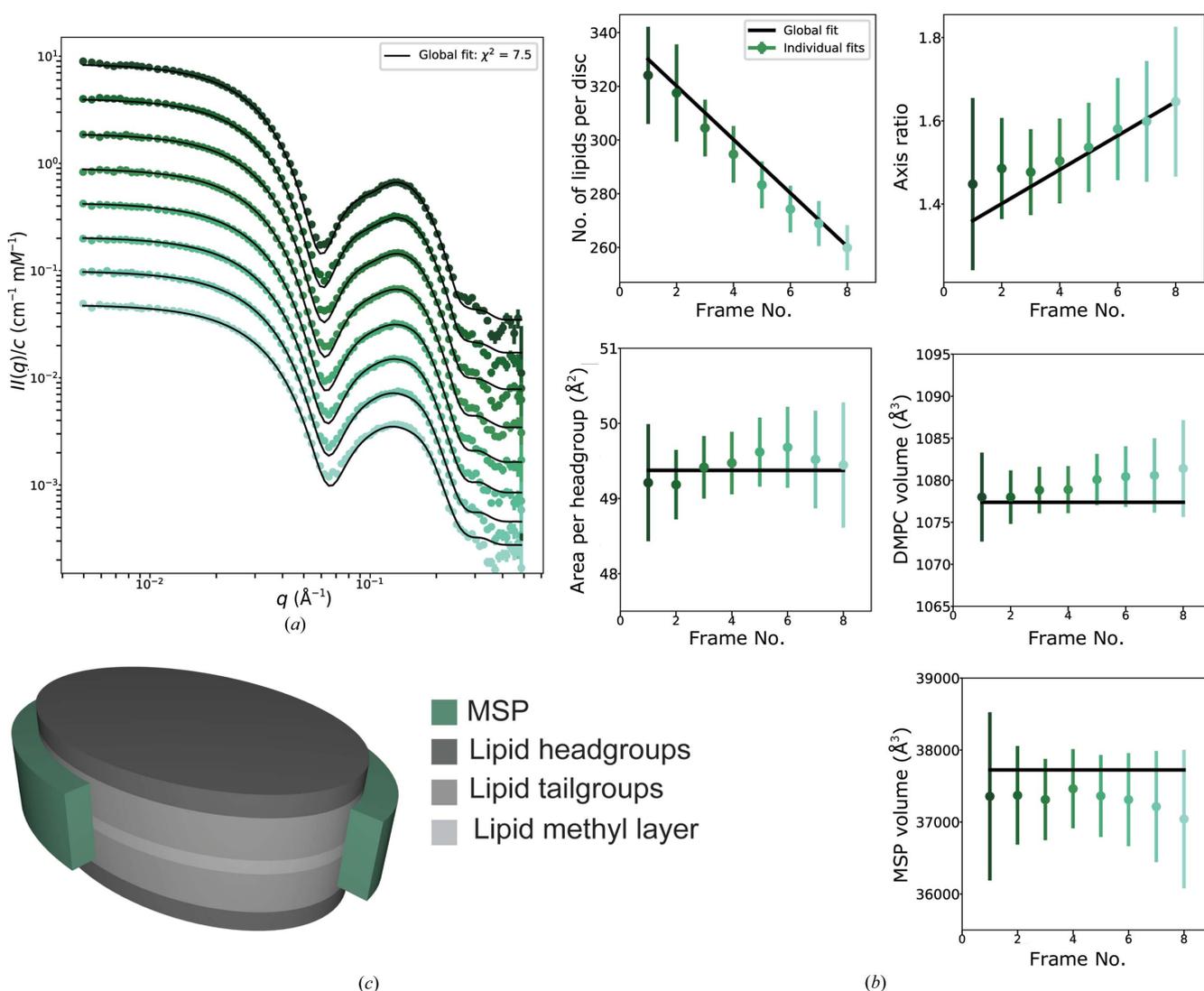


**Figure 3**
Model fit results for csMSP1E3D1 nanodiscs. (*a*) Global fit to experimental SAXS data sets from frames with increasing elution volumes/positions across the SEC peak. Data sets are the middle eight highlighted frames in Fig 2(*a*). (Individual fits are shown in Supplementary Fig. S3.) The topmost data set is on an absolute scale, while those below are scaled by $2^{-n}$, where $n$ is the frame number. (*b*) Refined structural parameters. The coloured data points indicate parameters refined from each data set individually. The black lines indicate parameters refined from the global fit, where one shared value is found for $A_{\mathrm{L}}$, $v_{\mathrm{P}}$ and $v_{\mathrm{L}}$, while $N_{\mathrm{L}}$ and $\varepsilon$ are both forced to follow a linear trend. (*c*) Representation of the nanodisc model used; a quarter of the MSP belt is not shown to highlight the interior structure of the lipid bilayer.

purposes; providing a more physical model to describe particles eluting from a SEC column could require a more complicated relationship and further investigation is necessary before drawing conclusions. More complicated relationships can readily be employed but become impractical (or simply useless) if they require a number of coefficients comparable to the number of data sets, unless there is a solid underlying theory to support their use.

## 3. Results and discussion

### 3.1. Co-calibration of the SEC–UV280 and the SEC–SAXS intensities

The SEC–SAXS setup is sketched in Fig. 1(b). Broadening of the elution peak often occurs during SEC–SAXS experiments due to Taylor dispersion (Taylor, 1953) and the difference in diameter between the HPLC tubing and the SAXS capillary (Bucciarelli *et al.*, 2018). Here, we introduce a novel approach for estimating and correcting for this broadening. The approach is illustrated in Figs. 1(c) and 1(d) and Supplementary Fig. S2. In Fig. 1(c) the normalized chromatogram for csMSP1D1ΔH5 nanodiscs is plotted with its corresponding scattergram, *i.e.* the scattering intensity per individual frame as a function of the elution volume. The centre of the peak of the scattergram is aligned with the centre of the peak of the chromatogram and the broadening of the scattergram is clearly visible.

Exponentially modified Gaussian (EMG) functions are good models for chromatographic peaks under a range of conditions (Naish & Hartwell, 1988; Busnel *et al.*, 2001), where broadening can be characterized by two parameters: the standard deviation (width), $\sigma$, and a relaxation parameter (skew), $\tau$. EMGs were fitted to the main peaks of the chromatogram and scattergram via nonlinear least-squares regression. A 'corrected' SEC profile was then calculated by keeping the area under the EMG fit of the chromatogram constant, but substituting in $\sigma$ and $\tau$ from the fit to the scattergram in order to take account of the change in the shape of the peak, which becomes wider and develops a tail on the right-hand side. Thus, the corrected profile approximates the UV absorption as if it were recorded directly on the SAXS capillary and should provide much more accurate concentration estimates. The original SEC peak and the corrected SEC peak can be compared in Fig. 1(d). Estimating the sample concentration directly from the raw HPLC absorption measurements may lead to underestimated concentrations in the tails of the peak and overestimated concentrations in the centre.

As a check, we calculated the forward scattering $I(0)$ divided by concentration for the SAXS data sets as a function of elution volume, as plotted in Fig. 1(e). The values calculated with the original concentrations show a prominent decrease and then an increase, which cannot be readily explained. For a fully homogeneous sample, $I(0)/c$ should remain constant. If there is some size variation $I(0)/c$ may decrease systematically towards the right-hand side, which is seen for the values

calculated with the corrected concentrations. These values also fall close to an estimate of $I(0)/c$ which we calculated for csMSP1D1ΔH5 nanodiscs loaded with 120 DMPC. We note that during modelling the nanodisc form factor multiplied by the new concentrations matches the experimental SAXS intensities perfectly without the need for an additional scaling factor.

One potential drawback of this method lies in the fact that scattering intensity scales with squared particle volume while protein UV absorption does not, meaning that some discrepancy between the shapes of the chromatograms and the scattergrams is to be expected. In this case, however, the corrected SEC profile performed better and the method could be considered for other SEC–SAXS studies in which accurate concentration estimates are desirable for absolute-scale modelling or molecular-weight determination.

### 3.2. SEC–SAXS data overview

The SEC–SAXS data and associated $p(r)$ distributions for all three nanodiscs species are shown in Fig. 2. For each nanodisc species the data indicate some systematic structural variation across the size-sorted fractions. For the smallest nanodiscs, $R_g$ stays constant across the SEC peak at ~40 Å; however, for csMSP1D1 nanodiscs there is a steady decrease from ~46 to 42 Å, and for csMSP1E3D1 nanodiscs the decrease from ~58 to 52 Å is even more apparent. Each of the scattering curves is compatible with that we typically observe for monodisperse nanodiscs: a flat Guinier region in the low-$q$ regime, followed by a trough and a broad bump at medium to high $q$. csMSP1D1ΔH5 and csMSP1E3D1 display the typical nanodisc double-bump feature (Skar-Gislinge *et al.*, 2010; Denisov *et al.*, 2005). For csMSP1D1, and even more significantly for csMSP1E3D1, as the position of the fraction in the elution profile progresses, the first minimum in the scattering curve shifts systematically to higher $q$ values, indicating a change in particle shape. The $p(r)$ distributions reaffirm this, showing a systematic loss of depth of the first minimum alongside a decrease in the maximum pair distance ($D_{max}$) as we move to larger elution volumes. Again, these variations are least prominent in the small discs and most prominent in the large discs, which may suggest that larger discs are more structurally disperse. Altogether, these observations suggest that even within a SEC-purified nanodisc population there is some size distribution which may be sorted by a SEC column so that larger particles elute first, but below some resolution it will not be separated into multiple elution peaks.

### 3.3. Modelling and data analysis

Analysing many data sets from the same SEC–SAXS experiment with the nanodisc model provides more detailed insights into the size and shape distributions underlying the populations. We select eight sequential SAXS data sets for each sample. Firstly, we refine the model against each data set independently as a precursor. Secondly, we refine the model against each data set *simultaneously* with both global and frame-specific free parameters in order to constrain the fits

# research papers

further and investigate the amount of information which can be extracted with a reduced number of free parameters. We note that although each of the individual data sets are collected from a narrow fraction of the SEC-purified sample, the data sets may still contain some slight overlap between different nanodisc sizes. The refined model therefore describes the average scattering from the nanodiscs present and does not account for polydispersity within a certain frame.

**3.3.1. Individual fits.** When fitted to the individual frames, the nanodisc model provides excellent fits to each of the SAXS data sets chosen for further analysis. The individual fits are plotted in Supplementary Figs. S3, S4 and S5. The refined model parameters from individual fits to the eight SAXS data sets for csMSP1E3D1 nanodiscs are plotted as coloured points in Fig. 3(b) and are further listed in Supplementary Table S1. The results for csMSP1D1 and csMSP1D1$\Delta$H5 nanodiscs are given in the supporting information.

For all three nanodisc samples the area per lipid headgroup, $A_L$, and the partial specific molecular volumes of the lipid, $v_L$, and MSP, $v_P$, generally fluctuate only slightly between frames. This is in line with our expectations since the volume of DMPC and of each MSP should be very stable for the entire sample, regardless of elution volume. Although prone to local fluctuations, the refined value of the area per headgroup should also remain stable. For the three nanodiscs, $A_L$ was refined to values of between 49.5 and 53.5 Å$^2$, which is in good agreement with previous values of 47.5 Å$^2$ for DMPC-loaded nanodiscs (Johansen *et al.*, 2021), 52.1 Å$^2$ for DMPC-loaded peptide discs (Midtgaard *et al.*, 2014) and 47.2 Å$^2$ for a pure DMPC bilayer (Tristram-Nagle *et al.*, 2002), all of which were recorded at 10°C. We mention that since the temperature is not controlled over the entire SEC–SAXS instrumentation, the temperature of the sample may be slightly above 10°C. This may affect the lipid packing slightly; however, as the temperature was kept below the melting temperature of DMPC at 24°C the effect will not be prominent (Johansen *et al.*, 2021). $v_L$ becomes up to 5% larger than the reported value of 1041 Å$^3$ (Tristram-Nagle *et al.*, 2002). $v_P$ stays within 5% below our pre-estimated values based on the molecular compositions, which are specific for each MSP. Prominent frame-to-frame fluctuations of these three free parameters could be the result of overfitting to the SAXS data and strong correlations between parameters in the model.

Rather, the systematic variations in the SAXS data sets are reflected in the steady decrease in the number of lipids per nanodisc, $N_L$, as a function of elution volume, likely coinciding with a general increase in the axis ratio, $\varepsilon$. Since the circumference of the nanodisc is determined by the length of the MSP and is therefore expected to remain constant, variation in the number of lipids (and thereby the bilayer surface area) must be compensated by some variation in the shape of the disc. Although $\varepsilon$ is poorly determined by this method, we assume that this dependency between $N_L$ and $\varepsilon$ is present across the sample. Each data set indicates elliptical discs, where discs with higher lipid-to-MSP stoichiometries appear to be slightly rounder, while discs with lower lipid-to-MSP stoichiometries become more elliptical. The same trend has

been observed many times (Skar-Gislinge *et al.*, 2010, 2018; Graziano *et al.*, 2018). According to our analysis, csMSP1E3D1 nanodiscs contain the largest underlying size distribution, with a difference in $N_L$ of 65 lipids between the size-sorted first and last frames, from 325 to 260 lipids. csMSP1D1 decreases by 35 lipids from 150 to 115 and csMSP1D1$\Delta$H5 decreases by 15 lipids from 130 to 115.

Unlike previous reports (Johansen *et al.*, 2021), we do not see a simple linear correlation between axis ratio and length of the MSP here, despite larger discs theoretically being more structurally flexible. csMSP1D1 nanodiscs persistently have the largest axis ratio, which varies between 1.6 and 1.8, whereas csMSP1D1$\Delta$H5 nanodiscs have the smallest, varying between 1.3 and 1.5, and csMSP1E3D1 lies in between with values varying between 1.45 and 1.65. Although seemingly incidental, this coincides with a recent course-grained molecular-dynamics study of the same circularized MSPs (cMSPs, non-supercharged; Kjølbye *et al.*, 2021), where cMSP1D1 was found to have the highest degree of anisotropy, with cMSP1D1$\Delta$H5 being the most circular and cMSP1E3D1 falling in between. These results suggest that there are other factors influencing the shape of nanodiscs besides the degree of lipid loading, especially the choice of MSP and its intrinsic rigidity.

**3.3.2. Global fits.** Fitting the nanodisc model to $M$ data sets requires $7M$ free parameters. Certain parameters, however, should be conserved when examining data sets from the same SEC–SAXS experiment and hence fitting the parameter $M$ times becomes redundant. The individual fits justify the introduction of global parameters for $A_L$, $v_L$, $v_P$ and $R$ to ensure that the model refinement is self-consistent and that these parameters are better determined. $N_L$ and $\varepsilon$, however, capture important trends between the data sets as a function of elution volume. This information would be lost if fitting using a constant rather than the two-parameter function that we utilized here.

A global model *could* be set up with $A_L$, $v_L$, $v_P$ and $R$ as global parameters and $N_L$, $\varepsilon$ and $b$ as frame-specific parameters, such that the number of free parameters is $4 + 3M$. However, to constrain the fit even further, frame-to-frame linear relationships are enforced for $N_L$ and $\varepsilon$, where the $y$ intercept and slope of the respective functions are global parameters as described in Section 2.4.2 and shown in the top row in Fig 3(b), capturing increasing or decreasing trends across the data series using only two free parameters per function. In this implementation of the model, the number of free parameters is $8 + M$, where the only frame-specific parameter is the background, $b$. In this case, where $M = 8$, swapping from individual modelling to the global modelling described here drastically reduces the number of free parameters from 56 ($7 \times 8$) to 16 ($8 + 8$).

Fig. 3(a) shows the global fit refined against the eight SAXS data sets simultaneously for csMSP1E3D1 nanodiscs. The refined model parameters are listed in Supplementary Table S1 and the frame-to-frame relationships are plotted in Fig. 3(b) as solid black lines. Global results for csMSP1D1 and csMSP1D1$\Delta$H5 nanodiscs are given in the supporting infor-

mation. Despite the extra constraints, the global model is able to describe the entire series of SAXS data sets excellently, with no features standing out visually as poorly captured. The global model achieves impressive $\chi_r^2$ values of 7.5, 5.4 and 5.9 for csMSP1E3D1, csMSP1D1 and csMSP1D1$\Delta$H5, respectively, as calculated by equation (2). Furthermore, reasonable structural parameters are maintained over the three samples and the important frame-to-frame trends are sustained.

For csMSP1E3D1 and csMSP1D1$\Delta$H5 the global fit parameters mimic the individual fit parameters very closely, which suggests that the results are reliable and the choice of frame-specific and global parameters are compatible. For csMSP1D1 the global fit parameters, although still satisfactory, are a slightly looser match to the individual fit parameters, especially the axis ratio, where the global model possibly determines a much steeper slope. We note that this could be explained by the fact that this data series has the poorest signal-to-noise ratio. We further comment that the large error on the $\varepsilon$ slope for all three experiments should be expected since it is clear in the individual fits that $\varepsilon$ is poorly determined and a range of slopes could be applicable. Refined global fit parameters should not be anticipated to emerge as the exact mean of the individual fit results, since the global fit minimizes the risk of overfitting to the SAXS data and constrains correlations between fit parameters.

We observe that the refined gradient of the straight line representing the fraction-dependent change in the number of lipids, $N_L$, further rationalizes the degree of polydispersity present in each respective nanodisc sample: the largest disc csMSP1E3D1 shows the greatest gradient of $-9.96N_L$ per frame, with csMSP1D1 showing a gradient of $-5.00N_L$ per frame and csMSP1D1$\Delta$H5 showing the most gentle gradient of $-1.47N_L$ per frame. These slopes can be compared with linear fits to $R_g$ as a function of position, where we calculate slopes of $-0.23$, $-0.18$ and $-0.03$ Å per frame for csMSP1E3D1, csMSP1D1 and csMSP1D1$\Delta$H5, respectively.

Summing up, employing frame-to-frame constraints in our analysis of the presented SEC–SAXS data seems to allow considerably more constrained fits of a large amount of data whilst still producing realistic models and capturing inter-frame trends in a quantitative manner. The most notable advantages are the considerable reduction in the total number of parameters refined from the data and the tractability of refining a single model accounting for all of the data sets rather than individual models from each data set, which are then to be compared at a later stage; both of which in the cases presented here seem to come at little expense in terms of the quality of the fits.

## 4. Conclusions and further perspectives

Often during SEC–SAXS analysis only a small fraction of the SEC peak is considered and a large amount of structural information is discarded. We perform a comprehensive investigation into three types of next-generation nanodiscs by analysing many SAXS data sets from the same SEC–SAXS experiment. The size-sorted SAXS data sets reveal some systematic polydispersity within the structure of the nanodisc populations. A global approach to model fitting provides a robust analysis to help characterize the polydispersity. We observe that the SEC column gradually splits the samples into discs with high and low lipid-to-MSP stoichiometries. We employ simple frame-to-frame linear functions to further reduce the number of free parameters in the fitting routine. Despite the extra constraints, the global model is able to describe the entire series of SAXS data sets excellently and provides a detailed overview of the nanodisc populations through frame-specific and global refined values.

The reduction in the number of parameters refined from the data sets is a particularly attractive attribute of the outlined modelling scheme. Like similar inference tasks, model refinement from small-angle scattering data is prone to overfitting, so these simplifications (in terms of number of parameters in the model) provide a convenient means of analyzing the extensive amount of data one obtains from, for example, a SEC–SAXS experiment in a somewhat constrained manner. Naturally, such schemes rely intrinsically on the validity of the assumed trends across the analyzed data sets. Here, we successfully employ constant and linear relationships and argue that they are indeed sufficient to capture the general behavior of our data; mostly as we observe little to no increase in our figure of merit and the overall quality of our fits by employing them.

Our method has general applicability for samples and systems with inherent polydispersity within the resolution of the SEC column, including cases where the SEC peak is asymmetric or where two peaks have merged together. These include nanodiscs, as presented here, as well as similar membrane-protein carrier systems, including di-block co-polymer lipid particles (Knowles *et al.*, 2009), saposin lipid particles (Frauenfeld *et al.*, 2016) and detergent micelles. Additionally, the method could be modified to analyse biological systems in different types of equilibrium and where distinct populations cannot be sufficiently separated on SEC for individual analysis (Vestergaard, 2016). These include, for example, protein monomer–dimer equilibria, protein–ligand equilibria, phase-separated disordered proteins or systems adopting different structural states. In these cases, our method could be complementary to the popular evolving factor analysis (EFA) programs where model-independent EFA can be employed to identify and isolate uncontaminated profiles of the distinct populations for further structural analysis, potentially including global fitting (although of only two or three data sets). With EFA it is possible to extract an overall picture of sources of extreme structural heterogeneity within a sample. Previous examples include identifying scattering contributions from massive contaminants (Meisburger *et al.*, 2016), separating protein monomers from dimers or oligomers (Hopkins *et al.*, 2017; Konarev *et al.*, 2022) and separating bound and unbound protein states (Tully *et al.*, 2021). Our presented method is more suitable, however, when the desired outcome is a continuous description of systematic polydispersity across a data series, particularly when there is an underlying distri-

# research papers

bution within a single population or when the amount of polydispersity is too small for EFA to detect. This is only possible by investigating many narrow fractions of the elution profile. Furthermore, EFA fails when the chromatographic peak is too asymmetrical or when two peaks are too close together (Konarev *et al.*, 2022). In this work we analyse data sets directly from SEC–SAXS and assume that each fraction contains only a single population; however, one should be cautious since this is not necessarily true under the resolution of the SEC column.

Furthermore, our global fitting scheme is readily suitable for SEC–SANS experiments, and would be a very powerful fitting platform if the model could be refined against series of SEC–SAXS data sets and series of SEC–SANS data sets simultaneously. Finally, issues with peak broadening are well acknowledged in the SEC–SAXS community (Ryan *et al.*, 2018) and efforts have been made to measure the absorption directly on the SAXS capillary (Bucciarelli *et al.*, 2018). As part of our overall method, we suggest a simple correction procedure for the online absorption measurement, which eliminates parts of the problem with peak broadening and thereby allows more accurate determination of the forward scattering and thereby parameters such as molecular weight.

### References

Als-Nielsen, J. & McMorrow, D. (2011). *Elements of Modern X-ray Physics*, 2nd ed. Chichester: John Wiley & Sons.
Arleth, L., Bergström, M. & Pedersen, J. S. (2002). *Langmuir*, **18**, 5343–5353.
Arleth, L. & Pedersen, J. S. (2001). *Phys. Rev. E*, **63**, 061406.
Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. (2002). *Nano Lett.* **2**, 853–856.
Bengtsen, T., Holm, V. L., Kjølbye, L. R., Midtgaard, S. R., Johansen, N. T., Tesei, G., Bottaro, S., Schiøtt, B., Arleth, L. & Lindorff-Larsen, K. (2020). *eLife*, **9**, e56518.
Bibow, S., Polyhach, Y., Eichmann, C., Chi, C. N., Kowal, J., Albiez, S., McLeod, R. A., Stahlberg, H., Jeschke, G., Güntert, P. & Riek, R. (2017). *Nat. Struct. Mol. Biol.* **24**, 187–193.
Blobel, J., Bernadó, P., Svergun, D. I., Tauler, R. & Pons, M. (2009). *J. Am. Chem. Soc.* **131**, 4378–4386.
Bucciarelli, S., Midtgaard, S. R., Pedersen, M. N., Skou, S., Arleth, L. & Vestergaard, B. (2018). *J. Appl. Cryst.* **51**, 1623–1632.
Busnel, J. P., Foucault, F., Denis, L., Lee, W. & Chang, T. (2001). *J. Chromatogr. A*, **930**, 61–71.
David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.
Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. (2004). *J. Am. Chem. Soc.* **126**, 3477–3487.

Denisov, I. G., McLean, M. A., Shaw, A. W., Grinkova, Y. V. & Sligar, S. G. (2005). *J. Phys. Chem. B*, **109**, 15580–15588.
Denisov, I. G. & Sligar, S. G. (2017). *Chem. Rev.* **117**, 4669–4713.
Frauenfeld, J., Löving, R., Armache, J.-P., Sonnen, A. F.-P., Guettou, F., Moberg, P., Zhu, L., Jegerschöld, C., Flayhan, A., Briggs, J. A. G., Garoff, H., Löw, C., Cheng, Y. & Nordlund, P. (2016). *Nat. Methods*, **13**, 345–351.
Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M., Appel, R. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, pp. 571–607. Totowa: Humana Press.
Gonnelli, A., Pieraccini, S., Baldassarri, E. J., Funari, S., Masiero, S., Ortore, M. G. & Mariani, P. (2020). *Nanoscale*, **12**, 1022–1031.
Graziano, V., Miller, L. & Yang, L. (2018). *J. Appl. Cryst.* **51**, 157–166.
Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. (2013). *J. Am. Chem. Soc.* **135**, 1919–1925.
Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.
Heller, W. T., Finley, N. L., Dong, W.-J., Timmins, P., Cheung, H. C., Rosevear, P. R. & Trewhella, J. (2003). *Biochemistry*, **42**, 7790–7800.
Herranz-Trillo, F., Groenning, M., van Maarschalkerweerd, A., Tauler, R., Vestergaard, B. & Bernadó, P. (2017). *Structure*, **25**, 5–15.
Hollamby, M. J., Aratsu, K., Pauw, B. R., Rogers, S. E., Smith, A. J., Yamauchi, M., Lin, X. & Yagai, S. (2016). *Angew. Chem.* **128**, 10044–10047.
Hopkins, J. B., Gillilan, R. E. & Skou, S. (2017). *J. Appl. Cryst.* **50**, 1545–1553.
Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E. & Svergun, D. I. (2016). *Nat. Protoc.* **11**, 2122–2153.
Jo, S., Kim, T., Iyer, V. G. & Im, W. (2008). *J. Comput. Chem.* **29**, 1859–1865.
Johansen, N. T., Luchini, A., Tidemand, F. G., Orioli, S., Martel, A., Porcar, L., Arleth, L. & Pedersen, M. C. (2021). *Langmuir*, **37**, 6681–6690.
Johansen, N. T., Pedersen, M. C., Porcar, L., Martel, A. & Arleth, L. (2018). *Acta Cryst.* D**74**, 1178–1191.
Johansen, N. T., Tidemand, F. G., Nguyen, T. T., Rand, K. D., Pedersen, M. C. & Arleth, L. (2019). *FEBS J.* **286**, 1734–1751.
Jordan, A., Jacques, M., Merrick, C., Devos, J., Forsyth, V. T., Porcar, L. & Martel, A. (2016). *J. Appl. Cryst.* **49**, 2015–2020.
Kjølbye, L. R., De Maria, L., Wassenaar, T. A., Abdizadeh, H., Marrink, S. J., Ferkinghoff-Borg, J. & Schiøtt, B. (2021). *J. Chem. Inf. Model.* **61**, 2869–2883.
Knowles, T. J., Finka, R., Smith, C., Lin, Y.-P., Dafforn, T. & Overduin, M. (2009). *J. Am. Chem. Soc.* **131**, 7484–7485.
Konarev, P. V., Graewert, M. A., Jeffries, C. M., Fukuda, M., Cheremnykh, T. A., Volkov, V. V. & Svergun, D. I. (2022). *Protein Sci.* **31**, 269–282.
Mariani, P., Spinozzi, F., Federiconi, F., Ortore, M. G., Amenitsch, H., Spindler, L. & Drevensek-Olenik, I. (2010). *J. Nucleic Acids*, **2010**, 472478.
Martinez, D., Decossas, M., Kowal, J., Frey, L., Stahlberg, H., Dufourc, E. J., Riek, R., Habenstein, B., Bibow, S. & Loquet, A. (2017). *ChemPhysChem*, **18**, 2651–2657.
Mathew, E., Mirza, A. & Menhart, N. (2004). *J. Synchrotron Rad.* **11**, 314–318.
Meisburger, S. P., Taylor, A. B., Khan, C. A., Zhang, S., Fitzpatrick, P. F. & Ando, N. (2016). *J. Am. Chem. Soc.* **138**, 6506–6516.
Midtgaard, S. R., Pedersen, M. C., Kirkensgaard, J. J. K., Sørensen, K. K., Mortensen, K., Jensen, K. & Arleth, L. (2014). *Soft Matter*, **10**, 738–752.
Mineart, K. P., Ryan, J. J., Appavou, M. S., Lee, B., Gradzielski, M. & Spontak, R. J. (2019). *Langmuir*, **35**, 1032–1039.
Naish, P. J. & Hartwell, S. (1988). *Chromatographia*, **26**, 285–296.
Nasr, M. L., Baptista, D., Strauss, M., Sun, Z. J., Grigoriu, S., Huser, S., Plückthun, A., Hagn, F., Walz, T., Hogle, J. M. & Wagner, G. (2017). *Nat. Methods*, **14**, 49–52.
Orthaber, D., Bergmann, A. & Glatter, O. (2000). *J. Appl. Cryst.* **33**, 218–225.

Ortore, M. G., Spinozzi, F., Vilasi, S., Sirangelo, I., Irace, G., Shukla, A., Narayanan, T., Sinibaldi, R. & Mariani, P. (2011). *Phys. Rev. E*, **84**, 061904.

Panjkovich, A. & Svergun, D. I. (2018). *Bioinformatics*, **34**, 1944–1946.

Pedersen, D. V., Pedersen, M. N., Mazarakis, S. M., Wang, Y., Lindorff-Larsen, K., Arleth, L. & Andersen, G. R. (2021). *eLife*, **10**, e63356.

Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.

Pedersen, M. C., Arleth, L. & Mortensen, K. (2013). *J. Appl. Cryst.* **46**, 1894–1898.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L. & McSweeney, S. (2013). *J. Synchrotron Rad.* **20**, 660–664.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s223–s228.

Qi, Y., Lee, J., Klauda, J. B. & Im, W. (2019). *J. Comput. Chem.* **40**, 893–899.

Ryan, T. M., Trewhella, J., Murphy, J. M., Keown, J. R., Casey, L., Pearce, F. G., Goldstone, D. C., Chen, K., Luo, Z., Kobe, B.,

McDevitt, C. A., Watkin, S. A., Hawley, A. M., Mudie, S. T., Samardzic Boban, V. & Kirby, N. (2018). *J. Appl. Cryst.* **51**, 97–111.

Savelyev, A. & Brookes, E. (2019). *Future Gener. Comput. Syst.* **94**, 929–936.

Sinibaldi, R., Ortore, M. G., Spinozzi, F., de Souza Funari, S., Teixeira, J. & Mariani, P. (2008). *Eur. Biophys. J.* **37**, 673–681.

Skar-Gislinge, N. & Arleth, L. (2011). *Phys. Chem. Chem. Phys.* **13**, 3161–3170.

Skar-Gislinge, N., Johansen, N. T., Høiberg-Nielsen, R. & Arleth, L. (2018). *Langmuir*, **34**, 12569–12582.

Skar-Gislinge, N., Simonsen, J. B., Mortensen, K., Feidenhans'l, R., Sligar, S. G., Lindberg Møller, B., Bjørnholm, T. & Arleth, L. (2010). *J. Am. Chem. Soc.* **132**, 13713–13722.

Taylor, G. (1953). *Proc. R. Soc. London Ser. A*, **219**, 186–203.

Tristram-Nagle, S., Liu, Y., Legleiter, J. & Nagle, J. F. (2002). *Biophys. J.* **83**, 3324–3335.

Tully, M., Tarbouriech, N., Rambo, R., Hutin, S., Tully, M. D. & Rambo, R. P. (2021). *J. Vis. Exp.*, e61578.

Vestergaard, B. (2016). *Arch. Biochem. Biophys.* **602**, 69–79.

Watanabe, Y. & Inoko, Y. (2009). *J. Chromatogr. A*, **1216**, 7461–7465.

Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trewhella, J. & Langley, D. B. (2007). *J. Mol. Biol.* **368**, 407–420.

1

# 1. Supporting Information

*SI.1. Choice of buffer frames*
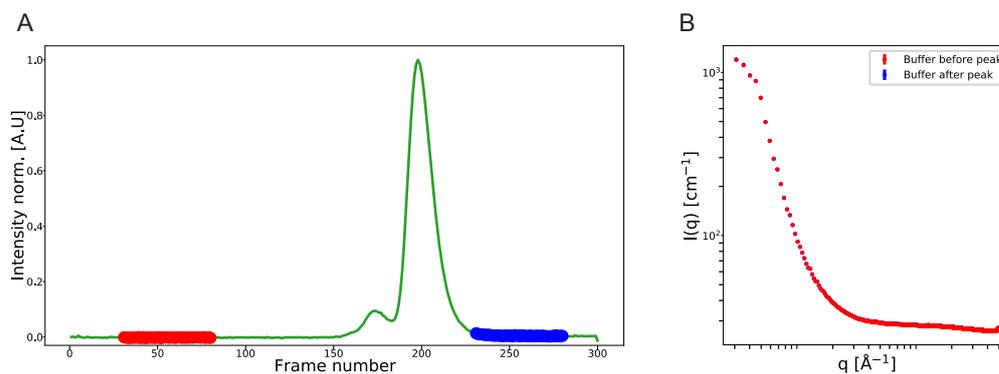


Fig. SI.1. Buffer stability in SEC-SAXS for csMSP1E3D1 nanodiscs. (A) The green profile is the full scattergram as a function of frame number, where each frame is 10 1s exposures. The red dots indicate the 50 SAXS datasets which were averaged for background subtraction. The baseline is restored after the peak. (B) Red: the average scattering profile corresponding with the red dots. : the average scattering profile of 50 datasets corresponding with the blue dots. The profiles are compared to check for potential complications regarding buffer selection. The difference between the two profiles is minimal and not visible in the plot.

2

*SI.2. Co-calibration the SEC-UV280 and the SEC-SAXS intensities*
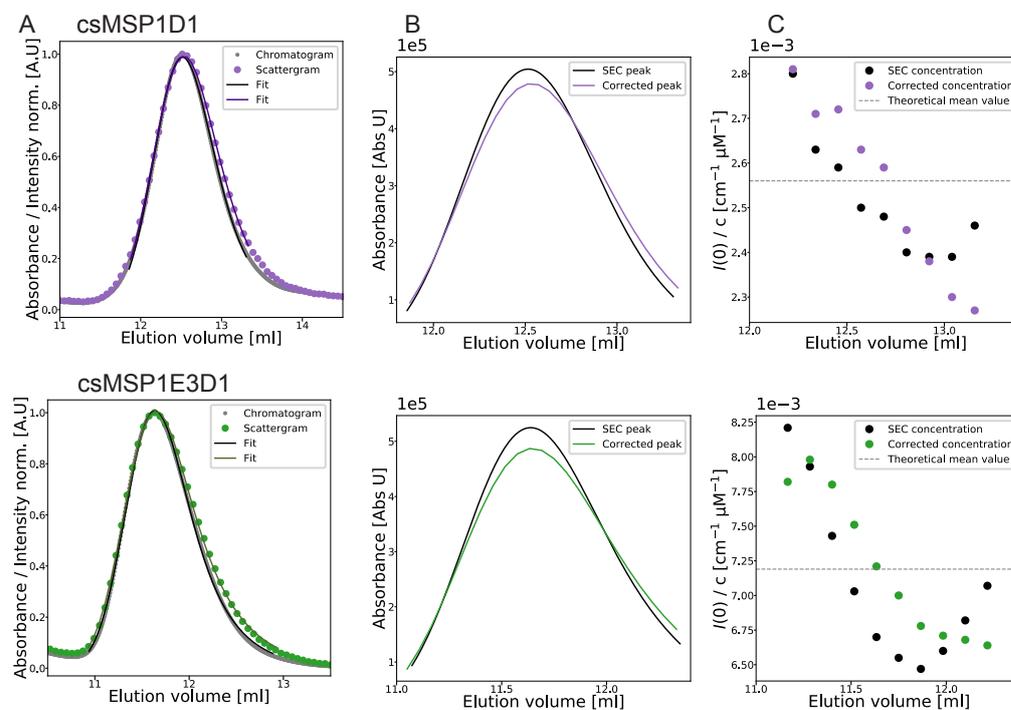


Fig. SI.2. (A) Normalised chromatogram and scattergram for csMSP1D1/ csMSP1E3D1 nanodiscs. The grey points indicate UV absorbance and the purple/ green points indicate the total intensity per SAXS frame. Solid lines are exponentially modified Gaussian (EMG)fits to the data. (B) The black profiles are the EMG fit to the chromatograms in Absorbance Units. The purple/green profiles are the corrected version substituting in $\sigma$ and $\tau$ from the fit to the scattergram while keeping the area under the curve constant. (C) $I(0)/c$ as a function of Elution volume. Black points are calculated from the original SEC profiles. Purple / green points are calculated from the corrected profiles. The dashed lines are the theoretical values estimated for 130 DMPC per nanodisc / 300 DMPC per nanodisc.

*SI.3. Results*

*SI.3.1. Model parameters for csMSP1E3D1 nanodiscs*  Tables SI.1 to SI.2 contain the

parameters refined from the presented data and fits in Figure 3.



Fig. SI.3. Individual fits to experimental SAXS datasets with increasing elution vol-
umes/ positions across the SEC-peak. Datasets are the middle eight highlighted
frames in Figure 2A. The topmost dataset is on absolute scale while those below
are scaled by $2^{-n}$ where $n$ is the frame number.

Individual fits

| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $N_L$ | 324 ±18 | 318 ±11 | 305 ±9 | 295 ±8 | 283 ±8 | 274 ±9 | 269 ±11 | 260 ±13 |
| $\epsilon$ | 1.45 ±0.21 | 1.49 ±0.12 | 1.48 ±0.10 | 1.50 ±0.10 | 1.53 ±0.11 | 1.58 ±0.12 | 1.60 ±0.15 | 1.65 ±0.18 |
| $A_L$ [Å$^2$] | 49.2 ±0.8 | 49.2 ±0.5 | 49.4 ±0.4 | 49.5 ±0.4 | 49.6 ±0.5 | 49.5 ±0.5 | 49.4 ±0.6 | 49.4 ±0.8 |
| $v_L$ [Å$^3$] | 1078 ±5 | 1078 ±3 | 1079 ±3 | 1079 ±3 | 1080 ±3 | 1080 ±3 | 1081 ±4 | 1081 ±6 |
| $v_p$ [Å$^3$] | 37360 ±1170 | 37370 ±680 | 37310 ±560 | 37460 ±550 | 37360 ±570 | 37310 ±650 | 37210 ±770 | 37040 ±960 |
| $R$ [Å] | 4.32 ±0.32 | 4.28 ±0.19 | 4.28 ±0.16 | 4.32 ±0.16 | 4.30 ±0.17 | 4.29 ±0.19 | 4.22 ±0.23 | 4.13 ±0.30 |
| $b$ [10$^{-5}$ cm$^{-1}$] | -17.0 ±5.4 | -20.3 ±3.9 | -24.0 ±3.9 | -25.1 ±3.9 | -23.6 ±3.9 | -21.1 ±3.9 | -16.8 ±3.9 | -15.0 ±4.0 |
| $\chi^2$ | 2.55 | 4.40 | 5.62 | 6.55 | 5.55 | 4.13 | 3.89 | 2.82 |

Table SI.1. *Structural nanodisc parameters refined from 8 SAXS datasets independently (csMSP1E3D1).*

*Symbols are described in Section 2.4.1*

4

|  | Global fit | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_L$ slope | -9.96 $\pm$2.67 | | | | | | | |
| $N_L$ intercept | 330 $\pm$13 | | | | | | | |
| $\epsilon$ slope | 0.04 $\pm$0.06 | | | | | | | |
| $\epsilon$ intercept | 1.36 $\pm$0.22 | | | | | | | |
| $A_L$ [Å$^2$] | 49.4 $\pm$0.5 | | | | | | | |
| $V_L$ [Å$^3$] | 1077 $\pm$3 | | | | | | | |
| $V_P$ [Å$^3$] | 37720 $\pm$620 | | | | | | | |
| $R$ [Å] | 4.34 $\pm$0.19 | | | | | | | |
| $\chi^2$ | 7.50 | | | | | | | |
| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $b$ [$10^{-5}$ cm$^{-1}$] | -16.4 $\pm$8.2 | -22.6 $\pm$8.5 | -25.5 $\pm$8.6 | -24.5 $\pm$8.6 | -22.9 $\pm$8.5 | -21.4 $\pm$8.4 | -19.3 $\pm$8.3 | -18.9 $\pm$8.2 |

Table SI.2. *Structural nanodisc parameters refined from 8 SAXS datasets globally (csMSP1E3D1). Symbols*

*are described in Section 2.4.1*

5

*SI.3.2. Model parameters for csMSP1D1 nanodiscs* Tables SI.3 to SI.4 contain the

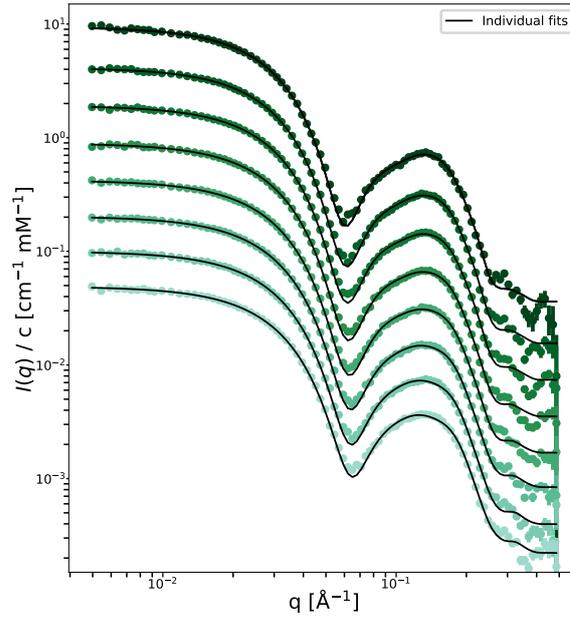parameters refined from the presented data and fits in Figure SI.5.



Fig. SI.4. Individual fits to experimental SAXS datasets from frames with increasing
   elution volumes/ positions across the SEC-peak. Datasets are the middle eight
   highlighted frames in Figure 2A. The topmost dataset is on absolute scale while
   those below are scaled by $2^{-n}$ where $n$ is the frame number.

Fig. SI.5. Model fit results for csMSP1D1 nanodiscs (A) Global fit to experimen-
tal SAXS datasets with increasing positions across the SEC-peak. Datasets are
the middle eight highlighted frames in Figure 2A. The bottom-most dataset is on
absolute scale while those above are scaled by $2^{-n}$ where $n$ is the frame number.
(B) Refined structural parameters. The colourful data points indicate parameters
refined from each dataset individually. The black lines indicate parameters refined
from the global fit, where one shared value is found for $A, v_P$ and $v_L$, while $N_L$ and
$\epsilon$ are both forced to follow a linear trend.

7

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Individual fits | | | | |
| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $N_L$ | 148 ±15 | 142 ±10 | 132 ±8 | 126 ±7 | 123 ±7 | 118 ±7 | 114 ±8 | 113 ±10 |
| $\epsilon$ | 1.60 ±0.28 | 1.63 ±0.21 | 1.65 ±0.17 | 1.68 ±0.16 | 1.67 ±0.16 | 1.74 ±0.18 | 1.72 ±0.21 | 1.79 ±0.27 |
| $A_L$ [Å$^2$] | 51.7 ±2.0 | 51.7 ±1.6 | 52.0 ±1.4 | 52.3 ±1.4 | 52.5 ±1.4 | 53.0 ±1.7 | 53.3 ±2.0 | 53.2 ±2.6 |
| $v_L$ [Å$^3$] | 1091 ±16 | 1092 ±12 | 1094 ±10 | 1097 ±10 | 1095 ±10 | 1099 ±11 | 1100 ±13 | 1099 ±17 |
| $v_p$ [Å$^3$] | 27470 ±1350 | 27050 ±980 | 27170 ±790 | 26980 ±690 | 27170 ±690 | 27020 ±730 | 27180 ±830 | 27320 ±1070 |
| $R$ [Å] | 4.35 ±0.46 | 4.18 ±0.36 | 4.15 ±0.30 | 4.13 ±0.27 | 4.13 ±0.28 | 4.23 ±0.30 | 4.24 ±0.36 | 4.43 ±0.44 |
| $b$ [10$^{-5}$ cm$^{-1}$] | -0.5 ±4.8 | -4.2 ±5.0 | -6.9 ±5.0 | -8.3 ±5.0 | -8.6 ±5.0 | -8.1 ±5.0 | -5.6 ±5.0 | -4.9 ±5.0 |
| $\chi^2$ | 1.62 | 2.36 | 2.15 | 2.40 | 2.00 | 1.86 | 1.55 | 1.75 |

Table SI.3. *Structural nanodisc parameters refined from 8 SAXS datasets independently (csMSP1D1).*

*Symbols are described in Section 2.4.1.*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Global fit | | | | |
| $N_L$ slope | | | | -5.00 ±2.37 | | | | |
| $N_L$ intercept | | | | 153 ±12 | | | | |
| $\epsilon$ slope | | | | 0.06 ±0.09 | | | | |
| $\epsilon$ intercept | | | | 1.36 ±0.38 | | | | |
| $A_L$ [Å$^2$] | | | | 51.3 ±1.4 | | | | |
| $V_L$ [Å$^3$] | | | | 1081 ±9 | | | | |
| $V_P$ [Å$^3$] | | | | 28290 ±730 | | | | |
| $R$ [Å] | | | | 4.40 ±0.33 | | | | |
| $\chi^2$ | | | | 5.38 | | | | |
| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $b$ [10$^{-5}$ cm$^{-1}$] | -0.8 ±9.5 | -5.5 ±9.8 | -10.7 ±9.9 | -13.4 ±10.0 | -13.3 ±10.0 | -11.0 ±10.0 | -7.0 ±9.9 | -4.0 ±9.7 |

Table SI.4. *Structural nanodisc parameters refined from 8 SAXS datasets globally (csMSP1D1). Symbols are*

*described in Section 2.4.1.*

8

*SI.3.3. Model parameters for csMSP1D1ΔH5 nanodiscs* Tables SI.5 to SI.6 contain
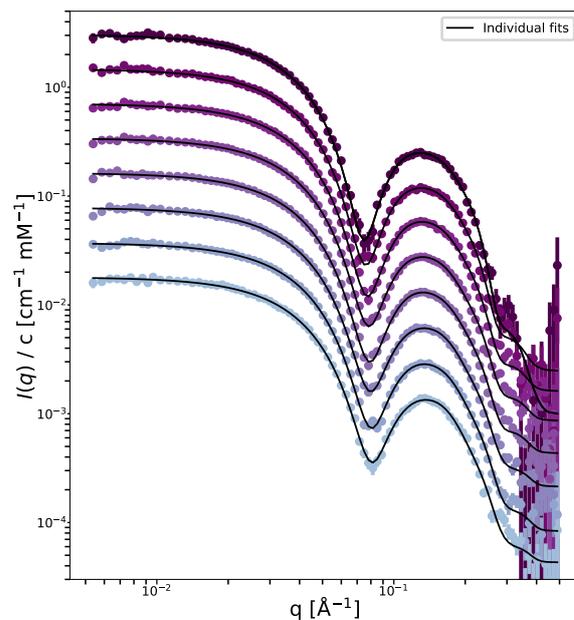the parameters refined from the presented data and fits in Figure SI.7.



Fig. SI.6. Individual fits to experimental SAXS datasets from frames with increasing
elution volumes/ positions across the SEC-peak. Datasets are the middle eight
highlighted frames in Figure **??**A. The topmost dataset is on absolute scale while
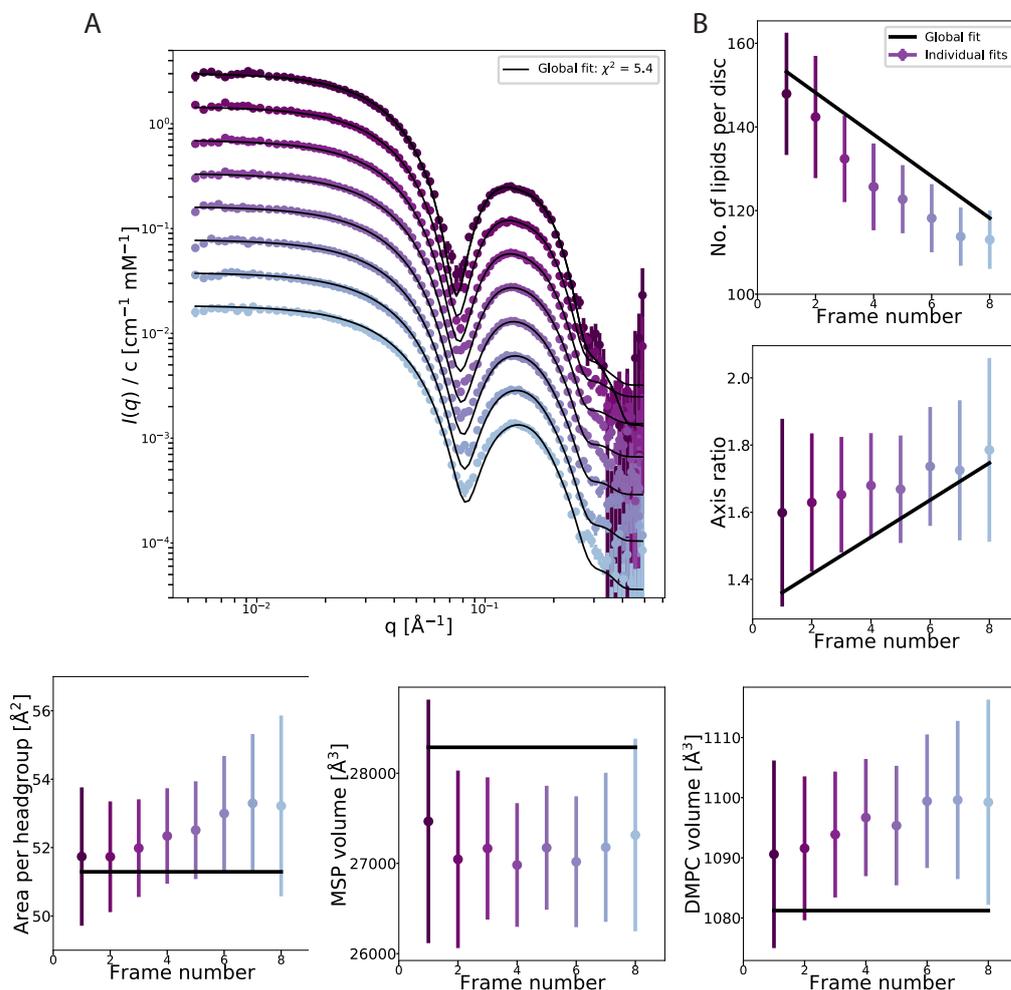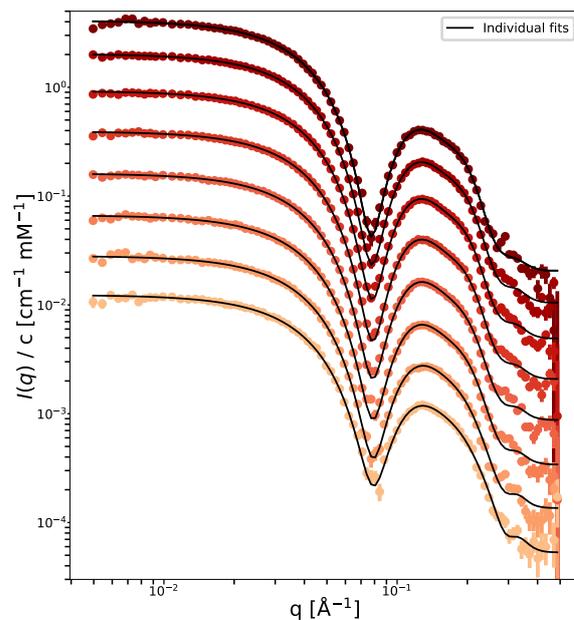those below are scaled by $2^{-n}$ where $n$ is the frame number.
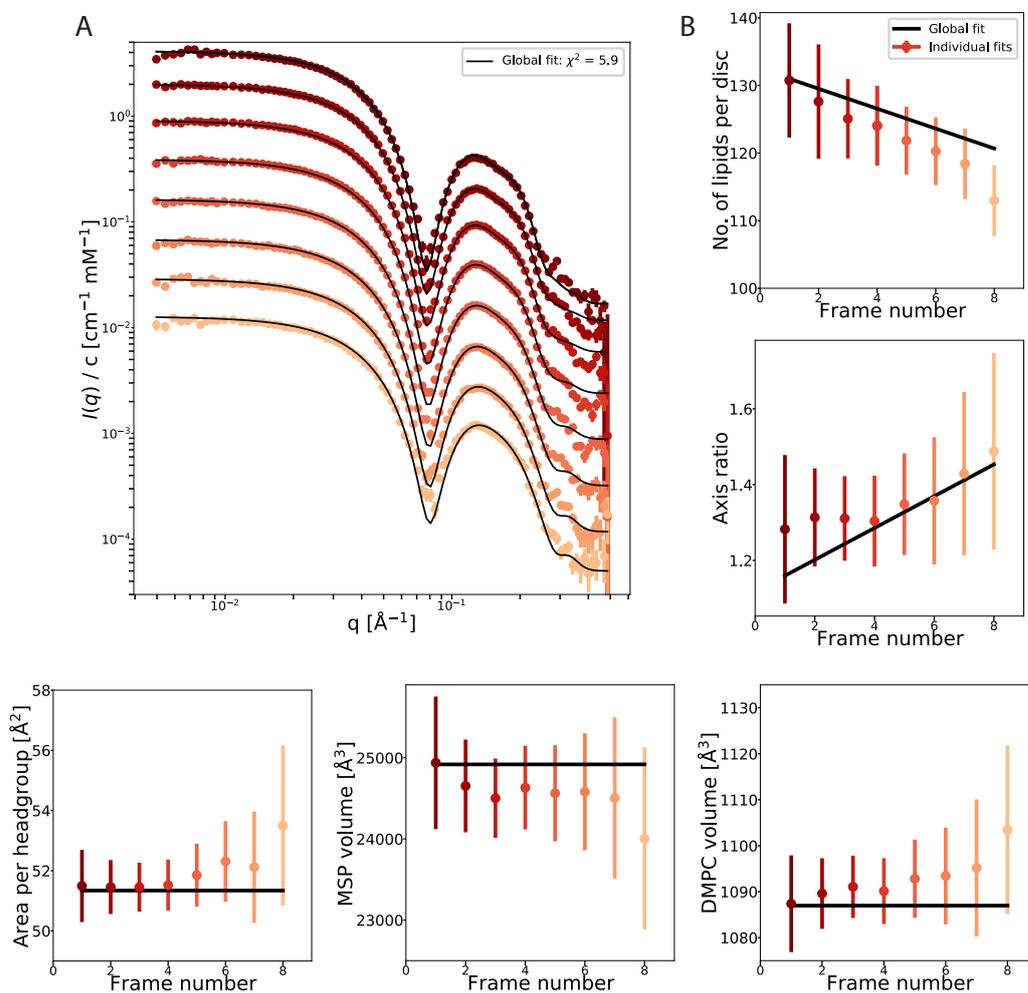
Fig. SI.7. Model fit results for csMSP1D1ΔH5 nanodiscs (A) Global fit to experimental SAXS datasets with increasing positions across the SEC-peak. Datasets are the middle eight highlighted frames in Figure 2 A. The bottom-most dataset is on absolute scale while those above are scaled by $2^{-n}$ where $n$ is the frame number. (B) Refined structural parameters. The colourful data points indicate parameters refined from each dataset individually. The black lines indicate parameters refined from the global fit, where one shared value is found for $A, v_P$ and $v_L$, while $N_L$ and $\epsilon$ are both forced to follow a linear trend.

10

| | Individual fits | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $N_L$ | 131 ±8 | 128 ±6 | 125 ±5 | 124 ±5 | 122 ±6 | 120 ±7 | 118 ±10 | 113 ±11 |
| $\epsilon$ | 1.28 ±0.20 | 1.31 ±0.13 | 1.31 ±0.11 | 1.30 ±0.12 | 1.35 ±0.13 | 1.36 ±0.17 | 1.43 ±0.22 | 1.49 ±0.26 |
| $A_L$ [Å$^2$] | 51.5 ±1.2 | 51.5 ±0.9 | 51.5 ±0.8 | 51.5 ±0.9 | 51.9 ±1.0 | 52.3 ±1.3 | 52.1 ±1.9 | 53.5 ±2.7 |
| $v_L$ [Å$^3$] | 1087 ±11 | 1090 ±8 | 1091 ±7 | 1090 ±7 | 1093 ±9 | 1093 ±11 | 1095 ±15 | 1103 ±18 |
| $v_p$ [Å$^3$] | 24940 ±820 | 24650 ±570 | 24500 ±490 | 24630 ±510 | 24560 ±590 | 24580 ±720 | 24500 ±1000 | 24000 ±1120 |
| $R$ [Å] | 4.30 ±0.29 | 4.09 ±0.21 | 4.06 ±0.19 | 4.06 ±0.20 | 4.15 ±0.23 | 4.16 ±0.28 | 4.23 ±0.37 | 4.14 ±0.49 |
| $b$ [10$^{-5}$ cm$^{-1}$] | -13.1 ±4.3 | -18.1 ±4.4 | -21.0 ±4.5 | -20.1 ±4.5 | -17.5 ±4.4 | -13.0 ±4.4 | -9.2 ±4.4 | -6.0 ±4.4 |
| $\chi^2$ | 2.29 | 2.16 | 2.80 | 2.74 | 2.67 | 2.03 | 1.89 | 1.37 |

Table SI.5. *Structural nanodisc parameters refined from 8 SAXS datasets independently (csMSP1D1ΔH5).*

*Symbols are described in Section 2.4.1.*

| | Global fit | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_L$ slope | -1.47 ±1.59 | | | | | | | |
| $N_L$ intercept | 131 ±9 | | | | | | | |
| $\epsilon$ slope | 0.04 ±0.08 | | | | | | | |
| $\epsilon$ intercept | 1.16 ±0.30 | | | | | | | |
| $A_L$ [Å$^2$] | 51.3 ±1.1 | | | | | | | |
| $V_L$ [Å$^3$] | 1087 ±9 | | | | | | | |
| $V_P$ [Å$^3$] | 24920 ±690 | | | | | | | |
| $R$ [Å] | 4.19 ±0.26 | | | | | | | |
| $\chi^2$ | 5.85 | | | | | | | |
| Frame number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $b$ [10$^{-5}$ cm$^{-1}$] | -10.6 ±8.5 | -20.3 ±8.9 | -25.4 ±9.1 | -13.2 ±9.1 | -17.6 ±8.7 | -12.2 ±8.6 | -7.9 ±8.3 | -5.6 ±8.2 |

Table SI.6. *Structural nanodisc parameters refined from 8 SAXS datasets globally (csMSP1D1ΔH5). Symbols*

*are described in Section 2.4.1.*

11

*SI.4. Scattering lengths*

The scattering lengths of the various molecules in our samples were assigned the scattering lengths in Table SI.7 during the refinement processes.

| Component | Chemical composition | X-ray scattering length [cm] |
|---|---|---|
| DMPC headgroups | $C_{10}H_{18}NO_8P$ | $4.62 \cdot 10^{-11}$ |
| DMPC tailgroups | $C_{24}H_{48}$ | $5.40 \cdot 10^{-11}$ |
| DMPC methyl | $C_2H_6$ | $5.08 \cdot 10^{-12}$ |
| csMSP1D1$\Delta$H5 | $C_{895}H_{1416}N_{244}O_{299}S_3$ | $3.08 \cdot 10^{-9}$ |
| csMSP1D1 | $C_{1002}H_{1593}N_{275}O_{333}S_3$ | $3.45 \cdot 10^{-9}$ |
| csMSP1E3D1 | $C_{1347}H_{2137}N_{373}O_{442}S_5$ | $4.64 \cdot 10^{-9}$ |
| Solvent | $H_2O$ | $2.82 \cdot 10^{-12}$ |

Table SI.7. *Chemical compositions and X-ray scattering lengths for the required components in the nanodisc modelling*

# 8

## ORDER AND DISORDER - AN INTEGRATIVE MODEL FOR THE FULL-LENGTH HUMAN GROWTH HORMONE RECEPTOR

### CONTRIBUTIONS

The growth hormone receptor is described as a 'method orphan' in structural biology because of its small size and high degree of structural disorder. In this study a plethora of experimental techniques were integrated into a molecular dynamics simulation and generate a structural ensemble of the protein.

The overall study and experiments were led by Noah Kassem and Birthe Kragelund. Raul Araya-Secchi performed all of the simulations. I processed and analysed the SAXS and SANS data presented in the article. I developed software in *WillItFit* to build a novel model for the full-length protein in a nanodisc, which was refined against SAXS data. I also helped to integrate the simulated trajectory with the SAXS data. I was mainly supervised by Lise Arleth and Martin Cramer Pedersen.

SCIENCE ADVANCES | RESEARCH ARTICLE

**BIOPHYSICS**

# Order and disorder—An integrative structure of the full-length human growth hormone receptor

Noah Kassem[1†], Raul Araya-Secchi[2†], Katrine Bugge[1], Abigail Barclay[2], Helena Steinocher[1], Adree Khondker[3], Yong Wang[1], Aneta J. Lenard[1], Jochen Bürck[4], Cagla Sahin[5], Anne S. Ulrich[4], Michael Landreh[5], Martin Cramer Pedersen[2], Maikel C. Rheinstädter[3], Per Amstrup Pedersen[6], Kresten Lindorff-Larsen[1]*, Lise Arleth[2]*, Birthe B. Kragelund[1]*

Because of its small size (70 kilodalton) and large content of structural disorder (>50%), the human growth hormone receptor (hGHR) falls between the cracks of conventional high-resolution structural biology methods. Here, we study the structure of the full-length hGHR in nanodiscs with small-angle x-ray scattering (SAXS) as the foundation. We develop an approach that combines SAXS, x-ray diffraction, and NMR spectroscopy data obtained on individual domains and integrate these through molecular dynamics simulations to interpret SAXS data on the full-length hGHR in nanodiscs. The hGHR domains reorient freely, resulting in a broad structural ensemble, emphasizing the need to take an ensemble view on signaling of relevance to disease states. The structure provides the first experimental model of any full-length cytokine receptor in a lipid membrane and exemplifies how integrating experimental data from several techniques computationally may access structures of membrane proteins with long, disordered regions, a widespread phenomenon in biology.

## INTRODUCTION

The human growth hormone receptor (hGHR) is ubiquitously expressed and is activated by human growth hormone (hGH), produced in the pituitary gland. hGHR is important for regulating growth at a cellular and systemic level (*1*, *2*) and is involved in the regulation of hepatic metabolism, cardiac function, bone turnover, and the immune system (*3*). Besides direct promotion of growth, its ligand hGH can also indirectly regulate growth by initiating the synthesis of insulin-like growth factor-I, an important factor in postnatal growth (*4*). Excess hGH production and mutations in the hGHR gene manifest in different diseases including cancer (*5*) and growth deficiencies (*6*), with associated cardiovascular, metabolic, and respiratory difficulties, and both hGH-based agonists and antagonists of the receptor exist as approved drugs (*7*).

The hGHR is 1 of ~40 receptors belonging to the class 1 cytokine receptor family. The family is topologically similar with a tripartite structure consisting of a folded extracellular domain (ECD), a single-pass transmembrane domain (TMD), and a disordered intracellular domain (ICD) (*8*, *9*). A characteristic trait of these receptors is the lack of intrinsic kinase activity, with the ICD, instead, forming a binding platform for a variety of signaling kinases and regulatory proteins (*8*, *10*), as well as of certain membrane lipids (Fig. 1A) (*9*). Within the ECD, the receptors share a characteristic cytokine receptor homology domain consisting of two fibronectin type III domains (D1, N-terminal and D2, C-terminal), each with a seven-stranded β-sandwich structure. Two hallmark disulfide bonds and a conserved WSXWS motif (X is any amino acid) (*11*) located in D1 and D2, respectively, are suggested to be important for cell surface localization and discrimination between signaling pathways (*11*, *12*). In hGHR, this motif is instead YGEFS (*10*), but the reason for this variation has remained enigmatic. Besides hGHR, group 1 of the class 1 cytokine receptors also encompasses the prolactin receptor (PRLR) and the erythropoietin (EPO) receptor. The group is considered to be the most structurally simple with one cytokine receptor homology domain and ligand binding in a homodimeric 2:1 complex (*10*).

Receptor activation is achieved by hGH binding to hGHR via two asymmetric binding sites, leading to structural rearrangements that are propagated through the TMD to the ICD. A recent study found that when hGH binds to a preformed hGHR dimer, structural rearrangements in the ECD leads to separation of the ICDs just below the TMD (*13*). This leads to activation through cross-phosphorylation of the Janus kinases 2 (JAK2) bound at the proline-rich Box1 motif in the juxtamembrane region (*13*). Furthermore, this study demonstrated that receptor dimerization in isolation is insufficient for receptor activation (*13*). Nonetheless, while recent single-particle tracking studies suggested dimerization to depend on expression levels (*14*), it is still debated to what extent the hGHR exists as preformed dimers in vivo (*15*) or whether the hGHR only dimerizes upon hGH binding.

From the viewpoint of structural biology, the hGH/hGHR system has a high molecular complexity with ordered and disordered domains joined by a minimal membrane embedded part. Hence, structural characterization has, so far, been based on a divide-and-conquer approach, where the domains have been studied in isolation. This includes the crystal structures of the ECD in the monomeric state (*15*), in 1:1 (*16*) and 2:1 (*17*) complexes with hGH, and of hGH alone (*18*). Furthermore, structures of the dimeric state of the hGHR-TMD in detergent micelles have been solved by nuclear magnetic resonance (NMR) (*19*) spectroscopy, while the hGHR-ICD was shown by NMR to adopt a fully intrinsically disordered region (IDR) (*9*). A recent approach that combined experimental data with computational
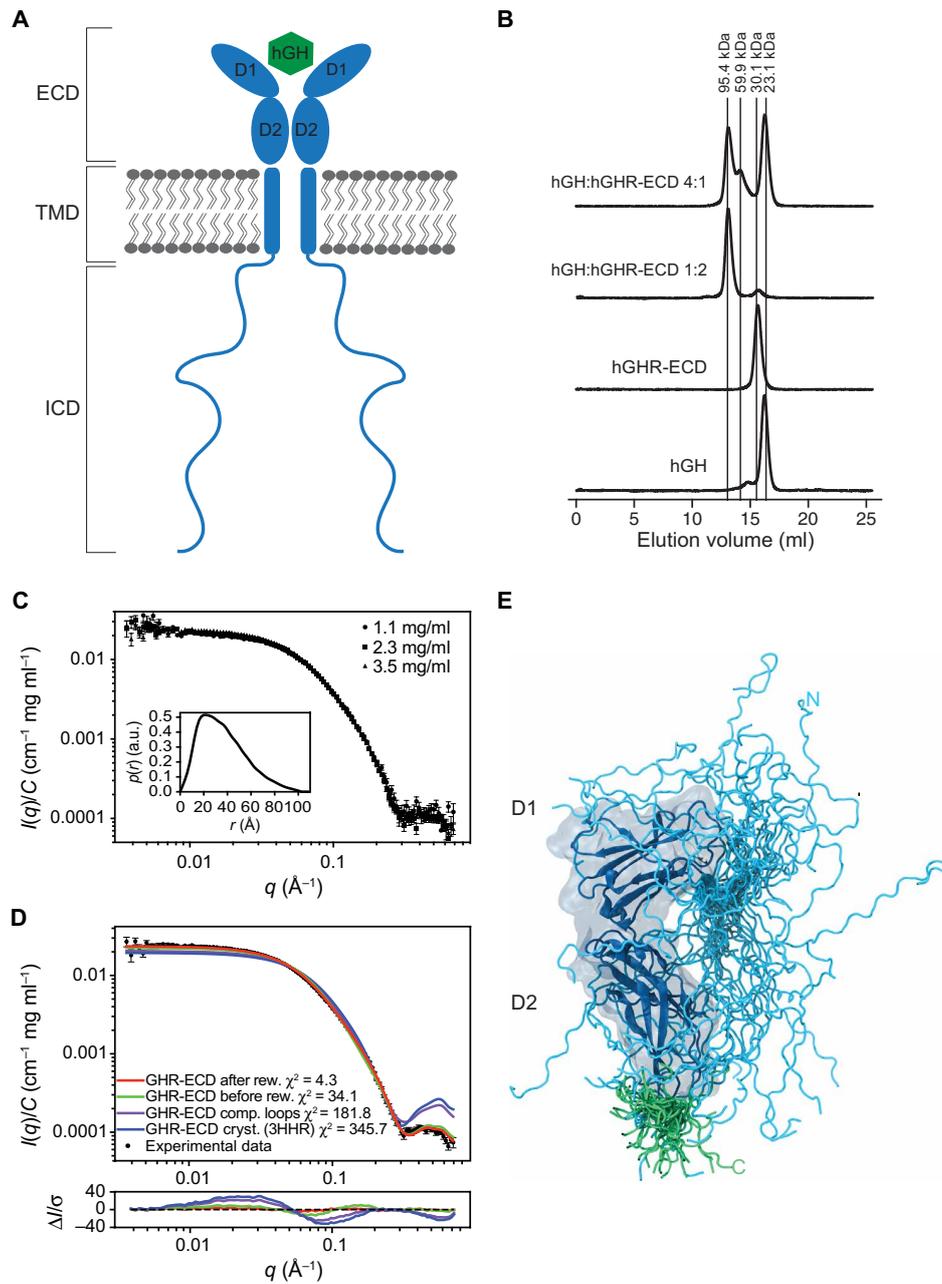
**Fig. 1. The hGHR has a dynamic ECD with a broad structural ensemble.** (**A**) A schematic representation of homodimeric hGHR (blue) in the membrane in complex with hGH (green). ECD, Extracellular domain; TMD, transmembrane domain; and ICD, intracellular domain. (**B**) SEC profiles of hGHR-ECD and hGH in 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4) and 150 mM NaCl at ratios 1:0 (hGH:hGHRECD 1:0), 0:1 (hGH:hGHR-ECD 0:1), 1:2 (hGH:hGHR-ECD 1:2), and 4:1 (hGH:hGHR-ECD 4:1). Absorption was measured at 280 nm. (**C**) Concentration-normalized SAXS data from hGHR-ECD (concentrations in legend) with the $p(r)$ from the sample (3.5 mg/ml) shown as inset. a.u., absorbance units. (**D**) SAXS data from hGHR-ECD at 3.5 mg/ml (black dots) together with fits of the theoretical scattering curves from a crystal structure of hGRH-ECD (blue; PDB 3HHR), the same crystal structure with missing loops completed (purple), and the average (green) and reweighted average [red; reweighted against the experimental data using the Bayesian maximum entropy approach (see Materials and Methods)] of scattering curves of the 500 hGHR-ECD models with added N- and C-terminal tails. Residuals are plotted below. (**E**) An ensemble model of the hGHR-ECD with a representative reweighted subensemble of 100 models highlighting the N-terminal (cyan) and C-terminal (green) dynamic tails.

efforts provided a model of the similar PRLR monomer built from integration of several individual sets of experimental data recorded on isolated domains (*20*). This work provided the view of a full-length class 1 cytokine receptor to scale. However, no structure or model based on data collected on an intact, full-length class I cytokine receptor exists, leaving a blind spot for how the domains affect each other and are spatially organized.

Even with advances in cryo–electron microscopy (EM), the full-length hGHR remains a challenge to structural biology. With 70 kDa, the receptor is a small target for cryo-EM, but adding to this, the

## SCIENCE ADVANCES | RESEARCH ARTICLE

fact that more than 50% of the protein is intrinsically disordered leaves only ~30 kDa visible. Likewise, the disorder of the ICD also hampers crystallographic studies. On the other hand, a 70-kDa protein together with membrane mimetics make up a too large target for NMR, where the combined molecular properties would lead to slow tumbling and severe line broadening. Hence, the hGHR appears to be an orphan in structural biology, along with a large group of other membrane proteins with long, disordered regions, including most of the ~1400 human single-pass membrane proteins (21).

Integrative structural modeling is a powerful approach to combine different sources of experimental information to study the structure and dynamics of biomolecules (22). Such approaches are particularly suitable for large and dynamic molecules whose structure cannot be solved by traditional means. When experimental information comes from lower-resolution techniques, such as solution small-angle x-ray and neutron scattering (SAXS/SANS), it is often necessary to combine the experimental information with molecular simulations or other modeling techniques (22, 23). The best choice of experiments and computational models depends on the system at hand, including its size and level of flexibility. For large and highly flexible systems, coarse-grained (CG) simulations provide a suitable balance between the ability to sample many different types of conformations and capturing key aspects of protein biophysics (23, 24). Recent advances building on the use of nanodiscs (25) have further proved its applicability in membrane protein structural biology when combined with computational modeling (26, 27). However, no membrane protein with the degree of disorder seen in hGHR has previously been studied in a nanodisc or approached by small-angle scattering.

Here, we applied an integrative approach to probe the structure of the monomeric hGHR from SAXS data recorded on the full-length receptor in a nanodisc. The data were validated and interpreted by combining SAXS, NMR, and x-ray diffraction (XRD) data obtained on the individual domains of hGHR through computational modeling. This has resulted in an experimentally supported structure based on studies of an intact, full-length, single-pass cytokine receptor in a lipid membrane, a topology that represents ~40 human cytokine receptors and many other membrane proteins. Our approach exemplifies that combining SAXS and computational modeling could be the bridge required for accessing structural information on the ~1400 single-pass receptors in humans (28).

### RESULTS

To arrive at the final result of this work, we took on a three-step approach. First, to aid the analysis of SAXS data on the full-length hGHR and qualify the integrity of the methodology, several different biophysical data were acquired and analyzed on isolated, individual parts of the hGHR. Second, SAXS and SANS data were acquired on the full-length hGHR in nanodiscs, expressed in yeast cells and carrying a C-terminal green fluorescent protein (GFP)–deca-histidine tag (GFP-$H_{10}$). Last, all the data were interpreted and integrated using molecular dynamics (MD) simulations.

### The binding-competent hGHR-ECD solution state ensemble contains disorder

While crystal structures of an N- and C-terminally truncated version of the hGHR-ECD exist (16–18), the complete hGHR-ECD has not previously been studied in solution. Therefore, to describe the ensemble of the full domain, we purified hGHR-ECD (residues 1 to

245, omitting the signal peptide) and hGH from expression in *Escherichia coli*. On the basis of circular dichroism (CD) data, the hGH was folded with the expected helicity (fig. S1A). The CD spectrum of hGHR-ECD had pronounced positive ellipticities around 230 nm stemming from aromatic exciton couplings, a trait of cytokine receptors, and showed additional contributions from disorder at 200 nm (fig. S1B). The functionality of the hGHR-ECD was confirmed from its ability to form complexes as determined by $K_{av}$ for hGH and its 1:1 and 1:2 complexes with hGHR-ECD from analytical size exclusion chromatography (SEC) (Fig. 1B and fig. S1, C and D). By varying the ratio of hGH to hGHR-ECD, we could isolate the 1:2 complex and the 1:1 complex (GH in four times excess) and obtain the mass of hGH, hGHR-ECD, and the 1:1 and 1:2 complexes using the forward scattering from SAXS approximated through the derived pair-distance distribution functions, $p(r)$'s (table S1 and fig. S1, E to G). Thus, the structural and functional integrity of both renatured proteins were confirmed. Last, to understand the ensemble properties of the hGHR-ECD in solution and generate a model, we acquired SAXS data on free hGHR-ECD at varying concentrations. The concentration-normalized SAXS data fully overlaid (Fig. 1C), showing no visible interaction effects. The derived $p(r)$ (Fig. 1C, inset) was skewed with a broad maximum around 30 Å and a maximum length ($D_{max}$) of ~100 Å, consistent with the hGHR-ECD having a nonglobular shape. Comparison of the SAXS data to a theoretical scattering profile obtained from one of the structures of hGHR-ECD (PDB 3HHR) (17) resulted in a poor fit (Fig. 1D, blue). However, the absence of the N- (1 to 30) and C-terminal (231 to 245) tails and two disordered loops (57 to 61; 74 to 77) may account for this and highlights that these tails contribute importantly to the ensemble properties. We therefore built a model of the hGHR-ECD, where the missing loops were added. The calculated scattering profile of this model provided a slightly improved fit to the SAXS data, which further suggested that a substantial contribution to the scattering comes from the conformational heterogeneity of the N- and C-terminal tails. To address this issue, we built an ensemble of 5000 models of the full-length hGHR-ECD including the N- and C-terminal tails using the "Floppy tail" (29) protocol from Rosetta (30). An average of the theoretical scattering intensities from these models was fitted to the experimental SAXS data (Fig. 1D, green) with $\chi^2$ of 34, about five times smaller than when using the crystal structure and a single conformation with completed flexible loops. Thus, an ensemble representation including the terminal tails provided a much better fit to the SAXS data. The ensemble was further refined by reweighting against the experimental data using the Bayesian maximum entropy (BME) approach (31), which brought $\chi^2$ to 4.3 using effectively 27% of the models (Fig. 1D, red). The $R_g$ distributions of the models before and after reweighting are shown in fig. S1H. A subensemble of 500 conformations, representative of the reweighted ensemble, was generated. Although any of these conformations would be a good representative, we selected a single structure with an $R_g$ most similar to the experimentally determined $R_g$ for building the model of the full-length hGHR (see below). A total of 100 conformations of this subensemble are shown in Fig. 1E, illustrating how the disordered regions contribute to the space-filling properties of the hGHR-ECD.

### The monomeric hGHR-TMD is organized parallel to the membrane normal

Structures of hGHR-TMD were recently solved in dimeric states (19) in micelles of the detergent $d_{38}$-dodecylphosphocholine (DPC).

To describe the structure and the tilt angle of the monomeric hGHR-TMD relative to the membrane, we designed this domain of hGHR with six- and five-residue overlap with hGHR-ECD and hGHR-ICD, respectively. The resulting 36-residue hGHR-TMD (F239-R274), including an N-terminal Gly-Ser, was produced with and without isotope labeling by a fast-track production method for single-pass TMDs (*32*). Subsequently, the peptides were reconstituted in either lipid bilayers (see below) or 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) micelles, previously used successfully for structure determination of the closely related hPRLR-TMD (*20*).

A schematic overview of the extent of the hGHR-TMD α helix determined by NMR spectroscopy or bioinformatics is shown in Fig. 2A. To compare the structural characteristics of this hGHR-TMD

with the previously published structures (*19*), we analyzed isotope-labeled hGHR-TMD in DHPC micelles by NMR and CD spectroscopy (Fig. 2B and fig. S2, A and B). From motif identification from chemical shifts (MICS) analysis (*33*) of NMR backbone chemical shifts and from backbone amide $R_2$ relaxation measurements, we observed that the hGHR-TMD populated a fully formed α helix in DHPC micelles from W249-K271 (Fig. 2B). This agrees with the findings for hGHR-TMD dimers in DPC micelles (*19*), suggesting the length of the TMD α helix to be maintained across different membrane mimetics. For building the full-length hGHR model (see below), we used the backbone chemical shift–derived dihedral angles (from TALOS) or the chemical shifts directly to calculate a low-resolution structure of the TMD, either by CYANA (*34*) or
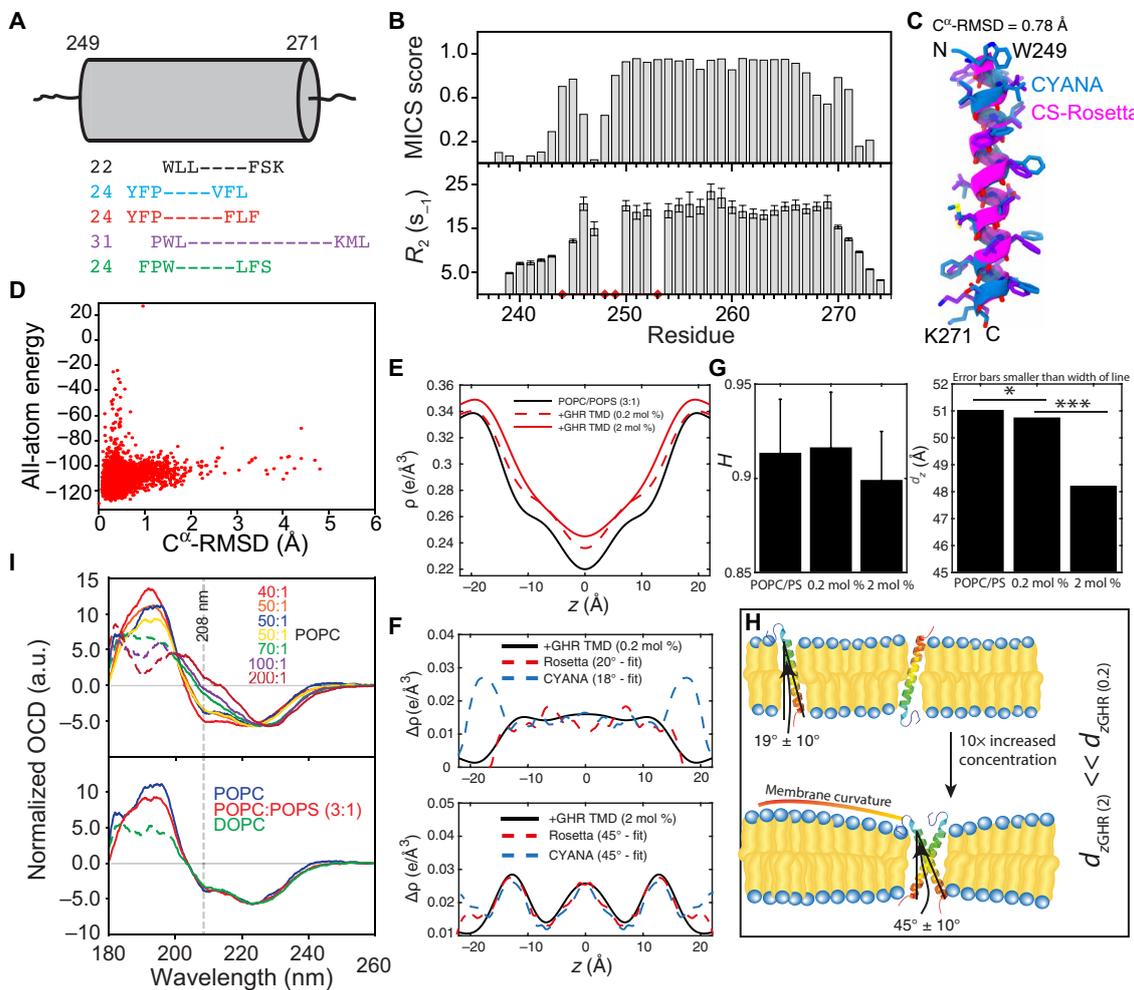


**Fig. 2. The single-pass α-helical hGHR-TMD.** (**A**) Extent of the hGHR-TMD α helix by NMR (black), TMHMM (blue), Phobius (red), METSAT-SVM (purple), and UniProt (green) (fig. S2). Cylinder represents length of the hGHR-TMD α helix by NMR with the first and last helical residue numbered. (Left) Predicted numbers. (**B**) MICS α helix probability (top) and $R_2$ relaxation rates (bottom) of hGHR-TMD in DHPC micelles. Red diamonds, insufficient data quality or prolines. (**C**) Models of hGHR-TMD α helix by CYANA (blue) or CS-Rosetta (magenta). (**D**) Energy versus $C^\alpha$-RMSD from CS-Rosetta modeling of hGHR-TMD. (**E**) Electron density profiles (EDPs) of lipid bilayers [POPC:POPS 3:1 mole percent (mol %)] with hGHR-TMD at 0.2 and 2 mol %, respectively. (**F**) Difference EDPs and best-fit profiles for the hGHR-TMD CS-Rosetta and the CYANA models, respectively. (**G**) Hermans orientation (left) and Lamellar spacing (right) of membranes at varying concentrations of hGHR-TMD. *, onefold change; ***, threefold change. (**H**) Illustration of membrane curvature. (**I**) Top: Oriented CD (OCD) spectra of 6 μg of hGHR-TMD in POPC, with L:P ratios varied from 1:40 to 1:200. Bottom: OCD spectra of 6 μg of hGHR-TMD in POPC, POPC:POPS (3:1), or DOPC at an L:P ratio of 50:1. The dashed lines represent nonreliable data due to too high high-tension voltage values.

CS-Rosetta (*35*), respectively. The helix covered W249-K271 in both structures, which aligned with a root mean square deviation (RMSD) of the $C^\alpha$s of 0.78 Å (Fig. 2, C and D).

To support the modeling, we reconstituted the hGHR-TMD in a more native-like membrane system of stacked bilayers. We used 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC):1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine (POPS) (3:1 molar ratio) and investigated structures and tilt angles by XRD, taking advantage of the oligomeric state of single-pass TMDs that may be manipulated through the detergent-to-protein or lipid-to-protein (L:P) ratio (*19*). The measured reflectivity Bragg peaks allowed us to determine the electron density profiles (EDPs), $\rho(z)$, of the different bilayer structures (Fig. 2E) and EDP difference plots, $\Delta\rho(z)$ (Fig. 2F), of the membranes with and without inserted hGHR-TMD helices at varying molar ratios. The EDPs contain information about the position in the membrane and tilt angle. The EDPs of the helices were calculated for the monomer based on the CS-Rosetta and CYANA structures, respectively, and for the dimer based on a previously determined NMR structure [PDB 5OEK (*19*)]. This was done for different tilt angles and fitted to the experimental densities (*36*).

Our XRD analysis showed that at monomer conditions for the hGHR-TMD (high L:P ratio of 500:1; Fig. 2F, top), the helix remained parallel to the membrane normal (tilt angle of 19° ± 10°) without effects on membrane thickness, $d_z$. The discrepancy between fit and experimental data at *z* values of ~18 Å (in the head group region of the membranes) for the CYANA model in 0.2 mole percent (mol %) results from the fact that the CYANA model includes the unstructured tail (which is not part of the Rosetta model). This tail partitions into the water phase and is not well picked up in the EDP structure. Within the membrane bilayer, the fitting to the CYANA model was, therefore, slightly better. At dimer conditions (low L:P ratio of 50:1), we found that the helix tilt angle changed to 45° ± 10° relative to the membrane normal, in accordance with the hGHR-TMD dimer structures (Fig. 2F, bottom) (*19*). While the membrane flatness and intactness, as measured by Hermans orientation function *H*, was unaffected by the presence of monomers or dimers (Fig. 2G, left), the dimer induced some membrane compression, giving rise to a slightly thinner bilayer with smaller laminar spacing, $d_z$ (Fig. 2G, right). An illustration of this behavior is shown in Fig. 2H.

To support these observations, we used oriented CD (OCD) spectroscopy with reconstitution of the hGHR-TMD in POPC, POPC:POPS (3:1), or 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) multilamellar bilayers (Fig. 2I and fig. S2C). In OCD, the ellipticity of the negative band at 208 nm, which is parallelly polarized to the helix axis, is strongly dependent on helix orientation, allowing distinction between a fully inserted state (I state, parallel to membrane normal), a tilted state (T state), and a surface-bound state (S state, perpendicular to the membrane normal). At dimer conditions (L:P ratio of 50:1), the OCD spectra showed two negative bands at 208 and 222 nm and a positive band at 190 nm in all types of membranes tested (Fig. 2I), indicating successful reconstitution with formation of helical structure. Furthermore, the negative ellipticity at 208 nm was smaller compared to that at 222 nm, demonstrating the hGHR-TMD to be either in a T state or in an equilibrium between an S state and an I state (*37*). Increasing the L:P ratio decreased the negative band intensity at 208 nm, which even became positive at an L:P ratio of 200:1 (Fig. 2I, top). This indicated that at monomer conditions, the hGHR-TMD populated the more parallel I state, fully supporting the results from XRD.

## A C-terminal GFP has no influence on the ICD ensemble

For purification of the full-length hGHR, a C-terminal GFP-$H_{10}$-tag had to be included (*21*). To ensure that this did not introduce intra- or intermolecular interactions interfering with the hGHR-ICD ensemble, we produced the hGHR-ICD (S270-P620) without and with GFP-$H_{10}$ (hGHR-ICD-GFP-$H_{10}$). The $^{15}$N–heteronuclear single-quantum coherence (HSQC) spectra of these two proteins were almost identical (Fig. 3A), confirming an unperturbed ICD ensemble. We also compared SAXS data acquired on both, which revealed a large increase in the forward scattering in the presence of GFP (Fig. 3B), reflecting the increase of the molar mass from 38.6 kDa for hGHR-ICD to 68.0 kDa for hGHR-ICD-GFP-$H_{10}$ (table S1). The derived $p(r)$ functions (Fig. 3D) showed not only an increased probability of short distances because of the folded GFP but also a conserved $D_{max}$ consistent with an overall unaffected ICD coil conformation. The addition of GFP did not give rise to a significant change in $R_g$ (65 Å for both) (Fig. 3B), whereas the hydrodynamic radius ($R_h$) obtained by NMR spectroscopy increased from 44 to 51 Å (Fig. 3C). We note that $R_g/R_h$ of ~1.5 for the hGHR-ICD falls in the range typically observed for linear chains in random coil conformations, while the smaller ratio obtained for the hGHR-ICD-GFP-$H_{10}$ is consistent with the hGHR-ICD-GFP-$H_{10}$ containing a larger fraction of folded protein. Together, these results indicate that the C-terminal addition of GFP-$H_{10}$ did not change the structural ensemble of hGHR-ICD.

## Scaling of the protein-water interactions is required to simulate the ensemble properties of hGHR-ICD

To aid in interpretation of the data of the full-length hGHR, the ensemble properties of the hGHR-ICD were modeled on the basis of the SAXS data following two approaches: (i) through fitting of the data by the form factor for simple (non–self-avoiding) Gaussian random coils (*38*) and (ii) using CG-MD simulations adapted to better represent the dynamics of intrinsically disordered proteins (IDPs), providing an ensemble that better describes the experimental data. Approach (i) provided an excellent fit to the full experimental SAXS *q* range, yielding an $R_g$ of 68 ± 4 Å (Fig. 3B, orange) with a $\chi^2$ of 1.4. This showed the average conformation of the hGHR-ICD to be very well described by a simple random coil model, which implicitly assumes a scaling exponent, ν = 0.5. We also obtained similar $R_g$ values (~65 Å) from fits using different scaling exponents (0.588 to 0.602) empirically predicted for unfolded proteins or IDPs or derived from computational analyses (table S2). Hence, the values agree closely, and the effect of assuming a simple idealized Gaussian random coil model has a negligible effect on the resulting $R_g$.

Protein-protein interactions may be overestimated in the Martini force field translating into unrealistic compaction of disordered regions and inability to reproduce experimentally obtained values for $R_g$ or $R_h$ (*24*). Recent reports investigating two multidomain proteins connected by flexible linkers suggested that this could be overcome by increasing the strength of protein-water interactions (*24*). In the case of hGHR with a long, disordered ICD, we performed unbiased and enhanced sampling metadynamics simulations using the Martini 3 force field modified by increasing the strength of the protein-water interactions in the range of 5 to 15%. Our goal was to search for a value that could provide an optimized description of the ensemble of GHR-ICD. Back-mapped atomistic conformations from these simulations were used to calculate their average $R_g$ and
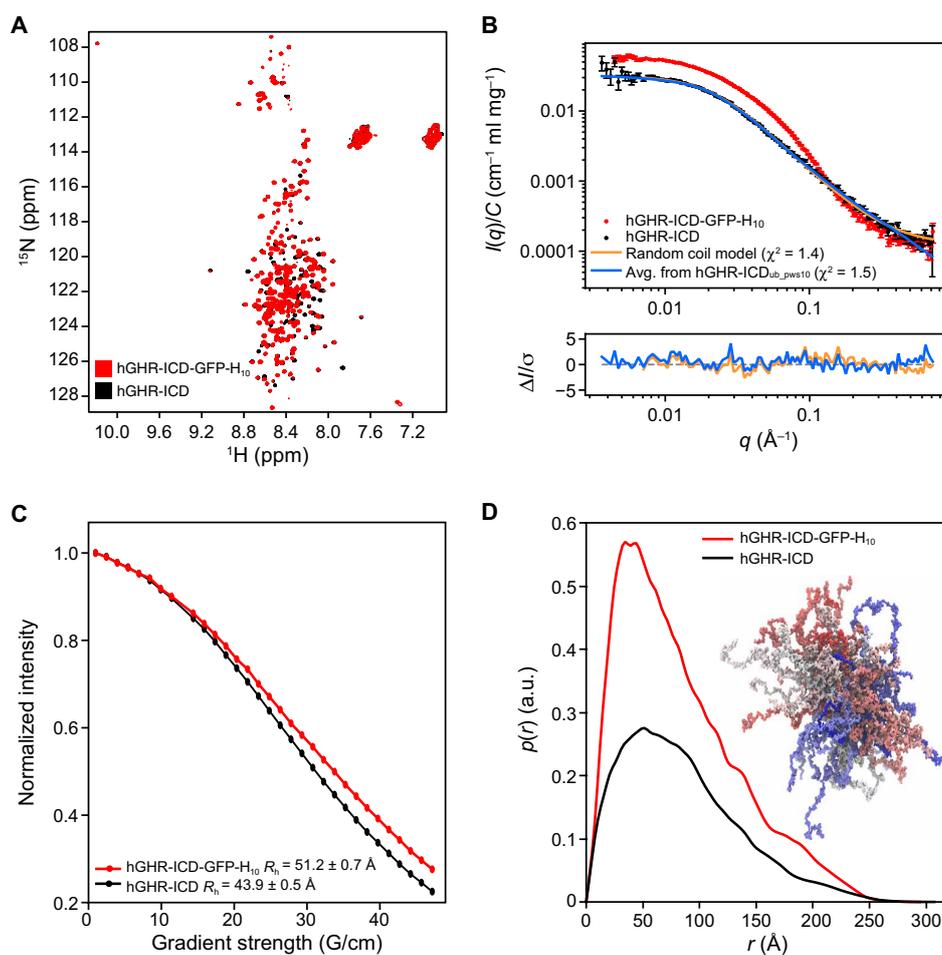
**Fig. 3. Properties of the hGHR-ICD ensemble.** (**A**) $^1$H-$^{15}$N-HSQC spectra at 5°C of hGHR-ICD (black) and hGHR-ICD-GFP-H$_{10}$ (red) at 150 and 100 μM, respectively. (**B**) Concentration normalized SAXS data from hGHR-ICD (black dots; 1.1 mg/ml) and hGHR-ICD-GFP-H$_{10}$ (red dots; 2.2 mg/ml). Fits to the data are shown for a Gaussian random coil model (orange) and from averaged scattering profiles from 5200 conformations taken from the hGHR-ICD$_{ub\_pws10}$ simulation (one every nanosecond) (blue). Residuals are plotted below. (**C**) $R_H$ of hGHR-ICD and hGHR-ICD-GFP-H$_{10}$ determined from pulse-field gradient NMR. Signal decays of hGHR-ICD (black) and hGHR-ICD-GFP-H$_{10}$ (red) are shown as a function of gradient strength together with the corresponding fits. (**D**) Concentration normalized $p(r)$'s derived from the above SAXS data from hGHR-ICD (black) and hGHR-ICD-GFP-H$_{10}$ (red). A subensemble of 200 conformations representative of the hGHR-ICD$_{metaD\_pws10}$ simulation is shown in the right side of the plot.

to obtain theoretical scattering intensities, which were then fitted to the SAXS data of hGHR-ICD (fig. S3). Our results indicate that an increase in the protein-water interaction strength of 10% produced optimal results (Fig. 3B and fig. S3). Thus, we settled on rescaling the protein-water interaction by 10% to obtain a reliable conformational ensemble of the hGHR-ICD and to be used in the simulation of the full-length hGHR-GFP system. We note that although lowly populated (<20%) transient helices were previously observed from NMR secondary chemical shifts (9), these are not to be captured by the CG simulations or the SAXS data.

**Full-length hGHR reconstituted in nanodiscs forms monomers and dimers**

The intact hGHR tagged with GFP-His$_{10}$ (hGHR-GFP) was expressed in the *Saccharomyces cerevisiae* strain PAP1500, purified, and reconstituted into POPC-loaded membrane scaffold protein 1D1 (MSP1D1) nanodiscs as described in Kassem *et al.* (*21*). Although an NMR solution

structure of the MSP1D1ΔH5 exists based on their reconstitution with a 1,2-dimyristoyl-sn-glycero-3-phosphocholine bilayer (*39*), we used the MSP1D1 nanodisc and POPC, as this combination is currently the most-applied and best-characterized carrier system by SAXS and SANS (*40–42*), making experimental handling and subsequent computation of the nanodisc-embedded full-length structure of hGHR more reliable. In SEC, the hGHR-GFP in MSP1D1 eluted over a broad peak from 10 to 14 ml (Fig. 4A). This suggested that the hGHR-GFP was reconstituted in the discs potentially as both monomers and dimers or as higher-order oligomers. To quantify the number of hGHR-GFP per disc, we performed an SDS–polyacrylamide gel electrophoresis (PAGE) analysis of hGHR-GFP and MSP1D1 standards along with hGHR-loaded MSP1D1 discs isolated from the SEC at different elution volumes (Fig. 4B). From gel quantifications of hGHR-GFP and MSP1D1, we found the ratio over the peak to vary from ~2 hGHR-GFP per disc (F1) to ~1 hGHR-GFP per disc (F3). Since reconstitution was conducted with a 10-time excess of discs to hGHR-GFP to minimize the
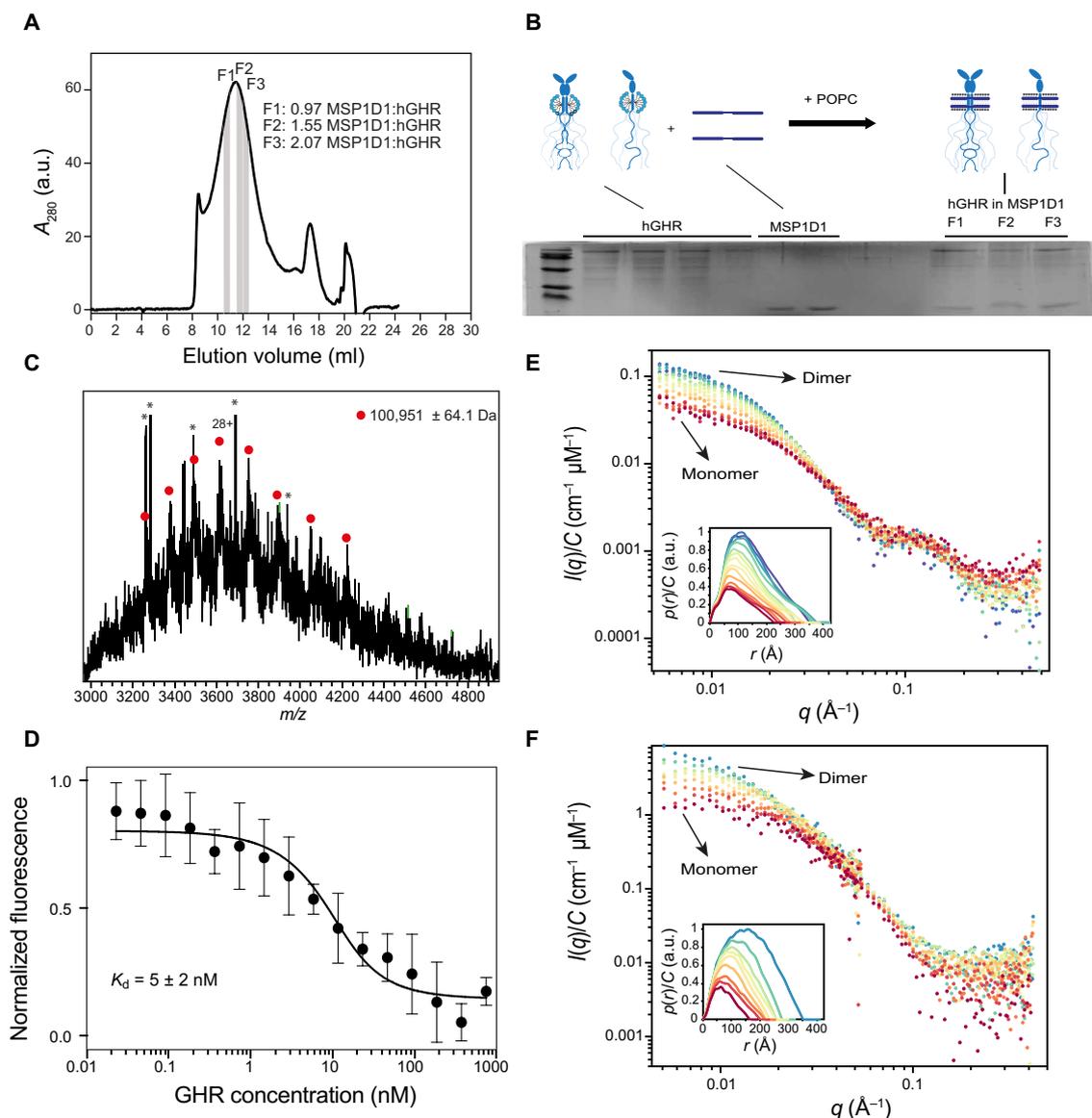
**Fig. 4. Incorporation of hGHR-GFP into MSP1D1, functional, and structural analysis.** (**A**) SEC profile of hGHR-GFP-loaded MSP1D1. The areas highlighted in gray indicate fractions (F1 to F3) used for the SDS-PAGE analysis in (B). (**B**) SDS-PAGE analysis of hGHR-GFP and MSP1D1 standards along with hGHR-GFP-loaded MSP1D1. Fractions F1 to F3 were taken from the indicated positions of the SEC-purified hGHR-GFP-loaded MSP1D1 shown in (A). The illustration above the gel shows the stoichiometry of the hGHR-GFP-loaded MSP1D1. (**C**) Intact mass spectra of hGHR-GFP show a single protein population with an average mass of $100,951 \pm 64$ Da. Asterisks denote detergent peaks. (**D**) MST determination of equilibrium binding constants for hGH to hGHR-GFP-loaded MSP1D1. The mean values and SD were obtained by fitting a 1:1 binding model (full line) as described in Materials and Methods. Concentration-normalized (**E**) SAXS data and (**F**) SANS data of the nanodisc-embedded hGHR-GFP corresponding to the highlighted SEC frames in fig. S4 (C and D).

probability of capturing more than one hGHR-GFP per disc, we argue that the distribution across the peak likely represents the equilibrium between dimeric and monomeric hGHR-GFP. These results also suggested that the hGHR-GFP can form dimers in the absence of hGH as previously suggested (*13*), most likely through the TMD region (*13*, *15*).

### The number of lipids in the hGHR-loaded MSP1D1 nanodiscs is as expected
We used phosphorus analysis (*43*) performed on samples across the SEC peak (fig. S4A) to quantify the number of POPC lipids in the

hGHR nanodiscs. In the fractions with dimers (F1), the ratio between MSP1D1 nanodiscs and POPC was $115 \pm 19$, and in the fraction with monomers (F3), it was $122 \pm 17$. The SD is based on two repetitive measurements each on two separate samples. This is comparable to results obtained in other studies of POPC nanodiscs with an α-helical membrane-anchored protein (*27*) and in good agreement with the values obtained for nanodiscs solely filled with POPC [~110 to 130 POPC per nanodisc (*41*)]. The number of lipids was used as input for the modeling of the SAXS data of hGHR-GFP-containing MSP1D1 nanodisc.

## hGHR retains its signal peptide but is not N-glycosylated when produced in yeast

The hGHR has five confirmed N-glycosylation sites at N28, N97, N138, N143, and N282 (44), whereas it is unknown if it is O-glycosylated. To assess whether the recombinant hGHR-GFP from *S. cerevisiae* was N-glycosylated, the electrophoretic mobility before and after treatment with endoglycosidase H was evaluated (fig. S4B). No mobility change was observed, and the band sharpness was equally high before and after treatment, suggesting the lack of N-glycosylations. This is in line with previous observations on other human membrane proteins produced in the same yeast expression system (45). To determine whether yeast-produced hGHR-GFP was O-glycosylated, we performed Western blot with horseradish peroxidase conjugated with concanavalin A, which binds to mannose residues in O-glycosylated proteins (45). A faint band corresponding to hGHR-GFP was seen, indicating minor O-glycosylation (fig. S4B). As a negative control, MSP1D1 purified from *E. coli* was not detected (fig. S4B). As a final control, we recorded native mass spectrometry (MS) on the purified, full-length hGHR-GFP in detergent, before reconstitution into nanodiscs. No glycosylation was detected. Instead, as expected from the cellular localization (21), the hGHR-GFP maintained its signal peptide (SP) (expected mass of 100,850.44 Da; the small deviation in mass caused by Na$^+$ adducts) (Fig. 4C).

## Recombinant full-length hGHR-GFP reconstituted in nanodiscs is fully binding competent

To ensure that full-length hGHR-GFP embedded in the MSP1D1 nanodisc was functional, we measured equilibrium binding constants for the interaction between hGH and hGHR-GFP(MSP1D1) by microscale thermophoresis (MST). In these studies, a 20 nM solution of fluorescently labeled (NT-647-NHS) hGH was incubated with increasing concentrations of hGHR-GFP(MSP1D1) (23 pM to 750 nM) using unlabeled hGH as the control. With this approach, the dissociation constant between hGH and hGHR-GFP(MSP1D1) was determined to be $K_d$ = 5 ± 2 nM (Fig. 4D). As another control, we previously showed that hGHR(MSP1D1) is unable to bind human prolactin (21), which cannot activate hGHR in vivo (46). The affinities of hGH for hGHR-ECD have previously been reported as 1.2 and 3.5 nM for the first and the second site of hGH, respectively (47). Taking all this into consideration, we find that our data agree well with previous findings and conclude that the nanodisc-reconstituted, yeast-produced full-length hGHR-GFP is fully binding competent.

## SEC-SAXS and SEC-SANS data of the full-length hGHR-GFP in nanodiscs

We obtained structural data of the reconstituted full-length hGHR-GFP in a POPC-loaded MSP1D1 nanodisc from SEC-SAXS (Fig. 4E and fig. S4C) and SEC-SANS (Fig. 4F and fig. S4D) with $p(r)$ functions in insets of Fig. 4 (E and F). As was the case for the initial analysis, the SEC profiles from both SEC-SAXS and SEC-SANS (fig. S4, C and D) were relatively broad and consistent with the underlying heterogeneity and systematic decrease of the particle size. Analysis of the data obtained over the SEC-SAXS and SEC-SANS elution peaks confirmed this picture, and SEC-SAXS showed $R_g$ decreasing from 120 to ~75 Å over the frames from 10 to 14 ml (fig. S4C). The SEC-SANS–derived $R_g$ also decreased steadily over the frames from 10 to 14 ml (fig. S4D) but were consistently ~10 Å smaller than in the SAXS experiment. The different $R_g$ values for SAXS and SANS are a

consequence of the nonhomogeneous and rather different excess scattering length density distributions of the nanodisc-embedded hGHR-GFP for x-rays and neutrons. The decrease in both the $R_g$, the low $q$ scattering intensity, and the development of the $p(r)$'s over the SEC peaks is fully consistent with the presence of discs containing the first two and then one hGHR-GFP, respectively, as also supported by our biochemical analysis (Fig. 4, A and B). In addition to dimerization, the large $R_g$ values obtained from the left side of the SEC peak could also arise from an overlap with the void volume (at 8 to 10 ml). We obtained a value of $D_{max}$ of ~200 to 250 Å from the SAXS data corresponding to discs with one hGHR. The low signal to noise at the right side of the SEC-SANS peak, corresponding to the monomeric fractions, did unfortunately not allow for a robust analysis of the SANS data. The dimeric fractions exhibited a significantly larger $D_{max}$ of ~350 Å in both SAXS and SANS. This larger size likely results from the larger extension of the two long uncorrelated ICDs. The shoulder around 0.1 Å$^{-1}$ of the SAXS data (Fig. 4E) is a typical signature of the lipid bilayer from the embedding nanodiscs (26, 27, 48).

Below, we focus our analyses on the structure of monomeric hGHR-GFP. First, while we can build an initial model of a monomeric hGHR from the chain connectivity and structures of the ECD and TMD, building models of the hGHR dimer would require further assumptions on the dimerization interface. Second, experimental complications arise both from the potential overlap with the void volume in the SEC-SAXS/SANS experiments and from possible structural heterogeneity. This may originate not only from a dynamic monomer-dimer equilibrium but also from different dimers in the nanodisc: the biologically relevant down-down dimer conformation, a trapped up-down conformation, or higher-order structures. We therefore concentrated on the reliable SEC-SAXS data representing monomeric hGHR-GFP in a nanodisc and used these to obtain the monomeric full-length hGHR-GFP structure embedded in a nanodisc bilayer.

## The structure of the monomeric full-length hGHR-GFP in a nanodisc

We followed a three-stage approach to derive a model of the structure of monomeric hGHR including the signal peptide and the GFP (hereafter named the hGHR-GFP) in the MSP1D1 nanodisc. First, we built a semianalytical model of the nanodisc-embedded full-length hGHR-GFP to refine the nanodisc parameters and to validate the overall structure of the complex. Second, we generated an ensemble of 6000 structures of the hGHR-GFP embedded in a POPC bilayer from 20 2-μs-long CG-MD simulations; these structures were back-mapped to all atoms and transferred to the model for the nanodisc. Last, from these models, we calculated the SAXS data of hGHR-GFP embedded in the nanodisc and compared to and refined against the SAXS data.

The semianalytical model of the SAXS scattering intensity from nanodisc-embedded hGHR-GFP (Fig. 5A) was calculated through the absolute squared sum of four scattering amplitude components arising from, respectively, the ECD-TMD, the ICD, the GFP, and the surrounding nanodisc (see Materials and Methods). The model, which implicitly includes the scattering cross-terms between the different components and take their internal phases into account, was implemented through the WillItFit platform (48). As in our previous work (26), we used a hybrid approach to the modeling, where different computational approaches were applied and combined for the four different components. Briefly, the ECD-TMD,

which is connected through a flexible linker between the ECD and TMD, was represented as a single rigid body through the atomic coordinates of one of the models produced with Rosetta (see Materials and Methods). The disordered ICD and its ensemble of conformations was modeled with a Gaussian random coil model parametrized by its $R_g$ with the center of mass displaced one $R_g$ from the disc surface, but where the connections at the two ends, respectively, to the TMD and the GFP, were disregarded. The GFP attached to the ICD was described through its atomic coordinates (PDB 1EMA) and allowed to take a random orientation in a certain "confusion volume" in extension of the disordered ICD. For the surrounding nanodisc, we allowed, as in our previous work (26, 27, 40), the lipid bilayer disc to take a slightly elliptical shape parametrized through its axis ratio. This accounts for the combined effect of less than maximal lipid loading and shape fluctuations. As in our previous work (26, 27, 40), we further constrained and reparametrized the underlying geometrical model into physically meaningful molecular parameters such as the number of POPC per disc and the area of POPC per headgroup. The scattering intensity corresponding to the model was calculated and fitted on an absolute scale. An excellent model fit to the experimental data ($\chi^2$ = 1.5; Fig. 5B, blue) was obtained using a nanodisc containing 122 POPC lipids, each with an area per headgroup of 63 Å$^2$ (49), an axis ratio of 1.4 of the elliptical bilayer, and an $R_g$ of the Gaussian random coil modeling the ICD of

73 Å (see details of the model in Materials and Methods and full account of model fit parameters in table S3). The number of lipids per disc was kept fixed at the value obtained from the experimental phosphorous analysis (fig. S4A). Likewise, the axis ratio of 1.4 was fixed on the basis of previous analyses (41). We note that the resulting fitted POPC area per headgroup agrees well with previous values obtained for POPC-loaded MSP1D1 nanodiscs (27, 40) and that the $R_g$ of the ICD accords with the value we determined for the isolated ICD. The analysis shows that the semianalytical model provides a low-resolution description of the nanodisc-embedded full-length hGHR-GFP and forms a basis for a more detailed molecular description.

In the next stage, a CG representation was built containing the full-length hGHR-GFP receptor embedded in a POPC bilayer (Fig. 5C). The simulation included the signal peptide and, hence, residues −18 to 620 (total of 638 residues) of the hGHR plus GFP. This full-length hGHR-GFP model was simulated with Martini 3 using the 10% increase in the strength of protein-water interactions found optimal for simulation of the hGHR-ICD. We ran 20 different 2-μs-long simulations and extracted 300 conformations (one every 5 ns) from each trajectory, discarding the first 500 ns of each run to allow for equilibration, providing a total of 6000 conformations of hGHR-GFP. These were back-mapped to all-atom representations and, one by one, embedded in the analytical nanodisc model that
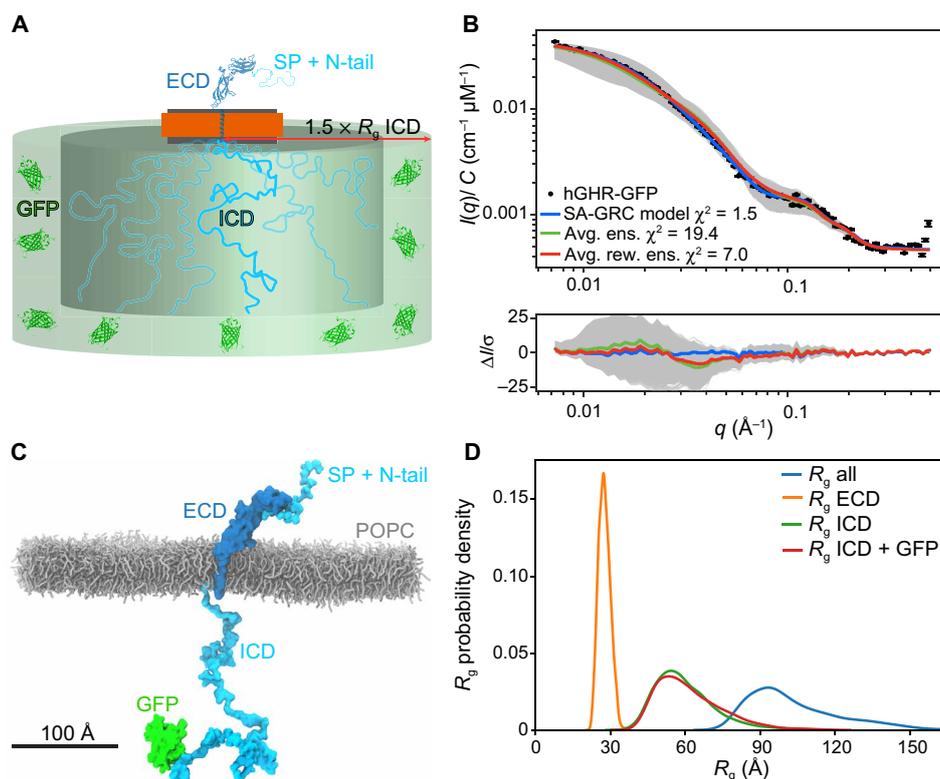
Fig. 5. Model of the full-length hGHR-GFP in nanodiscs. (A) Schematic representation of the semianalytical Gaussian random coil (SA-GRC) model. (B) Fits of the SA-GRC to the SAXS data of nanodisc-embedded hGHR-GFP (with GFP) (blue), the ensemble of 6000 conformations taken from the hGHR-GFP +POPC$_{pws10}$ simulations embedded in the nanodisc model (gray), their ensemble average (green), and reweighted ensemble average (red). (C) Representative snapshot from one of the hGHR-GFP + POPC$_{pws10}$ simulations (see Materials and Methods). POPC lipids are shown as gray sticks; protein is depicted in surface representation. Some lipids and all water and ions are omitted for clarity. (D) Distribution of $R_g$ from 6000 all-atom conformations obtained from the hGHR-GFP + POPC$_{pws10}$ simulation after reweighting against the SAXS data. The values are shown both for the full-length protein (blue) and for the individual structural components: ECD (orange), ICD (green), and ICD-GFP (red).

had been optimized through the semianalytical approach and following the WillItFit-based procedure previously described (*48*). SAXS scattering curves were calculated from the obtained ensemble (Fig. 5B, gray) and averaged, taking the mean $I(q)$ of the 6000 individual curves for each value of $q$ (Fig. 5B, green). Comparison to the experimental SAXS data showed that the MD-derived model, despite not being refined against the experimental data, provided a relatively good fit to the experiments (Fig. 5B, green) with a $\chi^2$ of 19. We used the BME approach (*31*) to improve agreement with the data further. We tuned the hyperparameter, $\theta$, which sets the balance between the information from the SAXS data and the force field to $\theta = 500$ (fig. S5, A and B). This resulted in an ensemble effectively containing $\approx 70\%$ of the conformations ($\phi_{eff} = 0.73$) from the simulation and obtained a considerable improvement of the fit ($\chi^2 = 7.0$; Fig. 5B, red). The resulting model thus combines information about the conformational preferences of hGHR-GFP encoded in both the martini model and SAXS data.

We proceeded to analyze the structural properties of hGHR-GFP described by the reweighted ensemble. Looking at the individual domains within the full-length protein, we found (reweighted) $R_g$ values of 60 ± 4 Å for hGHR-ICD, 62 ± 4 Å for hGHR-ICD-GFP, and 27 ± 2 Å for the hGHR-SP + ECD (Fig. 5D). Measurement of the reweighted average helix tilt angle (16° ± 1°) (see fig. S5C) shows that the TMD remains nearly parallel to the axis normal of the membrane plane as suggested by the XRD and OCD results obtained on the isolated hGHR-TMD. The ICD remained disordered and, for the most part, remained avoiding the membrane. Long-lived contacts and penetration of the bilayer were observed only for the intracellular juxtamembrane region (Q272-M277) and the Box1 motif (L278-K287) of the ICD, as well as for some residues from the ECD-TMD linker (fig. S5D, inset), in line with previous reports (*9*). Visual inspection of the trajectory (movie S1) showed that the ECD-TMD linker remained flexible, allowing the ECD to adopt a range of orientations while remaining mostly upright, as shown by the angle between the principal axis of the D2 domain and the $z$ axis (reweighted average, 34° ± 3°; fig. S5E). We note that the D1 domain remained far from the lipid surface. The N-terminal tail (N-tail) of the ECD remained disordered without long-lived contacts with the folded part of the ECD or the membrane.

In summary, our integrative model of the full-length monomeric hGHR in a nanodisc, containing almost equal amounts of structural order and disorder, combines information from molecular simulations and SAXS data recorded on the complex molecular system. In this way, the model provides the first experimentally derived detailed molecular insight into the structure of an intact, full-length class 1 cytokine receptor in a lipid membrane carrier system.

## DISCUSSION

Membrane proteins take on a variety of different topologies, sizes, and functions, and large portions of membrane proteins exist in tripartite structures that require different handling schemes and methodological studies. Such complexities are further amplified for membrane proteins having large fractions of structural disorder (*21, 50, 51*), which impose obstacles to classical structural biology. Thus, different topologies and order/disorder dispositions require different approaches, and one particular group of membrane proteins falls between the cracks by being too small and unstructured for cryo-EM, too large for NMR spectroscopy, and too dynamic for x-ray crystallography. An important subgroup of these membrane proteins, which plays key biological roles, is the cytokine receptor family.

In the present work, we examined the structure of an archetypal and particularly challenging membrane protein, the cytokine receptor hGHR, for which 50% of its chain is intrinsically disordered (Fig. 6). The structure of the monomeric hGHR revealed that when inserted
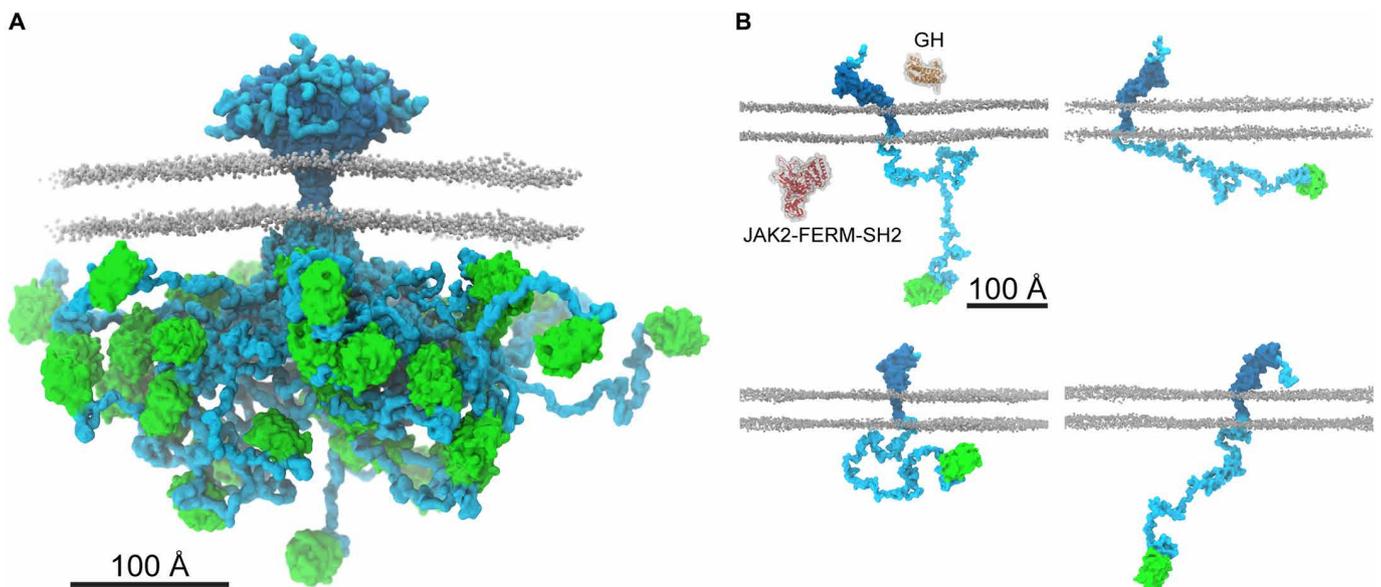
**Fig. 6. The ensemble structure of membrane-embedded full-length hGHR-GFP. (A)** Representative ensemble of conformations obtained from the last 1.5 μs of each of the 20 runs of 2-μs hGHR-GFP + POPC$_{pws10}$ simulations. Color scheme and representations as in Fig. 5C. **(B)** Examples of the multitude of domain orientations of hGHR-GFP in the membrane. In the first panel, the structures of hGH (PDB 3HHR_A; orange) and of JAK2-FERM-SH2 (PDB 4Z32; red) are shown. Color scheme and representation of hGHR and POPC as in Fig. 5C.

in a bilayer mimetic, neither the ECD nor the long, disordered ICD engage in long-lived contacts with the membrane. This is remarkable, although it should be noted that the lipids used in the current study do not fully mimic the complexity of native membranes by, e.g., lacking phosphoinositides or/and cholesterol, just as the proteoglycan layer on the extracellular side and the cytoskeleton on the inside are missing. We did, however, capture some lipid interactions by the intracellular juxtamembrane region (fig. S5D), which have been previously described (9). It is possible that the native composition of the bilayer may influence the conformation of the receptor, but, inherently, there is no affinity for the POPC bilayer. Thus, the intracellular, disordered domain protrudes from the bilayer and into the cytosol. Its average $R_g$ of 65-70 Å corresponds to an average end-to-end distance of about twice this value. This defines its capture distance and the large search volume (Fig. 6 and fig. S5F), which allows it to scout for and engage with kinases, phosphatases, and regulatory proteins such as the signal transducer and activator of transcription (STAT), suppressors of cytokine signaling, and the cytoskeleton (52).

A particular noteworthy observation from the structure of hGHR is the disordered, ~30-residue N terminus of the ECD, which has been neglected in all previous structural studies. The role of this N-terminal IDR in GHR function is unknown, but N-terminal IDRs are present in other family members, including the EPOR. An isoform of the GHR with a 22-residue deletion in the disordered N-tail (*d3*-GHR) shows altered extracellular signal–regulated kinase 1/2 signaling but unaltered STAT5 signaling, and *d3*-GHR individuals show increased life spans (53). Thus, key functional relevance is coupled to the N-tail. A search in the eukaryotic linear motifs database (54) suggests the presence of a glycosaminoglycan (GAG) attachment site, $_1$FGFS$_4$, in the N-tail. Of relevance to this, the WSXWS motif, which in hGHR is YGEFS, constitutes a C-mannosylation site linking the C1 atom of the α-mannose to the indole C2 atom of the tryptophan (3, 12). The WSXWS motifs has also been suggested to bind GAGs (12), so it is possible that the disordered N-tail of hGHR play similar roles as the WSXWS motif, and we notice a degenerate motif of this kind, also in the N-tail, given by the sequence $_{16}$WSLQS$_{20}$. Nonetheless, the function of the disordered N-tail of hGHR remains unestablished.

The integrative nature of our approach to determine the structure of the hGHR required development and optimization of several protocols. This was particularly necessary during the modeling and fitting of the SAXS data based on the combined semianalytical and experimentally driven molecular modeling approach to account for the structure and large flexibility of the hGHR. Key to the success was a scaling of the strength of the protein-water interaction in the CG-MD simulations of the ICD and full-length hGHR. This enabled reliable fits to the disordered chain in terms of $R_g$. On the semianalytical modeling side, we have expanded our previous approaches to interpret scattering data from bare nanodiscs and rigid membrane proteins incorporated into these discs (26, 27, 40), to now also allow for modeling membrane proteins with significant amounts of structural disorder. We emphasize that even if the parameters of the GHR model are custom-fitted to the hGHR-GFP system, the approach is fully generalizable and may be adapted to membrane proteins of similar topology provided that high-quality scattering data are available. Thus, the use of this integrative semianalytical and MD simulation–based approach suggests that SAXS (and SANS) in combination with MD simulations is a useful way of providing structural insight into

otherwise "method orphan" membrane proteins, particularly highlighting the interdomain orientations. This opens for more systematic investigations of, for example, single-pass transmembrane proteins in different environments, e.g., with respect to the lipid composition, the buffer environment, or with binding partners to understand how these very dynamic membrane proteins transduce information across the membrane. Furthermore, structures of not only other single-pass membrane proteins with similar complexity such as the cadherins and cell adhesion molecules (e.g., Down syndrome cell adhesion molecule) but also membrane proteins with long, disordered regions such as the solute carrier family 9, type II receptor serine/threonine family, and palmitoyl transferases now become accessible.

A key observation made possible from acquiring data on the full-length hGHR is the lack of restriction on the relative orientation of the domains (Fig. 6). Not only is the ICD and the N-tail disordered but also the flexible linker joining the ECD and TMD, which combined with the lack of membrane association allow them to freely reorient relative to each other, at least in the free state (Fig. 6). Thus, in addition to structure, it becomes important to consider how the flexibility of the entire chain takes on roles in signaling. From our studies, we were not able to derive whether correlated motions between the ECD and the ICD exist. However, once the hGH binds to the ECD, changes in conformation and flexibility may propagate along the chain, reaching the ICD and bound protein partners, eliciting signaling. Similar suggestions were put forward on the basis of data from solid-state NMR studies on the epidermal growth factor receptor, revealing increased dynamics in the bound state (55). Since the JAK2 binding site only constitutes ~6% of the ICD, and the STAT5 docking sites are ~200 to 300 residues away from it (56), conformational changes involving redistribution of the structural ensemble of the long, disordered region need to be achieved in a controlled manner. In addition, the ICD contains many short linear motifs (SLiMs), distributed along the chain in SLiM hotspots (8), and the space occupied by the free ICD (fig. S5, E and F) may therefore enable room for generation of larger, supramolecular signaling complexes. With the presence of two disordered chains in a dimer, the occupied space of each ICD chain may be reduced because of steric exclusion, affecting the formation and content of supramolecular complexes. With a structure of a full-length membrane protein embedded in a realistic membrane scaffold and containing a large disordered chain at hand, the understanding of regulation of signaling by disordered chains, often present in higher-order assemblies of several chains, now has a molecular platform from which new questions can be tackled.

## MATERIALS AND METHODS
### hGHR-ECD expression and purification
The DNA sequence coding for hGHR-ECD (1 to 245, C242S, no signal peptide) in a pET11a was bought from GenScript and transformed into competent Rosetta2 (DE3)pLysS cells. These were grown in 1 liter of LB medium with 3% (v/v) ethanol containing ampicillin and chloramphenicol (100 μg/ml) to $OD_{600}$ (optical density at 600 nm) = 0.6 to 0.8, and induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 4 hours at 37°C and 160 revolutions per minute (RPM). The cells were harvested by centrifugation (5000*g* for 15 min) and resuspended in one-time phosphate-buffered saline (PBS) (140 mM NaCl, 2.7 mM KCl, 10 mM $Na_2H_2PO_4$, and 1.8 mM $KH_2PO_4$) (pH 7.4) containing 25% (w/v) sucrose and 5 mM

EDTA. The cells were lysed on ice by sonication using an UP400S ultrasonic processor, $6 \times 30$-s sonication followed by 30-s rest at 50% amplitude. Following centrifugation (20,000$g$, 4°C) for 25 min, the pellet was resuspended in one-time PBS (pH 7.4) containing 25% (w/v) sucrose and 5 mM EDTA, repeated three times in total. The pellet was solubilized in 500 ml of 50 mM tris-HCl (pH 8.5), 10 mM β-mercaptoethanol (bME), and 6 M urea, heated for 5 min at 55°C, and left O/N (overnight) with slow stirring at 4°C. The amount of hGHR-ECD was estimated on an SDS-PAGE by comparing to the LMW (low molecular weight and diluted to a concentration less than 0.1 mg/ml in 50 mM tris-HCl (pH 8.5), 10 mM bME, and 6 M urea. To refold, hGHR-ECD was dialyzed against 4 liters of 150 mM NaCl, 50 mM tris-HCl (pH 8.5), 10/1 mM cysteamine/cystamin at 4°C, and a molecular weight cutoff of 12 kDa until the urea concentration was below 0.1 M. Following centrifugation at 20,000$g$ for 15 min, the sample was placed on ice and stirred slowly, while ammonium sulfate was added to a final concentration of 75% (w/v) and then left for 2 hours. The solution was centrifuged at 12,000$g$ at 4°C for 25 min, and the pellet was dissolved in 100 ml of Milli-Q water and left for 2 hours, followed by dialysis against 30 mM $NH_4HCO_3$ (pH 8.0) overnight at 4°C. After centrifugation at 13,000$g$ for 15 min, the supernatant was concentrated using a Millipore spin filter (10 kDa cutoff) and applied to a Superdex 75 16/85 column (GE Healthcare) at 4°C, 150 mM NaCl, and 30 mM $NH_4HCO_3$ (pH 8.5). Selected fractions were reapplied to a Superdex 200 increase 10/300 GL column in 20 mM $Na_2H_2PO_4$ (pH 7.5) and 150 mM NaCl before SAXS measurements.

### hGHR-ICD expression and purification

The coding region for hGHR-ICD (S270-P620) was cloned into a pGEX-4T-1 vector, containing an N-terminal glutathione $S$-transferase (GST)–tag followed by thrombin cleavage site and transformed into Bl21(DE3) cells. Expression was done in 1 liter of terrific broth (TB) medium containing ampicillin (100 μg/ml). At $OD_{600}$ = 0.6 to 0.8, cells were induced by 1 mM IPTG for 4 hours at 37°C and 160 RPM. Cells were harvested by centrifugation and resuspended in 40 ml of one-time PBS (pH 7.4), 0.1% (v/v) Triton X-100, and a tablet complete EDTA-free protease inhibitor cocktail. The cells were lysed on ice by sonication using an UP400S ultrasonic processor, four times 30-s sonication followed by 30-s rest at 100% amplitude. Following centrifugation (20,000$g$, 4°C) to remove cellular debris, the lysate was applied to a glutathione Sepharose 4 fast flow column (GE Healthcare) and incubated for 2 hours at 25°C. The column was washed with 50 ml of one-time PBS (pH 7.4) and eluted 20 ml of 50 mM tris-HCl and 10 mM reduced glutathione (pH 7.4). The eluted solution was dialyzed against 1 liter of 20 mM tris-HCl and 150 mM NaCl (pH 7.4) at 4°C. The GST-tag was cleaved off by the addition of 20 U of thrombin per liter of culture, leaving residues GS in the N terminus. The sample was then concentrated; 10 mM dithiothreitol was added and heated to 72°C for 5 min, incubated on ice, and centrifuged for 20,000$g$ at 4°C for 10 min. A final purification on a Superdex 200 increase 10/300 GL column (GE Healthcare) in 20 mM $Na_2H_2PO_4$ (pH 7.5) and 150 mM NaCl was done, and selected fractions were used for SAXS measurements.

### hGHR-ICD-GFP-H₁₀ expression and purification

The coding region for hGHR-ICD (S270-P620) including an N-terminal methionine, C-terminal Tobacco Etch Virus (TEV) cleavage (ENLYFQS) site followed by a yeast enhanced GFP (EGFP) (57),

and 10 histidines (hGHR-ICD-GFP-H₁₀) in a pET-11a vector was bought from GenScript. Expression was done in 1 liter of TB medium (for SAXS) and in $^{15}$N-labeled minimal medium [22 mM $KH_2PO_4$, 62.5 mM $NaH_2PO_4$, 85.6 mM NaCl, 1 mM $MgSO_4$, 1 ml of "trace element solution," 4 g of glucose, and 1.5 g of $NH_4Cl$ ($^{15}$N-labeled nitrogen)] (for NMR) containing ampicillin (100 μg/ml). At $OD_{600}$ = 0.6 to 0.8, expression was induced by 1 mM IPTG for 3 hours at 37°C and 160 RPM. Cells were harvested by centrifugation and resuspended in 40 ml of one-time PBS (pH 7.4) and a tablet of cOmplete EDTA-free protease inhibitor cocktail. The cells were lysed on ice by sonication using an UP400S ultrasonic processor, four times 30-s sonication followed by 30-s rest at 100% amplitude. Following centrifugation (20,000$g$, 4°C), the pellet containing hGHR-ICD-GFP-H₁₀ was solubilized by adding 40 ml of 20 mM $NaHCO_3$ (pH 8.0), 150 mM NaCl, and 8 M urea. Following centrifugation (20,000$g$, 4°C), the supernatant was refolded by dialysis in two steps: first, by dialysis in 4 liters of 20 mM $NaHCO_3$ (pH 8.0), 150 mM NaCl, and 4 M urea at 4°C using 3 kDa of molecular weight dialysis bag cutoff for 4 hours and then in 4 liters of 20 mM $NaHCO_3$ (pH 8.0) and 150 mM NaCl at 4°C overnight. Following centrifugation (20,000$g$, 4°C), the supernatant was applied to a prepacked 5 ml of Ni resin column. The column was washed with three column volumes (CVs) of 20 mM $NaCHO_3$ (pH 8), 150 mM NaCl, and 10 mM imidazole and eluted using 20 mM $NaCHO_3$ (pH 8.0), 150 mM NaCl, and 250 mM imidazole. Fractions containing hGHR-ICD-GFP-H₁₀ were concentrated and applied to a Superdex 200 16/60 increase column in 20 mM $NaH_2PO_4$/$Na_2H_2PO_4$ (pH 7.5) and 150 mM NaCl. Fractions containing hGHR-ICD-GFP-H₁₀ were analyzed by SDS-PAGE, and selected fractions were used for SAXS and NMR experiments.

### hGH purification

hGH in a pJExpress414 was bought from ATUM, USA (formerly DNA2.0) and transformed into competent BL21 (DE3) cells. These were grown in 1 liter of TB containing ampicillin (100 μg/ml) to $OD_{600}$ = 0.6 to 0.8 and induced by addition of 1 mM IPTG for 4 hours at 37°C and 160 RPM. Cells were harvested by centrifugation (5000$g$, at 4°C, 25 min) and resuspended in 50 ml of 50 mM tris, 0.5 mM EDTA (pH 8.0), and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were lysed on ice by sonication using an UP400S ultrasonic processor, five times 30-s sonication followed by 30-s rest at 50% amplitude. Following centrifugation at 10,000$g$ at 4°C for 15 min, the pellet was resuspended in 20 ml of 10 mM tris, 1 mM EDTA (pH 8.0), and 1 mM PMSF. The pellet was recentrifuged two times, and the supernatant was discarded and solubilized in 250 ml of 5 M guanidinium chloride (GuHCl), 200 mM $Na_2HPO_4$/$NaH_2PO_4$ (pH 7.0), and 15 mM bME. The solution was heated for 10 min at 55°C and stirred mildly for 2 hours at room temperature. The solution was diluted in denaturation buffer [5 M GuHCl, 200 mM $Na_2HPO_4$/$NaH_2PO_4$ (pH 7.0), and 15 mM bME] to reach an hGH protein concentration below 0.1 mg/ml. The solution was dialyzed in a 5-liter beaker, with a drain on the top, and filled with 5 M GuHCl, 200 mM $Na_2HPO_4$/$NaH_2PO_4$ (pH 7.0), and 15 mM bME. A peristaltic pump was used to add refolding buffer [20 mM $NH_4HCO_3$ (pH 8.0) and 200 mM NaCl] at the bottom of the beaker with a flow rate of 1.5 ml/min. After 3 days, when the GuHCl concentration was below 1.5 M, the dialysis bags were transferred to a new 5-liter beaker with 20 mM $NH_4HCO_3$ (pH 8.0) and 200 mM NaCl and dialyzed three times until the concentration of GuHCl was below 0.1 M. Following centrifugation for 18,000$g$ for 10 min, the

supernatant was concentrated using a Millipore Pellicon module to approximately 30 ml. The solution was applied to a Superdex 75 26/600 column in 20 mM $NH_4HCO_3$ and 100 mM NaCl (pH 8.0). Selected fractions were dialyzed against 5 liters of 20 mM tris (pH 8.0) twice and applied to a HiTrap QFF of 5 ml. The sample was eluted in 20 mM tris (pH 8.0) by a salt gradient from 0 to 1 M NaCl at a flow rate of 5 ml/min over 20 CV. Selected fractions were flash-frozen in liquid nitrogen and left at −20°C before use.

## Analytical SEC
Analytical SEC experiments of a set of samples with various ratios of hGH:hGHR-ECD were run on Superdex 200 increase 10/300 GL column in 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4) and 100 mM NaCl at room temperature with a flow rate of 0.5 ml/min. Protein sample concentration was in the micromolar range but varied. The column was calibrated using conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa), acetone, and blue dextran, and apparent partition coefficient, $K_{av}$, was determined for all peaks.

## Circular dichroism spectroscopy
Far-ultraviolet CD spectra were recorded on 10 µM hGHR-TMD in 2 mM DHPC, 5 µM hGH, and 5 µM hGHR-ECD in 10 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4). The spectra were recorded on a Jasco J-810 spectropolarimeter in a 1-mm quartz glass Suprasil cuvette (Hellma) at 20°C. A total of 10 scans were accumulated from 260 to 190 nm for each sample, and buffer background was recorded at identical setting and subtracted. For hGHR-TMD, the background included 2 mM DHPC. The scan mode was continuous with a speed of 10 nm/min and a data pitch of 0.1 nm. The spectra were processed and smoothened (means-movement method, convolution width of 25) and converted into mean residue ellipticity values.

## hGHR-TMD purification
hGHR-TMD was expressed in *E. coli* and purified as previously described (*32*).

## Oriented circular dichroism
hGHR-TMD was dried under a flow of $N_2$ and subsequently dissolved in MeOH:CHCl$_3$ (5:1) to reach a final stock solution of hGHR-TMD (0.4 mg/ml). To validate the concentration, 100 µl of the stock solution was dried under $N_2$ flow and resuspended in 100 µl of 50 mM SDS in phosphate buffer (pH 7.0), and the absorbance at 280 nm was measured. Lipid stock solutions of POPC, DOPC, and POPC/POPS (3:1) were prepared in MeOH:CHCl$_3$ (1:1) at 0.25 and 5 mg/ml. The protein and lipid stock solutions were mixed in the following L:P ratios: 40:1, 50:1, 70:1, 100:1, 150:1, and 200:1. Six micrograms of protein was applied to a quartz glass with a Hamilton pipette for each experiment. The sample was spread over a fixed circular area on the glass and subsequently dried under vacuum for 3 hours to remove the MeOH:CHCl$_3$. The dried sample was mounted in a sample holder and was hydrated overnight in a chamber with a saturated $K_2SO_4$ solution at 20°C. Last, the samples were loaded into a rotor in a Jasco J-810 spectropolarimeter, and the spectra were recorded from eight different angles: 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. Each spectrum was measured twice from 260 to 180 nm with a scanning speed of 20 nm/min, a data pitch of 0.1, and a response time of 8 s. The spectra were averaged, and the reference OCD spectra from samples with the same amount of lipid were

subtracted. The OCD spectra were recorded from eight different angles to even out linear dichroism (fig. S2C) (*37*). The spectra from different angles were averaged, background subtracted, and normalized to the intensity at 222 nm. High-voltage effects prevented the measurement of higher L:P ratios.

## X-ray diffraction
Highly oriented multilamellar membranes were prepared on single-side polished silicon wafers. POPC (Avanti), POPS (Avanti), and 1,2-dimyristoyl-*sn*-glycero-3-phospho-L-serine (Sigma-Aldrich) were mixed with hGHR-TMD at 2 and 20 mol % concentrations in 2,2,2-trifluoroethanol:chloroform [1:1 (v/v)] at a solution concentration of 18 mg/ml. The wafers were sonicated in 1,2-dichloromethane for 30 min and then rinsed with alternating methanol and water (18.2 megaOhm·cm). The wafers were dried, and 75 µl of the solution was deposited. After drying, the samples were placed in a vacuum for 24 hours at 37°C to allow for trace solvent evaporation and annealing. Samples were then hydrated in a closed chamber at 97% relative humidity with a separate $K_2SO_4$-saturated solution for 48 hours before scanning.

XRD data were obtained using the Biological Large Angle Diffraction Experiment (BLADE) at McMaster University. BLADE uses a 9-kW (45 kV, 200 mA) CuKα rotating anode at a wavelength of 1.5418 Å using a Rigaku HyPix-3000 two-dimensional semiconductor detector with an area of 3000 $mm^2$ and 100 $µm^2$ of pixel size (*58*). All samples were prepared and measured in replicates to check for consistency. EDPs were determined from specular reflectivity, as previously described (*36*). The lamellar spacing, $d_z$, was determined from the spacing of the reflectivity Bragg peaks. Hermans orientation function was determined by integrating the intensity of the third Bragg peak as function of the meridonal angle $\phi$ (the angle relative to the $q_z$ axis), as described in (*59*). The CYANA and CS-Rosetta models of the monomer and the Protein Data Bank (PDB) structure of the dimer (PDB 5OEK) were used and fitted to the difference electron density. The calculation uses the electronic distribution of each atom of the molecule. To account for thermal motions, a Gaussian distribution, with a half width of 3.5 Å, is placed at the correct position, and all the atomic contributions are added. The linear electron density is then calculated by projecting the calculated electronic distribution on a linear axis and fitting to the experimental data. Position and tilt angle were determined from the fit by shifting the molecule along the $z$ axis and varying the $z$ projection (*60*).

## NMR spectroscopy and structure of the TMD
NMR spectra were recorded on a 750-MHz ($^1$H) Bruker AVANCE spectrometer equipped with a cryogenic probe. Unless otherwise specified, all NMR samples contained 10% (v/v) $D_2O$ and 1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Proton chemical shifts were referenced internally to DSS at 0.00 parts per million (ppm), with heteronuclei referenced by relative gyromagnetic ratios. Free induction decays were transformed and visualized in NMRPipe or TopSpin (Bruker Biospin) and analyzed using CcpNmr Analysis software. For hGHR-TMD, all NMR spectra were recorded at 37°C in 2 mM tris(2-carboxyethyl)phosphine (TCEP), 0.05% (v/v) NaN$_3$, 50 mM NaCl, and 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4). The spectra for backbone assignments of hGHR-TMD [HNCO, HNCAHC, HNCA, HNCACB, CBCA(CO)NH, $^1$H, and $^{15}$N-HSQC] were measured on 1 mM $^{13}$C,$^{15}$N-hGHR-TMD solubilized in 210 mM DHPC. Secondary structure content was evaluated from backbone

chemical shifts using the MICS program (33). $R_2$ transverse relaxation rates of 0.5 mM $^{15}$N-hGHR-TMD in 110 mM DHPC were determined from a series of $^{1}$H,$^{15}$N-HSQC spectra with varying relaxation delays between 10 and 250 ms and triple replica at 130 ms. The relaxation decays were fitted to single exponentials and relaxation times determined using CcpNmr Analysis software. A low-resolution model of hGHR-TMD was calculated using CYANA (34) including only dihedral angle restraints derived from the backbone chemical shifts using TALOS (61). Standard settings were used calculating 50 conformers with 4000 torsion angle dynamics steps. The 10 best conformers, with the lowest CYANA target function score, was used for further modeling. In addition, a model of hGHR-TMD was calculated using CS-Rosetta (35). The chemical shifts obtained for the hGHR-TMD were used as an input for the CS-Rosetta server (https://spin.niddk.nih.gov/bax/nmrserver/csrosetta/) from which 3- and 9-nucleotide oligomer fragments were obtained and used for de novo modeling. We adapted previously described protocols (62) to the relevant parts of the calculation: (i) ab initio modeling using the NMR-derived chemical shifts, (ii) definition of the membrane-spanning region, and (iii) rescoring of the Rosetta energy considering the chemical shifts. A total of 1000 models were generated. Clustering of the 10 lowest-energy models around the lowest one, with $C^{\alpha}$-RMSD < 2 Å for the W249 – K271 region, indicated successful prediction (see Fig. 2D).

All NMR data of hGHR-ICD and hGHR-ICD-GFP-H$_{10}$ were acquired at 5°C to minimize amide exchange in 1 mM TCEP, 150 mM NaCl, and 20 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$ (pH 7.4). $^{1}$H,$^{15}$N-HSQC spectra were acquired at concentrations of 150 μM for $^{15}$N-hGHR-ICD and 100 μM for $^{15}$N-hGHR-ICD-GFP-H$_{10}$. The hydrodynamic radii ($R_h$) of hGHR-ICD and hGHR-ICD-GFP-H$_{10}$ were determined from a series of $^{1}$H,$^{15}$N-HSQC spectra with preceding pulse-field gradient stimulated-echo longitudinal encode-decode diffusion filter and with the gradient strength increasing linearly from 0.963 to 47.2 G cm$^{-1}$. To determine the diffusion coefficients ($D$), the decay curves of the amide peaks were plotted against the gradient strength and fitted in Dynamics Center (Bruker) using

$$I = I_0 \exp(-10^4 D_x^2 \gamma^2 \delta^2 (\Delta - \delta/3))$$

in which $I$ is the intensity of the NMR signal at the respective gradient strength, $I_0$ the intensity without applied gradient, $x$ the gradient strength in G cm$^{-1}$, $\gamma = 26{,}752$ rad G s$^{-1}$, $\delta = 3$ ms, and $\Delta = 250$ ms. $R_H$ was calculated from the diffusion coefficient using the Stokes-Einstein relation, $R_H = k_B T/(6\pi\eta D)$, with η being the viscosity of water at 5°C.

### Production of full-length hGHR
See Kassem *et al.* (21) for expression, purification, and reconstitution of hGHR-GFP in POPC-containing MSP1D1 nanodiscs. pMSP1D1 was a gift from S. Sligar (Addgene plasmid no. 20061; http://n2t.net/addgene:20061; RRID:Addgene_20061) (63).

### Native mass spectrometry
Purified hGHR-GFP in detergent was desalted and concentrated using C4 ZipTips (Merck Millipore), eluted into 50% acetonitrile/0.1% formic acid, and directly infused into an Orbitrap Fusion Tribrid mass spectrometer equipped with an offline nanospray source using borosilicate capillaries (Thermo Scientific). The capillary voltage was 1.5 kV in positive ionization mode, and the pressure in the ion-routing multipole was maintained at 0.11 torr. Ten percent of HCD activation energy in the ion trap was used to dissociate any residual detergents from the

protein. Spectra were acquired in the Orbitrap mass analyzer operated in high mass mode between 1500 to 6000 mass/charge ratio ($m/z$) with an injection time of 10 ms and a resolution of 60,000 full width at half maximum at 200 $m/z$. Data were analyzed using Excalibur (Thermo Scientific) and UniDec (unidec.chem.ox.ac.uk) software packages.

### Phosphorus analysis
The POPC-hGHR-GFP ratio of the formed nanodiscs with POPC-hGHR-GFP inserted was determined by phosphorus analysis (43). This was done by hydrolyzing POPC in H$_2$SO$_4$ to release free phosphate (PO$_4^{-3}$), which reacted with molybdate to produce a blue chromophore, absorbing at 812 nm. A series of phosphate standards from 0 to 80 nM Na$_2$HPO$_4$ and hGHR-GFP in MSP1D1 at approximately 1 μM were prepared. Aliquots of 175 μl of each sample were transferred to glass tubes. HClO$_4$ was added [400 μl, 72% (v/v)] to each sample, and the glass tubes were loosely closed using glass pearls. The samples were heated to 180°C in a heating block in a fume hood for 1 hour and then left at room temperature to cool for 30 min. Four milliliters of 125 mM (NH$_4$)$_6$Mo$_7$O$_{24}$ × four H$_2$O was added to each sample and vortexed, followed by addition of 500 μl of 10% (w/w) ascorbic acid and vortexed again. Samples were then heated to 80°C for 10 min in a water bath and subsequently cooled in ice water. Absorption was measured at 812 nm. A phosphate standard curve was generated, using the Na$_2$HPO$_4$ standards, by linear regression and was used to determine the content of phosphate in the hGHR-GFP in MSP1D1 samples.

### Gel quantification of hGHR-GFP-loaded nanodiscs
Standards of hGHR-GFP and MSP1D1 with a known absorption at 280 nM were prepared and loaded in different amounts of the same gel and three aliquots of hGHR-GFP-loaded nanodiscs taken from three different positions of the SEC elution profile (fractions 1, 2, and 3). The gels were stained with Coomassie brilliant blue G-250 (Bio-Rad) and subsequently destained in 15% (v/v) ethanol, 5% (v/v) acetic acid, and 5% glycerol (v/v). Gel images were obtained on a LAS 4000 imager (GE Healthcare, USA), and the images were quantified in ImageJ. The intensities of the standards were fitted by linear regression, and the amount of hGHR-GFP relative to MSP1D1 was quantified accordingly.

### Microscale thermophoresis
hGH was labeled with NT-647-NHS (64) using the Monolith NT Protein Labeling Kit RED-NHS (NanoTemper Technologies) for 1 hour at room temperature with NT-647-NHS at a molar ratio of 1:3 in labeling buffer following the protocol. These conditions favor the modification of the N-terminal amino group. Free dye was separated from reacted dye using the provided desalting column. The ratio between fluorophore and protein was 0.2. The equilibrium binding between 20 nM NT-647-NHS–labeled hGH and hGHR (MSP1D1) was calculated from the change in thermophoresis $\Delta F_{norm} = \Delta F_{hot}/\Delta F_{cold}$ measured on a Monolith NT.115 (NanoTemper Technologies). For hGH$_{G120R}$, the raw fluorescence change was used to determine the binding affinity. A twofold dilution series of monomeric hGHR-GFP from 750 nM to 23 pM was prepared in 20 mM Na$_2$HPO4/NaH$_2$PO$_4$ (pH 7.4) and 100 mM NaCl and measured in triplicates. Samples were loaded into the Monolith NT.115 Premium Capillaries (NanoTemper Technologies), and the thermophoresis and raw fluorescence signals were measured at 25°C with a light-emitting diode power of 80% and an infrared laser power of 100%. The dissociated constant $K_d$ was obtained by fitting the data by

$$Y = Y_0 + \frac{Y_f - Y_0}{2\,[P]_{total}} \times \left( K_d + [P]_{total} + X - \right.$$
$$\left. \sqrt{\left(K_d + [P]_{total} + X\right)^2 - 4\,[P]_{total}X} \right)$$

where $Y$ is the measured fluorescence/MST, $X$ is the ligand concentration, $[P]_{total}$ is the total concentration of the protein, $Y_f$ is the estimated end point of the titration, and $Y_0$ is the start point.

### N-glycosylation removal by endoglycosidase H

One microgram of purified full-length hGHR-GFP was incubated with 500 U of Endo-H (New England Biolabs, USA) at 4°C in 20 mM $NaH_2PO_4/Na_2HPO_4$ (pH 7.4), 150 mM NaCl, and 5% (v/v) glycerol. The sample was separated and analyzed on a 15% SDS-PAGE gel and visualized by in-gel fluorescence on a LAS 4000 imager (GE Healthcare, USA).

### Western blotting

hGHR-GFP was separated on a 15% SDS-PAGE gel and blotted to a polyvinylidene difluoride. Horseradish peroxidase–conjugated concanavalin A (Sigma-Aldrich, L6397) was used to identify O-glycosylations after Western blotting. Chemiluminescence was detected by using the Immobilon Western Chemiluminescent HRP Substrate from Millipore and the LAS 4000 imager (GE Healthcare, USA).

### Small-angle x-ray and neutron scattering

SAXS data on hGH, hGHR-ECD, and the hGH:hGHR-hECD 1:1 and 1:2 complexes were collected at the PETRA III, P12 beamline (DESY Synchrotron, Hamburg), following standard procedures at 8°C. All samples were concentrated and run on a Superdex 200 increase 10/300 GL in 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4) and 150 mM NaCl before measuring. The most concentrated top fractions were taken, except for 1:1 complex, where the fraction was taken to the right of the peak, to make sure the hGH:hGHR-ECD 1:2 complex was absent in the sample. hGH was measured at 1.8 mg/ml, ECD at 3.5 mg/ml, hGH:hGHR-ECD 1:1 complex at 0.3 mg/ml, and the hGH:hGHR-ECD 1:2 complex at 1.3 mg/ml. The scattering curves, each of which has an average of 40 frames, were recorded, and the buffer was measured before and after each sample. The processing and preliminary data analysis was done using the ATSAS package (65). As a part of the process, buffer scattering curves before and after the sample were averaged and subtracted from the scattering curve of the sample. The scattering curves were scaled into units of 1/cm using a measurement of water as secondary standard and subsequently logarithmically rebinned. For the full-length hGHR in MSP1D1, in-line SEC-SAXS of the sample in 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4) and 150 mM NaCl was performed at BM29 (European Synchrotron Radiation Facility, Grenoble) equipped with a Superose 6 increase 10/300 GL (GE Healthcare) running at a flow rate of 0.75 ml/min. In-line SEC-SANS data on the full-length hGHR in MSP1D1 were recorded on the D22 small-angle scattering diffractometer at Institute Laue-Langevin, Grenoble, France. The in-line SEC was performed using a the recently commissioned and described modular HPLC system (Serlabo) in 20 mM $Na_2HPO_4/NaH_2PO_4$ (pH 7.4) and 150 mM NaCl on a Superose 6 increase 10/300 GL (GE Healthcare) (42, 66). The flow rate was lowered from the 0.75 ml/min used in the SEC-SAXS measurements to 0.05 ml/min when the peak was reached in the lower-intensity SANS to get as good counting statistics on the individual frames as possible. Two settings were used,

11.2 and 2.0 m (with collimation lengths of 11.2 and 2.8 m, respectively), giving a $q$ range between 0.0044 and 0.46 $Å^{-1}$. The intensities were binned into 30-s frames.

### Modeling of the hGHR-ECD

To build a model of the full-length hGHR-ECD that covers the same sequence of the construct used in the experimental procedures, the following steps were performed: (i) We selected a previously determined crystal structure of the GHR-ECD (chain C of PDB entry 3HHR, residues 32 to 236) (17). (ii) We built models for the missing loops (residues 57 to 61 and 74 to 77) using the MODELLER interface of Chimera (67). (iii) The missing N-terminal (residues 1 to 31) and C-terminal (residues 237 to 245) tails were modeled as ensembles to capture their flexibility in the fitting of SAXS data. The Rosetta (30) routine Floppy tail (29) was used to generate 5000 conformations of both tails.

### Modeling of the hGHR ECD-TMD linker

The linker between the hGHR-ECD and hGHR-TMD (S237-W249) is not present in the available structures of hGHR-ECD, and its structure may play a relevant role in determining the proper ECD-TMD orientation. Thus, this linker was modeled to provide a starting conformation of the hECD-TMD region of hGHR for further use in the modeling of the full-length hGHR structure. To do this, the recently developed mp_domain_assembly protocol (68) in Rosetta_MP was used. We used the model described above of the ECD with missing loops completed (residues 32 to 236, i.e., without the SP and N- and C-terminal tails) and the TMD structure corresponding to an NMR-derived CYANA model (residues 250 to 272). A total of 5000 models were built with the best 10 (according to their Rosetta score) selected for further analysis, and the best-ranked model was used as a rigid body in the semianalytical models of hGHR-GFP in a nanodisc and as starting conformation in the building of the full-length hGHR-GFP CG model (see below).

### Modeling of the hGHR-ICD

We built a model of the hGHR-ICD (residues 288 to 638) as a random coil from its sequence using PyMOL. This all-atom model was used to build a CG system using the martinize.py and insane.py scripts to obtain a system of protein, water, and 150 mM NaCl with the martini 3 (m3.b3.2) (69) topology. The system contained a total of 455,457 beads and size of 383 Å by 383 Å by 383 Å. This system was used as the initial structure to perform all the GHR-ICD CG simulations (see below).

### Structural model for the full length hGHR-GFP

A full-length model of intact hGHR-GFP with the signal peptide on its N terminus and a GFP on its C terminus was built using the different parts modeled separately. A representative conformation from the reweighted subensemble of the full-length ECD (residues 1 to 237) was aligned to the best model of the ECD-TMD to obtain a complete ECD-TMD structure (residues 1 to 272). A representative structure of the ICD (residues 273 to 620) was taken from the back-mapped conformation from the CG-MetaD-$R_g$ simulation with 10% increase in the protein-water interaction strength (see below). Rotations of the peptide bond between residues 273 and 274 had to be adjusted to allow the correct orientation of the ICD with respect to the TMD and the membrane plane. EGFP (PDB 1EMA) was added at residue 620. The signal peptide (residues −18 to −1) was added to the

N terminus of this model as a coil using MODELLER. This all-atom model was used to build a CG system using the martini Maker module (70) of CHARMM-GUI to obtain a system of protein (hGHR-GFP) + POPC + water + 150 mM NaCl for the martini 2 force field. The topology was later adapted to open the beta version of martini 3 (m3.b3.2) (69). The final system contains 453,662 beads and has a size of 361 Å by 361 Å by 406 Å.

## CG-MD simulations

MD simulations were performed with Gromacs 2016, 2018, or 2019 (71) using the open beta version of the martini 3 (3.b3.2) force field (69). In some simulations, we increased the strength of interactions between protein and water to avoid excessive compaction of the disordered regions. To find the optimal factor to change these protein-water interactions, we performed two sets of simulations of the GHR-ICD system with different values of the protein-water interaction strengths (increased between 5 and 15%). Unbiased simulations were performed with a 5% (5.6 μs), 6% (3.1 μs), 8% (3.3 μs), and 10% (5.2 μs) increase, while metadynamics simulations (see below) were performed with a 10% (10.1 μs), 11% (10.1 μs), 12% (10.1 μs), 13% (10.1 μs), 14% (9.9 μs), and 15% (9.4 μs) increase. On the basis of the best reproduction of $R_g$ and best fit to the hGHR-ICD SAXS data (see fig. S3), we chose a 10% increase in interaction strength and used this also for the simulation of the hGHR-GFP + POPC system.

We performed 20 independent unbiased MD simulations (2 μs each) of the hGHR-GFP + POPC system with a time step of 5 fs. Other simulation parameters, common to all the CG simulations performed, were chosen following the recommendations in (72). Briefly, the Verlet cutoff scheme was used considering a buffer tolerance of 0.005 kJ/(mol ps atom). The reaction field method was used for Coulomb interactions with a cutoff of 11 Å and a dielectric constant of $\varepsilon_r$ = 15 for water. For van der Waals interactions, the cutoff scheme with a cutoff of 11 Å was used. The velocity rescaling thermostat was used with a reference temperature of $T$ = 300 and 310 K for the hGHR-ICD and hGHR-GFP + POPC simulations, respectively, with a coupling constant of $\tau_T$ = 1 ps. For the equilibrations, the Berendsen barostat was used ($p$ = 1 bar, $\tau_p$ = 3 ps), whereas the production runs were performed with a Parrinello-Rahman barostat ($p$ = 1 bar, $\tau_p$ = 12 ps) (73). A semi-isotropic pressure coupling was used for the hGHR-GFP system embedded on a lipid bilayer. For all systems, an initial round of equilibrations with decreasing constraints applied to the protein beads (hGHR-ICD) and protein beads and lipid beads (hGHR-GFP) was performed.

Sampling of the hGHR-ICD simulations with an increase in the protein-water interactions of 10, 11, 12, 13, 14, and 15% was enhanced using a well-tempered metadynamics (74) protocol applied with PLUMED 2.5 (75). The $R_g$ of the protein was used as collective variable within the boundaries of 30 to 110 Å. The metadynamics parameters used were a bias factor of 50, Gaussian height of 4.2 kJ/mol, and collective variable space Gaussian widths equal to 0.3.

Analysis of the MD trajectories was performed using plugins and analysis tools implemented in VMD, GROMACS, and PLUMED together with in-house-prepared tcl and python scripts. All molecular renderings were done with VMD.

## Fitting of the SAXS data of the hGHR-ECD and hGHR-ICD

Similar protocols were used to fit the SAXS data of the hGHR-ECD and hGHR-ICD with the conformations obtained from the modeling

of hGHR-ECD and MD simulations of hGHR-ICD, respectively: (i) For the hGHR-ECD, the SAXS profile of each model was directly calculated and fitted to the SAXS data using Pepsi-SAXS (76), with all parameters free. For the conformations obtained from different hGHR-ICD simulations, an initial round of back-mapping was performed to go from CG to all atoms as described in (58), before calculating and fitting its SAXS profile with Pepsi-SAXS. (ii) From the fits, the average value of the hydration shell contrast was calculated (hGHR-ECD = 7.4%; hGHR-ICD = 4%) and used as a fixed parameter in a second round of fitting. (iii) The average scattering profile of the ensemble was calculated directly from the Pepsi-SAXS intensity files and compared to the data. (iv) In the case of hGHR-ECD, the BME (31) procedure was used to reweight the ensemble against the experimental data. From the reweighted ensemble, a representative subensemble of 500 conformations was obtained.

## Semianalytical model for the ND-embedded hGHR-GFP

To generate the semianalytical model for the full-length hGHR-GFP, a hybrid approach, which combines analytical approaches to describe the nanodisc and the hGHR-ICD with rigid body modeling for the ECD-TMD and the GFP, was implemented in the WillItFit (48) framework. The mathematical model for hGHR-GFP in nanodiscs, illustrated in Fig. 5A, is composed of four distinct amplitude components arising from the ECD-TMD, the ICD, the attached GFP, and the surrounding nanodisc. The final expression for the total scattering intensity was calculated on the absolute scale as the orientationally averaged scattering amplitude squared according to standard theory of small-angle scattering

$$I(q) = n \cdot \langle \, | \, A_{ECD-TMD}(\vec{q}) + A_{ICD}(\vec{q}) + A_{GFP}(\vec{q}) + A_{ND}(\vec{q}) \, |^2 \rangle_\Omega$$

where $\langle \ldots \rangle_\Omega$ denotes the orientational average, $| \ldots |$ denotes the complex norm, $n$ is the number density of particles, and $A(\vec{q})$ is the scattering amplitude of each component for a single particle, each of which is equipped with a phase factor accounting for its relative position in the complex. Subscript ECD-TMD refers to the ECD with the TMD, ICD refers to the intrinsically disordered ICD, and GFP refers to the green fluorescent protein, which is fused to the ICD. ND refers to the surrounding POPC-loaded nanodisc. For each amplitude term, $A(\vec{q})$, we furthermore have ensured correct normalization through $A(\vec{q}) = \Delta\rho \cdot V \cdot F(\vec{q})$, where $\Delta\rho$ is the average excess scattering length density, $V$ is the molecular volume, and $F(\vec{q})$ is the normalized form factor amplitude for the relevant component. The model for the surrounding nanodisc, $A_{ND}(\vec{q})$, is the same as we have described previously (40): A stack of five elliptical cylinders representing the phospholipid bilayer is surrounded by a hollow elliptical cylinder representing the two stacked MSPs. As done previously (40), molecular constraints were systematically implemented to constrain the nanodisc solution space. As a part of this, the height of the MSP was fixed to a value of 25.8 Å, as derived from a high-resolution structure of nanodiscs (39), and values for the molecular volume (ν) for the POPC lipids were taken from the literature (49). The scattering amplitudes of the ECD-TMD and the GFP of the hGHR were calculated from their atomic coordinates as a part of our WillItFit (48) framework as outlined in previous work (26) and incorporated into the ND as rigid bodies. PDB 1EMA was used for the GFP atomic coordinates, while those of the flexible ECD-TMD were obtained from one of the back-mapped structures from one of the (full-length) GHR-GFP + POPC$_{pws10}$ simulations,

## SCIENCE ADVANCES | RESEARCH ARTICLE

but considering only the residues that comprise the SP, ECD, ECD-TMD linker, and TMD (residues −18 to 270). We used our previously developed formalism (*26*) to ensure that the TMD displace lipids in the ND and for adjusting the excess scattering lengths of the lipid-embedded residues by considering their lipid environment rather than the solvent (*26*). To gain computational speed, we described the scattering from the ICD as a Gaussian random coil where the form factor for the average scattering intensity is readily described through the analytical Debye function (*38*) parameterized through the average $R_g$ of the coil. The center of mass of the Gaussian random coil was placed one $R_g$ below the center of the lower interface between nanodisc lipid headgroups and solvent. The averaged form factor amplitude for a Gaussian random coil required for the off-diagonal cross-terms in the calculation of $I(q)$ in the above equation is given by the so-called Hammouda function (*77*), which is also a function of the $R_g$ of the coil. Hence, we used the same modeling principle as previously applied for polymer-modified micelles (*78*) to take into account the scattering contribution from a flexible ICD, which is connected to a nanodisc-embedded TMD in the model. Following a similar philosophy, the scattering from the GFP was randomly oriented and located within a certain allocated confusion volume. This way, the model captures the dynamically evolving position of the GFP with respect to the rest of the system. For the modeling of the shape of the confusion volume, we attempted to mimic the bowl-like distribution of GFP below the bilayer as observed in the CG-MD simulation of hGHR in a lipid bilayer (see Fig. 6F) by placing the GFP randomly in a thick cylindrical shell below the nanodisc (see Fig. 6A, inner and outer shell radii equal to, respectively, 1 and 1.5 times the $R_g$ of hGHR-ICD). However, we found that the actual shape of the confusion volume, whether it was bowl shaped or simply spherical and centered under the disc, only had a minor effect. The WillItFit implementation of the model can be downloaded free for use (see Code availability section in Acknowledgments).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/27/eabh3805/DC1

View/request a protocol for this paper from *Bio-protocol*.

### REFERENCES AND NOTES

1. C. Liongue, A. C. Ward, Evolution of Class I cytokine receptors. *BMC Evol. Biol.* **7**, 120 (2007).
2. K. Madsen, U. Friberg, P. Roos, S. Edén, O. Isaksson, Growth hormone stimulates the proliferation of cultured chondrocytes from rabbit ear and rat rib growth cartilage. *Nature* **304**, 545–547 (2005).
3. M. J. Waters, A. J. Brooks, Growth hormone receptor: Structure function relationships. *Horm. Res. Paediatr.* **76**, 12–16 (2011).
4. S. Yakar, C. J. Rosen, W. G. Beamer, C. L. Ackert-Bicknell, Y. Wu, J.-L. Liu, G. T. Ooi, J. Setser, J. Frystyk, Y. R. Boisclair, D. LeRoith, Circulating levels of IGF-1 directly regulate bone growth and density. *J. Clin. Invest.* **110**, 771–781 (2002).
5. Y. Chhabra, H. Y. Wong, L. F. Nikolajsen, H. Steinocher, A. Papadopulos, K. A. Tunny, F. A. Meunier, A. G. Smith, B. B. Kragelund, A. J. Brooks, M. J. Waters, A growth hormone receptor SNP promotes lung cancer by impairment of SOCS2-mediated degradation. *Oncogene* **37**, 489–501 (2018).
6. P. E. Mullis, Genetics of isolated growth hormone deficiency. *J. Clin. Res. Pediatr. Endocrinol.* **2**, 52–62 (2010).
7. A. Shimatsu, M. Nagashima, S. Hashigaki, N. Ohki, K. Chihara, Efficacy and safety of monotherapy by pegvisomant, a growth hormone receptor antagonist, in Japanese patients with acromegaly. *Endocr. J.* **63**, 337–347 (2016).
8. P. Seiffert, K. Bugge, M. Nygaard, G. W. Haxholm, J. H. Martinsen, M. N. Pedersen, L. Arleth, W. Boomsma, B. B. Kragelund, Orchestration of signaling by structural disorder in class 1 cytokine receptors. *Cell Commun. Signal* **18**, 132 (2020).
9. G. W. Haxholm, L. F. Nikolajsen, J. G. Olsen, J. Fredsted, F. H. Larsen, V. Goffin, S. F. Pedersen, A. J. Brooks, M. J. Waters, B. B. Kragelund, Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. *Biochem. J.* **468**, 495–506 (2015).
10. M. J. Waters, The growth hormone receptor. *Growth Hormon. IGF Res.* **28**, 6–10 (2016).
11. J. W. Baumgartner, C. A. Wells, C. M. Chen, M. J. Waters, The role of the WSXWS equivalent motif in growth hormone receptor function. *J. Biol. Chem.* **269**, 29094–29101 (1994).
12. J. G. Olsen, B. B. Kragelund, Who climbs the tryptophan ladder? On the structure and function of the WSXWS motif in cytokine receptors and thrombospondin repeats. *Cytokine Growth Factor Rev.* **25**, 337–341 (2014).
13. A. J. Brooks, W. Dai, M. L. O'Mara, D. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton, M. W. Parker, E. Sierecki, Y. Gambin, G. A. Gomez, K. Alexandrov, I. A. Wilson, M. Doxastakis, A. E. Mark, M. J. Waters, Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science* **344**, 1249783 (2014).
14. S. Wilmes, M. Hafer, J. Vuorio, J. A. Tucker, H. Winkelmann, S. Löchte, T. A. Stanly, K. D. Pulgar Prieto, C. Poojari, V. Sharma, C. P. Richter, R. Kurre, S. R. Hubbard, K. Christopher Garcia, I. Moraga, I. Vattulainen, I. S. Hitchcock, J. Piehler, Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations. *Science* **367**, 643–652 (2020).
15. R. J. Brown, J. J. Adams, R. A. Pelekanos, Y. Wan, W. J. McKinstry, K. Palethorpe, R. M. Seeber, T. A. Monks, K. A. Eidne, M. W. Parke, M. J. Waters, Model for growth hormone receptor activation based on subunit rotation within a receptor dimer. *Nat. Struct. Mol. Biol.* **12**, 814–821 (2005).
16. M. Sundström, T. Lundqvist, J. Rödin, L. B. Giebel, D. Milligan, G. Norstedt, Crystal structure of an antagonist mutant of human growth hormone, G120R, in complex with its receptor at 2.9 Å resolution. *J. Biol. Chem.* **271**, 32197–32203 (1996).
17. A. M. de Vos, M. Ultsch, A. A. Kossiakoff, Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science* **255**, 306–312 (1992).
18. L. Chantalat, N. D. Jones, F. Korber, J. Navaza, A. G. Pavlovsky, The crystal-structure of wild-type growth-hormone at 2.5 Angstrom resolution. *Protein Pept. Lett.* **2**, 333–340 (1995).
19. E. V. Bocharov, D. M. Lesovoy, O. V. Bocharova, A. S. Urban, K. V. Pavlov, P. E. Volynsky, R. G. Efremov, A. S. Arseniev, Structural basis of the signal transduction via transmembrane domain of the human growth hormone receptor. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 1410–1420 (2018).
20. K. Bugge, E. Papaleo, G. W. Haxholm, J. T. S. Hopper, C. V. Robinson, J. G. Olsen, K. Lindorff-Larsen, B. B. Kragelund, A combined computational and structural model of the full-length human prolactin receptor. *Nat. Commun.* **7**, 11578 (2016).
21. N. Kassem, M. M. Kassem, S. F. Pedersen, P. A. Pedersen, B. B. Kragelund, Yeast recombinant production of intact human membrane proteins with long intrinsically disordered intracellular regions for structural studies. *Biochim. Biophys. Acta Biomembr.* **1862**, 183272 (2020).
22. A. B. Ward, A. Sali, I. A. Wilson, Biochemistry. Integrative structural biology. *Science* **339**, 913–915 (2013).
23. R. Cárdenas, J. Martínez-Seoane, C. Amero, Combining experimental data and computational methods for the non-computer specialist. *Molecules* **25**, 4783 (2020).
24. A. H. Larsen, Y. Wang, S. Bottaro, S. Grudinin, L. Arleth, K. Lindorff-Larsen, Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLoS Comput. Biol.* **16**, e1007870 (2020).
25. T. H. Bayburt, Y. V. Grinkova, S. G. Sligar, Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**, 853–856 (2002).
26. S. A. R. Kynde, N. Skar-Gislinge, M. C. Pedersen, S. R. Midtgaard, J. B. Simonsen, R. Schweins, K. Mortensen, L. Arleth, Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 371–383 (2014).
27. N. Skar-Gislinge, S. A. R. Kynde, I. G. Denisov, X. Ye, I. Lenov, S. G. Sligar, L. Arleth, Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2412–2421 (2015).
28. K. Bugge, K. Lindorff-Larsen, B. B. Kragelund, Understanding single-pass transmembrane receptor signaling from a structural viewpoint-what are we missing? *FEBS J.* **283**, 4424–4451 (2016).
29. G. Kleiger, A. Saha, S. Lewis, B. Kuhlman, R. J. Deshaies, Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* **139**, 957–968 (2009).
30. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, ROSETTA3: An object-oriented software suite

SCIENCE ADVANCES | RESEARCH ARTICLE

for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

31. S. Bottaro, T. Bengtsen, K. Lindorff-Larsen, Integrating molecular simulation and experimental data: A bayesian/maximum entropy reweighting approach. *Methods Mol. Biol.* **2112**, 219–240 (2020).

32. K. Bugge, H. Steinocher, A. J. Brooks, K. Lindorff-Larsen, B. B. Kragelund, Exploiting hydrophobicity for efficient production of transmembrane helices for structure determination by NMR spectroscopy. *Anal. Chem.* **87**, 9126–9131 (2015).

33. Y. Shen, A. Bax, Identification of helix capping and b-turn motifs from NMR chemical shifts. *J. Biomol. NMR* **52**, 211–232 (2012).

34. P. Güntert, Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).

35. Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, A. Bax, Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4685–4690 (2008).

36. A. Khondker, A. K. Dhaliwal, S. Saem, A. Mahmood, C. Fradin, J. Moran-Mirabal, M. C. Rheinstädter, Membrane charge and lipid packing determine polymyxin-induced membrane damage. *Commun. Biol.* **2**, 67 (2019).

37. J. Bürck, P. Wadhwani, S. Fanghänel, A. S. Ulrich, Oriented circular dichroism: A method to characterize membrane-active peptides in oriented lipid bilayers. *Acc. Chem. Res.* **49**, 184–192 (2016).

38. P. Debye, Molecular-weight determination by light scattering. *J. Phys. Colloid Chem.* **51**, 18–32 (1947).

39. S. Bibow, Y. Polyhach, C. Eichmann, C. N. Chi, J. Kowal, S. Albiez, R. A. McLeod, H. Stahlberg, G. Jeschke, P. Güntert, R. Riek, Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I. *Nat. Struct. Mol. Biol.* **24**, 187–193 (2017).

40. N. Skar-Gislinge, J. B. Simonsen, K. Mortensen, R. Feidenhans'l, S. G. Sligar, B. Lindberg Møller, T. Bjørnholm, L. Arleth, Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the lipids and the protein. *J. Am. Chem. Soc.* **132**, 13713–13722 (2010).

41. N. Skar-Gislinge, N. T. Johansen, R. Høiberg-Nielsen, L. Arleth, Comprehensive study of the self-assembly of phospholipid nanodiscs: What determines their shape and stoichiometry? *Langmuir* **34**, 12569–12582 (2018).

42. N. T. Johansen, M. C. Pedersen, L. Porcar, A. Martel, L. Arleth, Introducing SEC-SANS for studies of complex self-organized biological systems. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 1178–1191 (2018).

43. G. Rouser, A. N. Siakotos, S. Fleischer, Quantitative analysis of phospholipids by thin-layer chromatography and phosphorus analysis of spots. *Lipids* **1**, 85–86 (1966).

44. P. A. Harding, X. Z. Wang, B. Kelder, S. Souza, S. Okada, J. J. Kopchick, In vitro mutagenesis of growth hormone receptor Asn-linked glycosylation sites. *Mol. Cell. Endocrinol.* **106**, 171–180 (1994).

45. F. B. Bjørkskov, S. L. Krabbe, C. N. Nurup, J. W. Missel, M. Spulber, J. Bomholt, K. Molbaek, C. Helix-Nielsen, K. Gotfryd, P. Gourdon, P. A. Pedersen, Purification and functional comparison of nine human Aquaporins produced in Saccharomyces cerevisiae for the purpose of biophysical characterization. *Sci. Rep.* **7**, 16899 (2017).

46. H. Wennbo, M. Gebre-Medhin, A. Gritli-Linde, C. Ohlsson, O. G. P. Isaksson, J. Törnell, Activation of the prolactin receptor but not the growth hormone receptor is important for induction of mammary tumors in transgenic mice. *J. Clin. Invest.* **100**, 2744–2751 (1997).

47. J.-L. K. Kouadio, J. R. Horn, G. Pal, A. A. Kossiakoff, Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. *J. Biol. Chem.* **280**, 25524–25532 (2005).

48. M. C. Pedersen, L. Arleth, K. Mortensen, WillItFit : A framework for fitting of constrained models to small-angle scattering data. *J. Appl. Crystallogr.* **46**, 1894–1898 (2013).

49. N. Kucerka, S. Tristram-Nagle, J. F. Nagle, Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **208**, 193–202 (2005).

50. Y. Minezaki, K. Homma, K. Nishikawa, Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *J. Mol. Biol.* **368**, 902–913 (2007).

51. M. Kjaergaard, B. B. Kragelund, Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.* **74**, 3205–3224 (2017).

52. A. J. Brooks, F. Dehkhoda, B. B. Kragelund, in *Principles of Endocrinology and Hormone Action*, A. Belfiore, D. LeRoith, Eds. (Springer International Publishing, Cham, 2018), pp. 157–185; https://doi.org/10.1007/978-3-319-44675-2_8.

53. D. Ben-Avraham, D. R. Govindaraju, T. Budagov, D. Fradin, P. Durda, B. Liu, S. Ott, D. Gutman, L. Sharvit, R. Kaplan, P. Bougnères, A. Reiner, A. R. Shuldiner, P. Cohen, N. Barzilai, G. Atzmon, The GH receptor exon 3 deletion is a marker of male-specific exceptional longevity associated with increased GH sensitivity and taller stature. *Sci. Adv.* **3**, e1602025 (2017).

54. M. Gouw, S. Michael, H. Sámano-Sánchez, M. Kumar, A. Zeke, B. Lang, B. Bely, L. B. Chemes, N. E. Davey, Z. Deng, F. Diella, C. M. Gürth, A. K. Huber, S. Kleinsorg, L. S. Schlegel, N. Palopoli, K. V. Roey, B. Altenberg, A. Reményi, H. Dinkel, T. J. Gibson, The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* **46**, D428–D434 (2018).

55. M. Kaplan, S. Narasimhan, C. de Heus, D. Mance, S. van Doorn, K. Houben, D. Popov-Čeleketić, R. Damman, E. A. Katrukha, P. Jain, W. J. C. Geerts, A. J. R. Heck, G. E. Folkers, L. C. Kapitein, S. Lemeer, P. M. P. van Bergen En Henegouwen, M. Baldus, EGFR dynamics change during activation in native membranes as revealed by NMR. *Cell* **167**, 1241–1251.e11 (2016).

56. X. Wang, C. J. Darus, B. C. Xu, J. J. Kopchick, Identification of growth hormone receptor (GHR) tyrosine residues required for GHR phosphorylation and JAK2 and STAT5 activation. *Mol. Endocrinol.* **10**, 1249–1260 (1996).

57. B. P. Cormack, G. Bertram, M. Egerton, N. A. R. Gow, S. Falkow, A. J. P. Brown, Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in Candida albicans. *Microbiology* **143**, 303–311 (1997).

58. A. Khondker, D. J. Malenfant, A. K. Dhaliwal, M. C. Rheinstädter, Carbapenems and lipid bilayers: Localization, partitioning, and energetics. *ACS Infect. Dis.* **4**, 926–935 (2018).

59. S. Himbert, M. J. Blacker, A. Kihm, Q. Pauli, A. Khondker, K. Yang, S. Sinjari, M. Johnson, J. Juhasz, C. Wagner, H. D. H. Stöver, M. C. Rheinstädter, Hybrid erythrocyte liposomes: Functionalized red blood cell membranes for molecule encapsulation. *Adv. Biosyst.* **4**, 1900185 (2020).

60. R.-C. Bider, T. Lluka, S. Himbert, A. Khondker, S. M. Qadri, W. P. Sheffield, M. C. Rheinstädter, Stabilization of lipid membranes through partitioning of the blood bag plasticizer di-2-ethylhexyl phthalate (DEHP). *Langmuir* **36**, 11899–11907 (2020).

61. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).

62. K. Reichel, O. Fisette, T. Braun, O. F. Lange, G. Hummer, L. V. Schäfer, Systematic evaluation of CS-Rosetta for membrane protein structure prediction with sparse NOE restraints. *Proteins* **85**, 812–826 (2017).

63. I. G. Denisov, Y. V. Grinkova, A. A. Lazarides, S. G. Sligar, Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).

64. M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).

65. M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, D. I. Svergun, New developments in the *ATSAS* program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).

66. A. Jordan, M. Jacques, C. Merrick, J. Devos, V. T. Forsyth, L. Porcar, A. Martel, SEC-SANS: Size exclusion chromatography combined in situ with small-angle neutron scattering. *J. Appl. Crystallogr.* **49**, 2015–2020 (2016).

67. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

68. J. Koehler Leman, R. Bonneau, A novel domain assembly routine for creating full-length models of membrane proteins from known domain structures. *Biochemistry* **57**, 1939–1944 (2018).

69. P. C. T. Souza, S. J. S. Marrink, Martini 3 - Open Beta-Release (2020); http://cgmartini.nl.

70. Y. Qi, H. I. Ingólfsson, X. Cheng, J. Lee, S. J. Marrink, W. Im, CHARMM-GUI martini maker for coarse-grained simulations with the martini force field. *J. Chem. Theory Comput.* **11**, 4486–4494 (2015).

71. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

72. D. H. de Jong, S. Baoukina, H. I. Ingólfsson, S. J. Marrink, Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **199**, 1–7 (2016).

73. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

74. A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).

75. PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).

76. S. Grudinin, M. Garkavenko, A. Kazennov, Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 449–464 (2017).

77. B. Hammouda, Structure factor for starburst dendrimers. *J. Polym. Sci. Part B Polym. Phys.* **30**, 1387–1390 (1992).

78. J. S. Pedersen, M. C. Gerstenberg, Scattering form factor of block copolymer micelles. *Macromolecules* **29**, 1363–1365 (1996).

# SCIENCE ADVANCES | RESEARCH ARTICLE

79.  Y. Wei, A. A. Thyparambil, R. A. Latour, Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim. Biophys. Acta Proteins Proteomics* **1844**, 2331–2337 (2014).

80.  R. F. Sommese, S. Sivaramakrishnan, R. L. Baldwin, J. A. Spudich, Helicity of short E-R/K peptides. *Protein Sci.* **19**, 2001–2005 (2010).

81.  S. Hansen, BayesApp: A web site for indirect transformation of small-angle scattering data. *J. Appl. Crystallogr.* **45**, 566–567 (2012).

82.  E. L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).

83.  L. Käll, A. Krogh, E. L. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).

84.  L. Käll, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).

85.  D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538–544 (2007).

86.  UniProt Consortium, UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).

87.  W. Zheng, R. B. Best, An extended guinier analysis for intrinsically disordered proteins. *J. Mol. Biol.* **430**, 2540–2553 (2018).

88.  J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, T. R. Sosnick, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).

89.  J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, K. W. Plaxco, Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491–12496 (2004).

90.  J. A. Marsh, J. D. Forman-kay, Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).

91.  N. T. Johansen, F. G. Tidemand, T. T. T. N. Nguyen, K. D. Rand, M. C. Pedersen, L. Arleth, Circularized and solubility-enhanced MSPs facilitate simple and high-yield production of stable nanodiscs for studies of membrane proteins in solution. *FEBS J.* **286**, 1734–1751 (2019).

**Citation:** N. Kassem, R. Araya-Secchi, K. Bugge, A. Barclay, H. Steinocher, A. Khondker, Y. Wang, A. J. Lenard, J. Bürck, C. Sahin, A. S. Ulrich, M. Landreh, M. C. Pedersen, M. C. Rheinstädter, P. A. Pedersen, K. Lindorff-Larsen, L. Arleth, B. B. Kragelund, Order and disorder—An integrative structure of the full-length human growth hormone receptor. *Sci. Adv.* **7**, eabh3805 (2021).

# Science Advances

## Order and disorder—An integrative structure of the full-length human growth hormone receptor

Noah Kassem, Raul Araya-Secchi, Katrine Bugge, Abigail Barclay, Helena Steinocher, Adree Khondker, Yong Wang, Aneta J. Lenard, Jochen Brck, Cagla Sahin, Anne S. Ulrich, Michael Landreh, Martin Cramer Pedersen, Maikel C. Rheinstdter, Per Amstrup Pedersen, Kresten Lindorff-Larsen, Lise Arleth, and Birthe B. Kragelund

**View the article online**
https://www.science.org/doi/10.1126/sciadv.abh3805
**Permissions**
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service

# Science Advances

**AAAS**

## Supplementary Materials for

## Order and disorder—An integrative structure of the full-length human growth hormone receptor

Noah Kassem, Raul Araya-Secchi, Katrine Bugge, Abigail Barclay, Helena Steinocher, Adree Khondker, Yong Wang,
Aneta J. Lenard, Jochen Bürck, Cagla Sahin, Anne S. Ulrich, Michael Landreh, Martin Cramer Pedersen,
Maikel C. Rheinstädter, Per Amstrup Pedersen, Kresten Lindorff-Larsen*, Lise Arleth*, Birthe B. Kragelund*

*Corresponding author. Email: bbk@bio.ku.dk (B.B.K.); lindorff@bio.ku.dk (K.L.-L.);
arleth@nbi.ku.dk (L.A.)

**The PDF file includes:**

> Figs. S1 to S5
> Tables S1 to S3
> Supplementary Text
> References

**Other Supplementary Material for this manuscript includes the following:**

(available at advances.sciencemag.org/cgi/content/full/7/27/eabh3805/DC1)

> Movie S1

## Supplementary Figures

Supplementary Fig. S1
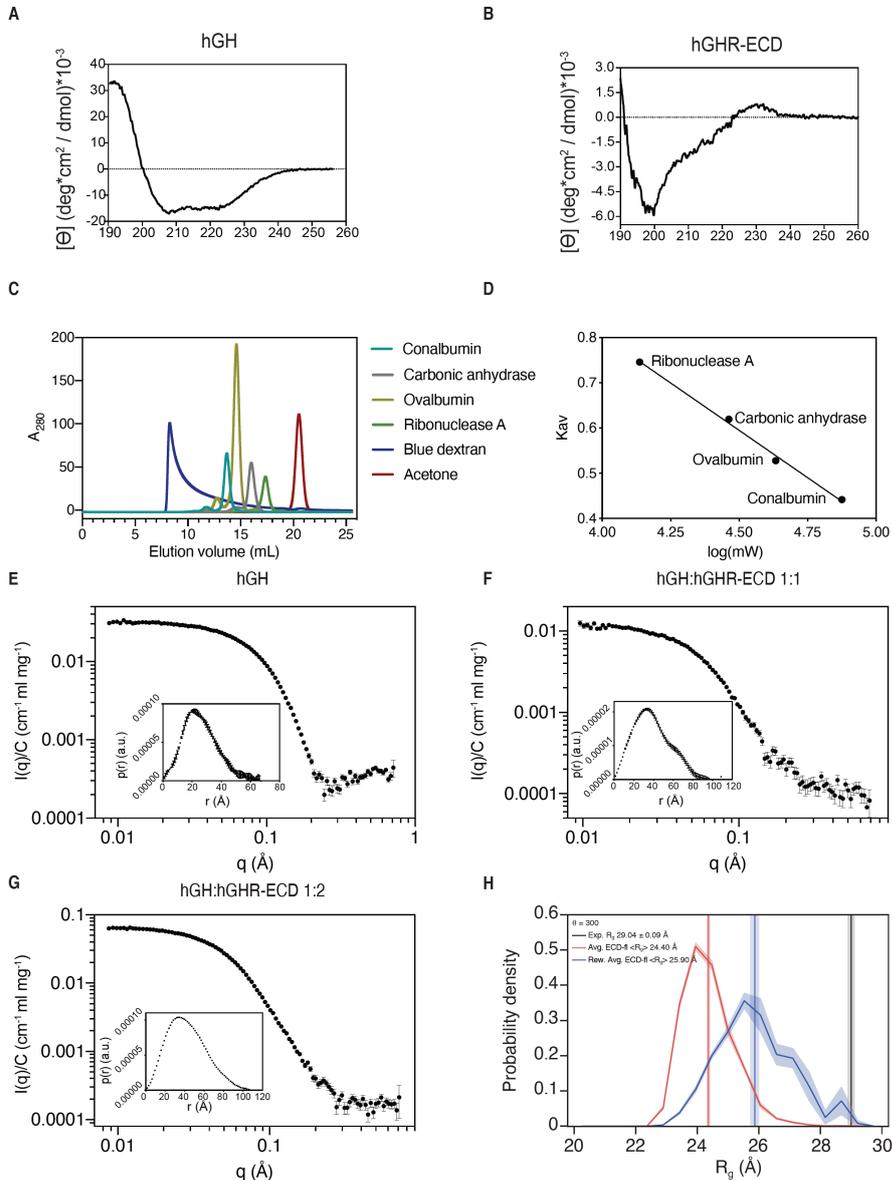


**Suppl. Fig. S1: Structural analyses of hGH and hGHR-ECD.** (A) Far-UV CD spectrum of hGH with a helicity calculated to be 32%(*79*) and 38%(*80*) (44% from the crystal structure PDB 1HGU). (B) Far-UV-CD spectrum of hGHR-ECD. (C) Six proteins standards; conalbumin, carbonic anhydrase, ovalbumin, ribonuclease A, blue dextran and acetone were run on a Superdex 200 increase 10/300 (GE Healthcare). (D) Partition coefficients were calculated for ribonuclease A, carbonic anhydrase, ovalbumin and conalbumin. Linear regression was performed, where *Kav* = -

0.4187*log($MW$)+2.4779, $R^2$ = 0.98. SAXS curves of (E) hGH, (F) hGH:hGHR-ECD 1:1, and (G) hGH:hGHR-ECD 1:2. Inserts: Pair-distance distribution functions, *p(r)'s*, obtain through Bayesian Indirect Fourier-Transformation(*81*). (H) $R_g$ distribution of hGHR-ECD models before (red) and after (blue) reweighting with BME. The vertical lines indicate the experimental $R_g$ values obtained from hGHR-ECD SAXS data (black) and the ensemble average $R_g$ from the hGHR-ECD models before (red) and after reweighting (blue). We note that the average (dry) $R_g$ of 26 Å, is in good agreement with the experimental value of 29 Å, considering an expected 5-10% difference due to solvation shell scattering in the experiments.

Supplementary Fig. S2
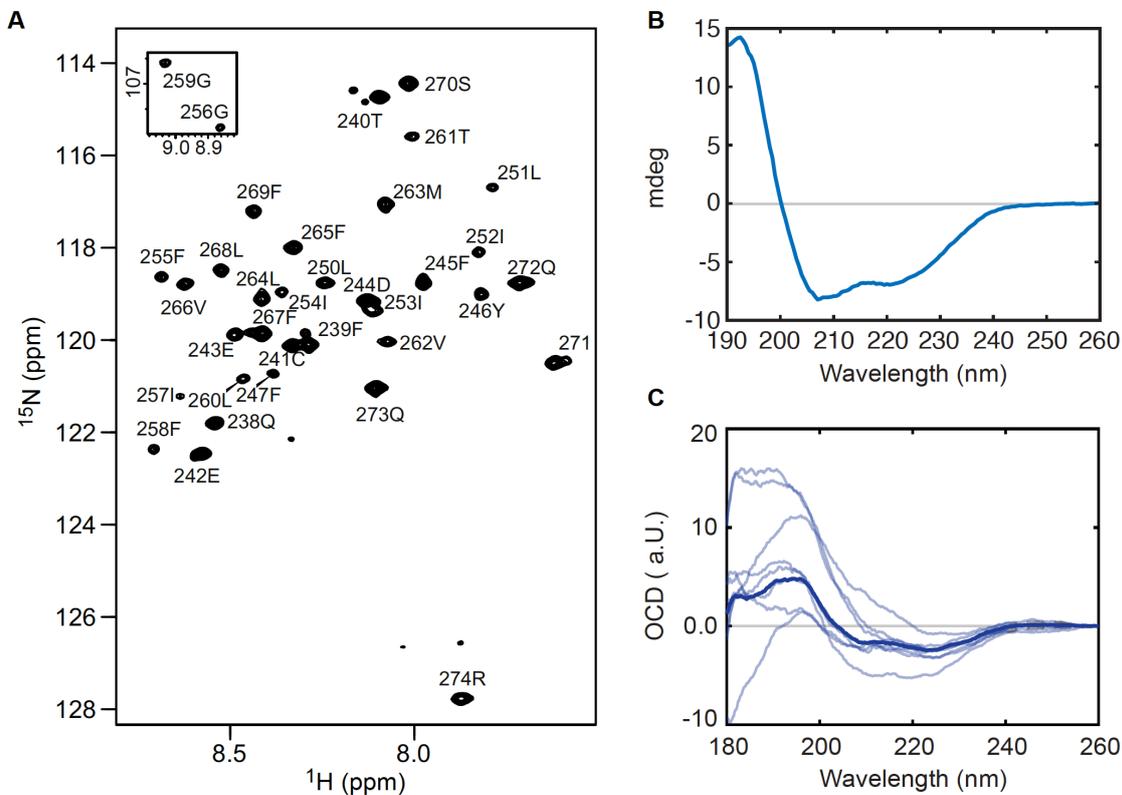


**Suppl. Fig. S2: Structural analysis of hGHR-TMD.** (A) Assigned $^1$H,$^{15}$N-HSQC spectrum at 37°C of 1 mM $^{13}$C,$^{15}$N-hGHR-TMD in 210 mM DHPC. (B) Far-UV CD spectrum of 10 μM hGHR-TMD in 2 mM DHPC. (C) OCD spectra of 6 μg hGHR-TMD in 50 μg POPC measured from 8 different angles (light blue) and the average spectrum (blue).

Prediction of the transmembrane regions of the hGHRTMD shown in Fig. 2A was done using TMHMM(*82*) (light blue), Phobius(*83, 84*) (red), METSAT-SVM(*85*) (purple), and Uniprot annotations(*86*) (green).

Supplementary Fig. S3
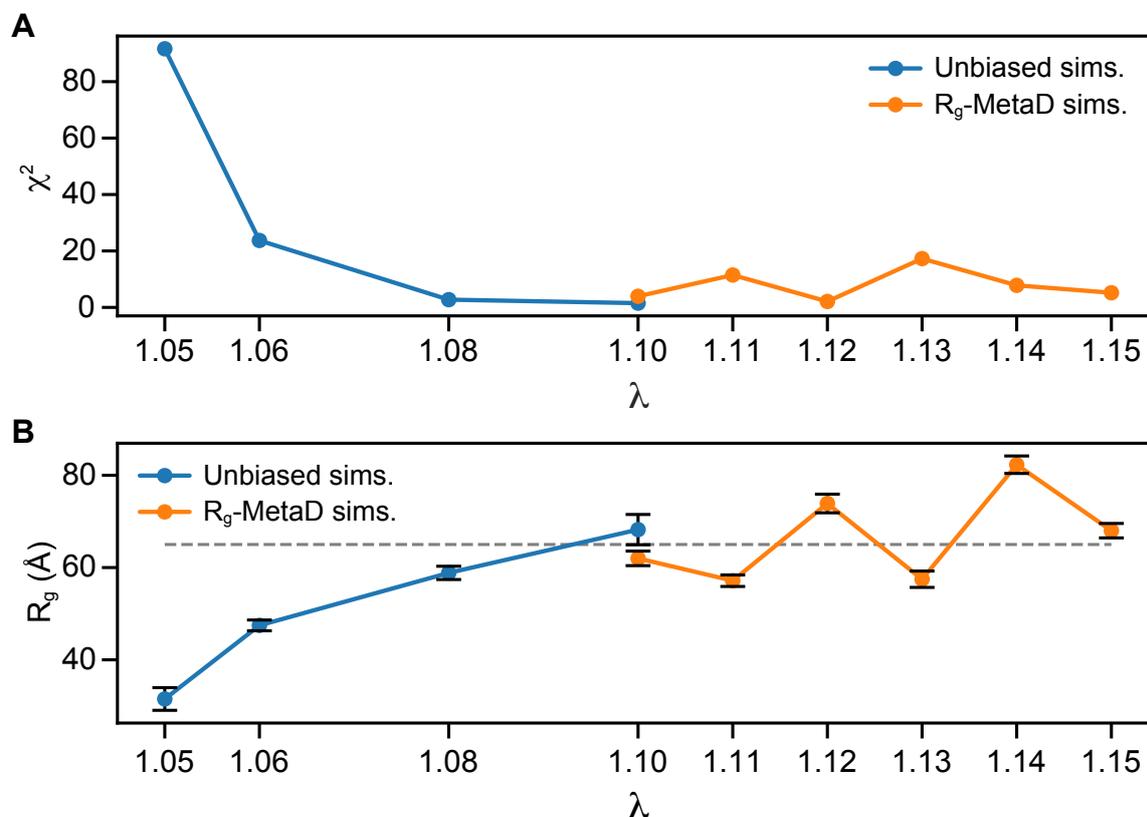


**Suppl. Fig. S3: Calibration of the protein-water interactions strength on the Martini3 forcefield for the simulation of hGHR-ICD.** The strength of the protein-water interactions was varied by a factor ($\lambda$) in the range between 5 and 10% in unbiased MD simulations (5%: 5 µs; 6%: 3 µs; 8%: 3µs and 10%: 5µs) and 10%–15% in metadynamics simulations (all of them ~ 10 µs) to improve the sampling. Back-mapped (CG → all-atom) conformations (1/ns) from each simulation were used to fit the SAXS data of hGHR-ICD 1.1 mg/mL as described in the *Materials and Methods* section. (A) Average $\chi^2$ obtained from fitting to the SAXS data and (B) Average $R_g$ measured from the conformations taken from the unbiased (blue) and metadynamics (orange) simulations. The dashed gray line in B corresponds to the hGHR-ICD $R_g$ obtained from the SAXS data.

Supplementary Fig. S4

**A**

| Protein | POPC lipid |
|---|---|
| Monomer: hGHR(MSP1D1) (F3) | 122 ± 17 |
| Dimer: hGHR(MSP1D1) (F1) | 115 ± 19 |
| MSP1D1* | 120 ± 10 |

**B**



**C** **D**
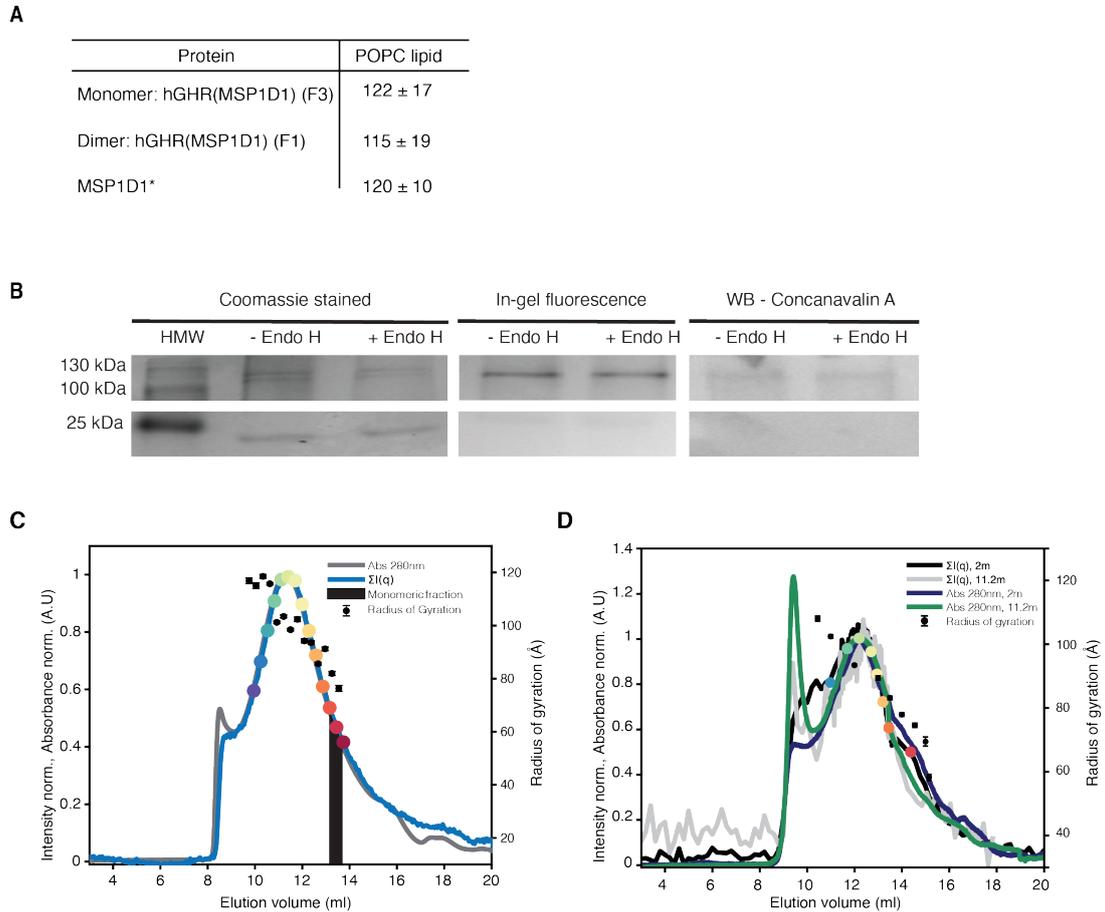


**Suppl. Fig. S4: Biochemical and biophysical analyses of hGHR in POPC loaded nanodiscs**. (A) POPC/hGHR-loaded MSP1D1 ratios were derived from fraction 1 (F1, mostly dimer) and 3 (F3, monomer) by phosphate analysis (see *Materials and Methods*). (B) hGHR-loaded MSP1D1 was treated with Endo-H overnight (see *Materials and Methods*). Endo-H (+Endo-H) and untreated (-Endo-H) were separated by SDS-PAGE and analysed by in-gel fluorescence, Coomassie staining, and western blotting using horse-radish peroxidase conjugated Concanavalin A that binds mannose residues. (C) SEC-SAXS data for hGHR in MSP1D1 with POPC. Absorption at 280 nm (black line), scaled to the total scattering intensities of individual frames (blue line). Calculated $R_g$ values are plotted as black dots. The black area indicates the fraction chosen for further analysis of the monomer. The SAXS data corresponding to this fraction is plotted in Figure 5B. (D) SEC frames corresponding to the two SEC-SANS data sets, one for each detector setting. Absorption at 280 nm (navy and green lines), scaled to the total scattering intensities obtained for the two settings, (black and grey lines). Calculated $R_g$ values are plotted as black circles.

Supplementary Fig. S5
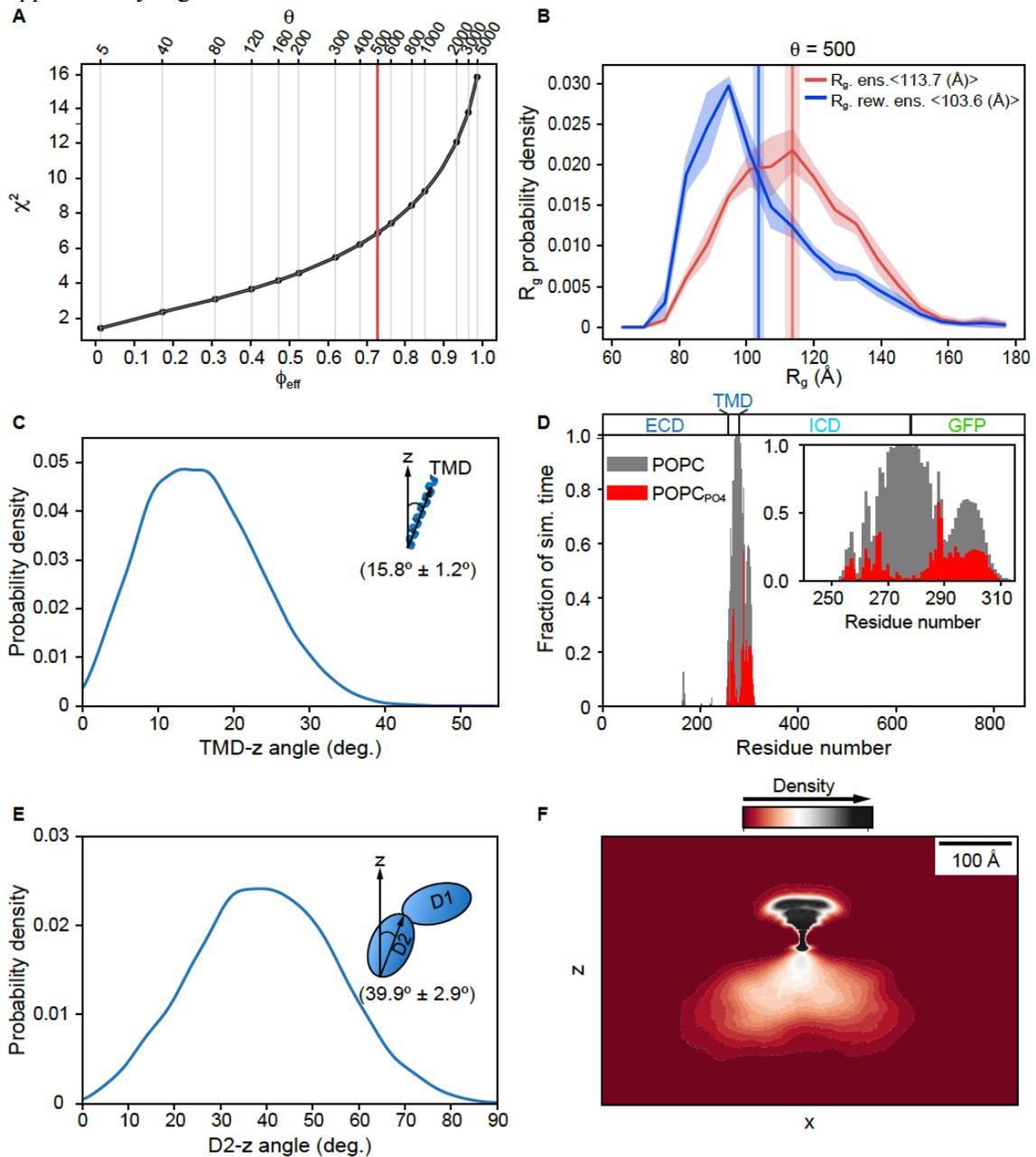


**Suppl. Fig. 5. Additional insights of hGHR-GFP relative orientations from the CG-MD simulation of hGHR-GFP in a POPC bilayer.** (A) Selection of the hyperparameter θ for the reweighting against the SAXS data using BME. The solid red line shows the value θ = 500 selected for further processing. (B) $R_g$ distribution of the hGHR-GFP conformations before (red) and after (blue) reweighting with BME. The vertical lines indicate the ensemble average $R_g$ from the conformations before (red) and after reweighting (blue). (C) Probability density of the angle between the principal axis of the TMD and the z-axis (perpendicular to the membrane plane) after re-weighting

against the SAXS data. (D) Protein-lipid contact profiles. A contact is defined when the distance between a BB bead of the protein and a $PO_4$ bead (red) or any bead (gray) of POPC is $\leq$ 7 Å. The inset shows a detail of the profile highlighting the interactions between the intracellular juxtamembrane (ICJM) and BOX1 regions of GHR-ICD with the lipids. (E) Probability density of the angle between the principal axis of the D1 domain and the z-axis after re-weighting against the SAXS data. (F) Average volumetric protein (hGHR-GFP) density map projected on the *xz* plane. This map provides a visualization of the large volume that the ICD+GFP can occupy below the membrane in contrast with the ECD. In all panels the measurements were performed considering the last 20 μs of the hGHR-GFP+POPC$_{pws10}$ simulation, hence the first μs was discarded as equilibration time.

**Supplementary Tables**

**Suppl. Table S1.** Molecular weights of the hGHR-ECD and hGHR-ICD from the SAXS data.

| Protein sample | $I(0)$ (1/(mg/ml)) | Concentration (mg/mL) | Theoretical molecular weight (kDa) | Calculated molecular mass (kDa) |
|---|---|---|---|---|
| hGHR-ECD | 0.083 | 3.45 | 28.1 | 33.3 |
| hGH | 0.032 | 1.83 | 24.2 | 22.1 |
| hGH:hGHR-ECD1:2 | 0.067 | 1.28 | 72.5 | 78.2 |
| hGH:hGHR-ECD1:1 | 0.012 | 0.30 | 55.4 | 50.2 |
| hGHR-ICD | 0.037 | 1.14 | 38.6 | 45.0 |
| hGHR-ICD-GFP-H$_{10}$ | 0.05 | 0.90 | 68.0 | 75.8 |

**Suppl. Table S2.** Dimensions of hGHR-ICD (1.14 mg/ml sample) – Comparisons of different analysis approaches

| Analysis | $R_g$ (Å) | $R_h$ (Å) | $\nu$ | Reference |
|---|---|---|---|---|
| **SAXS** | | | | |
| Guinier | 64.5± 1.3 | | **n.a.** | |
| IFT | 62.0 ± 0.1 | | **n.a.** | (*81*) |
| GRC model fit | 68 ± 4 | | 0.5** | (*38*) |
| Extended Guinier | 65.9 | | 0.618±0.002 | (*87*) |
| Sosnick | 64.9 ± 0.4 | | 0.603±0.002 | (*88*) |
| Kohn* | 64.2 | | 0.598** | (*89*) |
| **NMR** | | | | |
| Diffusion NMR | | 43.9±0.5 | **n.a.** | |
| Forman-Kay* | | 49.2 | 0.509** | (*90*) |

*: Empirical prediction based on the number of residues of hGHR-ICD (352).
**: Parameter prefixed in the model.
SAXS analysis is based on the hGHR-ICD sample of 1.14 mg/ml (see main text).
See main text and Materials and Methods for description of the NMR conditions.

**Suppl. Table S3:** SAXS-refined fit parameters of the hGHR-GFP in nanodisc

| Fitting Parameters | |
|---|---|
| $A_{head}$ ($\text{Å}^2$) | $63 \pm 11$ |
| $R_g$ of coil (Å) | $73 \pm 13$ |
| $v_{POPC}$ ($\text{Å}^3$) | $1220 \pm 30$ |
| $v_{MSP}$ ($\text{Å}^3$) | $27\,900 \pm 2700$ |
| $v_{GHR\text{-}GFP}$ ($\text{Å}^3$) | $119\,000 \pm 2300$ |
| $H_{belt}$ (Å) * | 25.8 |
| $\varepsilon$ * | 1.4 |
| $N_{POPC}$ * | 122 |
| Roughness (Å) * | 6.0 |
| **Deduced parameters** | |
| $H_{ND}$ (Å) | 39 |
| $D_{minor}$ (Å) | 30 |
| $D_{major}$ (Å) | 42 |
| $d_{belt}$ (Å) | 8.5 |

*Parameter not fitted.

$A_{head}$: area taken up by one POPC headgroup.

$R_g$ of coil: ensemble average $R_g$ of the Gaussian random coil representing the ICD.

$v_{POPC}$: partial specific molecular volume of one POPC molecule.

$v_{GHR\text{-}GFP}$: partial specific volume of the hGHR-GFP.

$H_{belt}$: height of the MSP-belt.

$\varepsilon$: axis ratio of the elliptical phospholipid bilayer patch ($D_{major}$ / $D_{minor}$) .

$N_{POPC}$: the average number of POPC per nanodisc.

Roughness: Interface rougness correcting for the fact the interfaces are not perfectly smooth.

$H_{ND}$: total height of the phospholipid bilayer.

$D_{minor}$ and $D_{major}$: major and minor diameter of the phospholipid bilayer.

$D_{belt}$: thickness of the MSP belt.

**Supplementary text**

**Detailed account of the results of the semi-analytical model fit of hGHR-GFP in POPC nanodiscs**

To refine our model against the data from full-length GHR-GFP in nanodiscs using a semi-analytical model, a total of six parameters were fitted: the area per one POPC headgroup; the ensemble average $R_g$ of the Gaussian random coil representing the disordered intracellular domain; the partial specific molecular volumes ($v$) of the POPC lipids ($v_{POPC}$), the membrane scaffold proteins ($v_{MSP}$) and the hGHR-GFP membrane protein ($v_{GHR-GFP}$), respectively and finally a small constant background in order to correct for small errors in the background subtraction. $v_{POPC}$ and $v_{GHR-GFP}$ were allowed to vary within a few percent of their pre-estimated values. The initial value for $v_{POPC}$ was taken from literature as 1260 Å[76] whereas the initial partial specific molecular volumes of one MSP and the ICD of the GHR, $v_{MSP}$ and $v_{GHR-ICD}$, were calculated using an average mass density of proteins of 1.35 cm$^2$/g to be 27 100 Å$^3$ and 47 300 Å$^3$, respectively. The partial specific molecular volumes of the remaining part of the GHR, *i.e.* the GHR-ECD-TMD and the GFP were calculated as part of *WillItFit*[48], giving a total volume of $v_{GHR-GFP}$=116 600Å$^3$ for the full-length GHR-GFP.

In the upper part of supplementary Table S3, the parameter values of the six main parameters are listed along with the fixed model parameters. In the lower part of the table, parameters deduced from the fit parameters due to the systematic use of molecular constraints are listed. Aside from the listed parameters, a small constant value was added to the model in order to correct for errors in the background subtraction.

In the fits, the belt-height was fixed to 25.8 following the same arguments as previously outlined[91]. The value for the interface roughness was fixed to 6Å and the value for the axis ratio of the elliptical cylinder representing the phospholipid bilayer was fixed to 1.4, both in line with the results from refinements of the nanodisc model on comparable nanodisc systems [26,27,91].

## REFERENCES AND NOTES

1. C. Liongue, A. C. Ward, Evolution of Class I cytokine receptors. *BMC Evol. Biol.* **7**, 120 (2007).

2. K. Madsen, U. Friberg, P. Roos, S. Edén, O. Isaksson, Growth hormone stimulates the proliferation of cultured chondrocytes from rabbit ear and rat rib growth cartilage. *Nature* **304**, 545–547 (2005).

3. M. J. Waters, A. J. Brooks, Growth hormone receptor: Structure function relationships. *Horm. Res. Paediatr.* **76**, 12–16 (2011).

4. S. Yakar, C. J. Rosen, W. G. Beamer, C. L. Ackert-Bicknell, Y. Wu, J.-L. Liu, G. T. Ooi, J. Setser, J. Frystyk, Y. R. Boisclair, D. LeRoith, Circulating levels of IGF-1 directly regulate bone growth and density. *J. Clin. Invest.* **110**, 771–781 (2002).

5. Y. Chhabra, H. Y. Wong, L. F. Nikolajsen, H. Steinocher, A. Papadopulos, K. A. Tunny, F. A. Meunier, A. G. Smith, B. B. Kragelund, A. J. Brooks, M. J. Waters, A growth hormone receptor SNP promotes lung cancer by impairment of SOCS2-mediated degradation. *Oncogene* **37**, 489–501 (2018).

6. P. E. Mullis, Genetics of isolated growth hormone deficiency. *J. Clin. Res. Pediatr. Endocrinol.* **2**, 52–62 (2010).

7. A. Shimatsu, M. Nagashima, S. Hashigaki, N. Ohki, K. Chihara, Efficacy and safety of monotherapy by pegvisomant, a growth hormone receptor antagonist, in Japanese patients with acromegaly. *Endocr. J.* **63**, 337–347 (2016).

8. P. Seiffert, K. Bugge, M. Nygaard, G. W. Haxholm, J. H. Martinsen, M. N. Pedersen, L. Arleth, W. Boomsma, B. B. Kragelund, Orchestration of signaling by structural disorder in class 1 cytokine receptors. *Cell Commun. Signal* **18**, 132 (2020).

9. G. W. Haxholm, L. F. Nikolajsen, J. G. Olsen, J. Fredsted, F. H. Larsen, V. Goffin, S. F. Pedersen, A. J. Brooks, M. J. Waters, B. B. Kragelund, Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. *Biochem. J.* **468**, 495–506 (2015).

10. M. J. Waters, The growth hormone receptor. *Growth Hormon. IGF Res.* **28**, 6–10 (2016).

11. J. W. Baumgartner, C. A. Wells, C. M. Chen, M. J. Waters, The role of the WSXWS equivalent motif in growth hormone receptor function. *J. Biol. Chem.* **269**, 29094–29101 (1994).

12. J. G. Olsen, B. B. Kragelund, Who climbs the tryptophan ladder? On the structure and function of the WSXWS motif in cytokine receptors and thrombospondin repeats. *Cytokine Growth Factor Rev.* **25**, 337–341 (2014).

13. A. J. Brooks, W. Dai, M. L. O'Mara, D. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton, M. W. Parker, E. Sierecki, Y. Gambin, G. A. Gomez, K. Alexandrov, I. A. Wilson, M. Doxastakis, A. E. Mark, M. J. Waters, Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science* **344**, 1249783 (2014).

14. S. Wilmes, M. Hafer, J. Vuorio, J. A. Tucker, H. Winkelmann, S. Löchte, T. A. Stanly, K. D. Pulgar Prieto, C. Poojari, V. Sharma, C. P. Richter, R. Kurre, S. R. Hubbard, K. Christopher Garcia, I. Moraga, I. Vattulainen, I. S. Hitchcock, J. Piehler, Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations. *Science* **367**, 643–652 (2020).

15. R. J. Brown, J. J. Adams, R. A. Pelekanos, Y. Wan, W. J. McKinstry, K. Palethorpe, R. M. Seeber, T. A. Monks, K. A. Eidne, M. W. Parke, M. J. Waters, Model for growth hormone receptor activation based on subunit rotation within a receptor dimer. *Nat. Struct. Mol. Biol.* **12**, 814–821 (2005).

16. M. Sundström, T. Lundqvist, J. Rödin, L. B. Giebel, D. Milligan, G. Norstedt, Crystal structure of an antagonist mutant of human growth hormone, G120R, in complex with its receptor at 2.9 Å resolution. *J. Biol. Chem.* **271**, 32197–32203 (1996).

17. A. M. de Vos, M. Ultsch, A. A. Kossiakoff, Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science* **255**, 306–12 (1992).

18. L. Chantalat, N. D. Jones, F. Korber, J. Navaza, A. G. Pavlovsky, The crystal-structure of wild-type growth-hormone at 2.5 Angstrom resolution. *Protein Pept. Lett.* **2**, 333–340 (1995).

19. E. V. Bocharov, D. M. Lesovoy, O. V. Bocharova, A. S. Urban, K. V. Pavlov, P. E. Volynsky, R. G. Efremov, A. S. Arseniev, Structural basis of the signal transduction via transmembrane domain of the human growth hormone receptor. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 1410–1420 (2018).

20. K. Bugge, E. Papaleo, G. W. Haxholm, J. T. S. Hopper, C. V. Robinson, J. G. Olsen, K. Lindorff-Larsen, B. B. Kragelund, A combined computational and structural model of the full-length human prolactin receptor. *Nat. Commun.* **7**, 11578 (2016).

21. N. Kassem, M. M. Kassem, S. F. Pedersen, P. A. Pedersen, B. B. Kragelund, Yeast recombinant production of intact human membrane proteins with long intrinsically disordered intracellular regions for structural studies. *Biochim. Biophys. Acta Biomembr.* **1862**, 183272 (2020).

22. A. B. Ward, A. Sali, I. A. Wilson, Biochemistry. Integrative structural biology. *Science* **339**, 913–5 (2013).

23. R. Cárdenas, J. Martínez-Seoane, C. Amero, Combining experimental data and computational methods for the non-computer specialist. *Molecules* **25**, 4783 (2020).

24. A. H. Larsen, Y. Wang, S. Bottaro, S. Grudinin, L. Arleth, K. Lindorff-Larsen, Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLoS Comput. Biol.* **16**, e1007870 (2020).

25. T. H. Bayburt, Y. V. Grinkova, S. G. Sligar, Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**, 853–856 (2002).

26. S. A. R. Kynde, N. Skar-Gislinge, M. C. Pedersen, S. R. Midtgaard, J. B. Simonsen, R. Schweins, K. Mortensen, L. Arleth, Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 371–383 (2014).

27. N. Skar-Gislinge, S. A. R. Kynde, I. G. Denisov, X. Ye, I. Lenov, S. G. Sligar, L. Arleth, Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2412–21 (2015).

28. K. Bugge, K. Lindorff-Larsen, B. B. Kragelund, Understanding single-pass transmembrane receptor signaling from a structural viewpoint-what are we missing? *FEBS J.* **283**, 4424–4451 (2016).

29. G. Kleiger, A. Saha, S. Lewis, B. Kuhlman, R. J. Deshaies, Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* **139**, 957–68 (2009).

30. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

31. S. Bottaro, T. Bengtsen, K. Lindorff-Larsen, Integrating molecular simulation and experimental data: A bayesian/maximum entropy reweighting approach. *Methods Mol. Biol.* **2112**, 219–240 (2020).

32. K. Bugge, H. Steinocher, A. J. Brooks, K. Lindorff-Larsen, B. B. Kragelund, Exploiting hydrophobicity for efficient production of transmembrane helices for structure determination by NMR spectroscopy. *Anal. Chem.* **87**, 9126–9131 (2015).

33. Y. Shen, A. Bax, Identification of helix capping and b-turn motifs from NMR chemical shifts. *J. Biomol. NMR* **52**, 211–32 (2012).

34. P. Güntert, Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).

35. Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, A. Bax, Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4685–4690 (2008).

36. A. Khondker, A. K. Dhaliwal, S. Saem, A. Mahmood, C. Fradin, J. Moran-Mirabal, M. C. Rheinstädter, Membrane charge and lipid packing determine polymyxin-induced membrane damage. *Commun. Biol.* **2**, 67 (2019).

37. J. Bürck, P. Wadhwani, S. Fanghänel, A. S. Ulrich, Oriented circular dichroism: A method to characterize membrane-active peptides in oriented lipid bilayers. *Acc. Chem. Res.* **49**, 184–192 (2016).

38. P. Debye, Molecular-weight determination by light scattering. *J. Phys. Colloid Chem.* **51**, 18–32 (1947).

39. S. Bibow, Y. Polyhach, C. Eichmann, C. N. Chi, J. Kowal, S. Albiez, R. A. McLeod, H. Stahlberg, G. Jeschke, P. Güntert, R. Riek, Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I. *Nat. Struct. Mol. Biol.* **24**, 187–193 (2017).

40. N. Skar-Gislinge, J. B. Simonsen, K. Mortensen, R. Feidenhans'l, S. G. Sligar, B. Lindberg Møller, T. Bjørnholm, L. Arleth, Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the lipids and the protein. *J. Am. Chem. Soc.* **132**, 13713–13722 (2010).

41. N. Skar-Gislinge, N. T. Johansen, R. Høiberg-Nielsen, L. Arleth, Comprehensive study of the self-assembly of phospholipid nanodiscs: What determines their shape and stoichiometry? *Langmuir* **34**, 12569–12582 (2018).

42. N. T. Johansen, M. C. Pedersen, L. Porcar, A. Martel, L. Arleth, Introducing SEC-SANS for studies of complex self-organized biological systems. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 1178–1191 (2018).

43. G. Rouser, A. N. Siakotos, S. Fleischer, Quantitative analysis of phospholipids by thin-layer chromatography and phosphorus analysis of spots. *Lipids* **1**, 85–86 (1966).

44. P. A. Harding, X. Z. Wang, B. Kelder, S. Souza, S. Okada, J. J. Kopchick, In vitro mutagenesis of growth hormone receptor Asn-linked glycosylation sites. *Mol. Cell. Endocrinol.* **106**, 171–180 (1994).

45. F. B. Bjørkskov, S. L. Krabbe, C. N. Nurup, J. W. Missel, M. Spulber, J. Bomholt, K. Molbaek, C. Helix-Nielsen, K. Gotfryd, P. Gourdon, P. A. Pedersen, Purification and functional comparison of nine human Aquaporins produced in Saccharomyces cerevisiae for the purpose of biophysical characterization. *Sci. Rep.* **7**, 16899 (2017).

46. H. Wennbo, M. Gebre-Medhin, A. Gritli-Linde, C. Ohlsson, O. G. P. Isaksson, J. Törnell, Activation of the prolactin receptor but not the growth hormone receptor is important for induction of mammary tumors in transgenic mice. *J. Clin. Invest.* **100**, 2744–2751 (1997).

47. J.-L. K. Kouadio, J. R. Horn, G. Pal, A. A. Kossiakoff, Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. *J. Biol. Chem.* **280**, 25524–32 (2005).

48. M. C. Pedersen, L. Arleth, K. Mortensen, WillItFit : A framework for fitting of constrained models to small-angle scattering data. *J. Appl. Crystallogr.* **46**, 1894–1898 (2013).

49. N. Kucerka, S. Tristram-Nagle, J. F. Nagle, Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **208**, 193–202 (2005).

50. Y. Minezaki, K. Homma, K. Nishikawa, Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *J. Mol. Biol.* **368**, 902–913 (2007).

51. M. Kjaergaard, B. B. Kragelund, Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.* **74**, 3205–3224 (2017).

52. A. J. Brooks, F. Dehkhoda, B. B. Kragelund, in *Principles of Endocrinology and Hormone Action*, A. Belfiore, D. LeRoith, Eds. (Springer International Publishing, Cham, 2018), pp. 157–185; https://doi.org/10.1007/978-3-319-44675-2_8.

53. D. Ben-Avraham, D. R. Govindaraju, T. Budagov, D. Fradin, P. Durda, B. Liu, S. Ott, D. Gutman, L. Sharvit, R. Kaplan, P. Bougnères, A. Reiner, A. R. Shuldiner, P. Cohen, N. Barzilai, G. Atzmon, The GH receptor exon 3 deletion is a marker of male-specific exceptional longevity associated with increased GH sensitivity and taller stature. *Sci. Adv.* **3**, e1602025 (2017).

54. M. Gouw, S. Michael, H. Sámano-Sánchez, M. Kumar, A. Zeke, B. Lang, B. Bely, L. B. Chemes, N. E. Davey, Z. Deng, F. Diella, C. M. Gürth, A. K. Huber, S. Kleinsorg, L. S. Schlegel, N. Palopoli, K. V. Roey, B. Altenberg, A. Reményi, H. Dinkel, T. J. Gibson, The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* **46**, D428–D434 (2018).

55. M. Kaplan, S. Narasimhan, C. de Heus, D. Mance, S. van Doorn, K. Houben, D. Popov-Čeleketić, R. Damman, E. A. Katrukha, P. Jain, W. J. C. Geerts, A. J. R. Heck, G. E. Folkers, L. C. Kapitein, S. Lemeer, P. M. P. van Bergen En Henegouwen, M. Baldus, EGFR dynamics change during activation in native membranes as revealed by NMR. *Cell* **167**, 1241–1251.e11 (2016).

56. X. Wang, C. J. Darus, B. C. Xu, J. J. Kopchick, Identification of growth hormone receptor (GHR) tyrosine residues required for GHR phosphorylation and JAK2 and STAT5 activation. *Mol. Endocrinol.* **10**, 1249–1260 (1996).

57. B. P. Cormack, G. Bertram, M. Egerton, N. A. R. Gow, S. Falkow, A. J. P. Brown, Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in Candida albicans. *Microbiology* **143**, 303–311 (1997).

58. A. Khondker, D. J. Malenfant, A. K. Dhaliwal, M. C. Rheinstädter, Carbapenems and lipid bilayers: Localization, partitioning, and energetics. *ACS Infect. Dis.* **4**, 926–935 (2018).

59. S. Himbert, M. J. Blacker, A. Kihm, Q. Pauli, A. Khondker, K. Yang, S. Sinjari, M. Johnson, J. Juhasz, C. Wagner, H. D. H. Stöver, M. C. Rheinstädter, Hybrid erythrocyte liposomes: Functionalized red blood cell membranes for molecule encapsulation. *Adv. Biosyst.* **4**, 1900185 (2020).

60. R.-C. Bider, T. Lluka, S. Himbert, A. Khondker, S. M. Qadri, W. P. Sheffield, M. C. Rheinstädter, Stabilization of lipid membranes through partitioning of the blood bag plasticizer di-2-ethylhexyl phthalate (DEHP). *Langmuir* **36**, 11899–11907 (2020).

61. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).

62. K. Reichel, O. Fisette, T. Braun, O. F. Lange, G. Hummer, L. V. Schäfer, Systematic evaluation of CS-Rosetta for membrane protein structure prediction with sparse NOE restraints. *Proteins* **85**, 812–826 (2017).

63. I. G. Denisov, Y. V. Grinkova, A. A. Lazarides, S. G. Sligar, Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).

64. M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).

65. M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, D. I. Svergun, New developments in the *ATSAS* program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).

66. A. Jordan, M. Jacques, C. Merrick, J. Devos, V. T. Forsyth, L. Porcar, A. Martel, SEC-SANS: Size exclusion chromatography combined in situ with small-angle neutron scattering. *J. Appl. Crystallogr.* **49**, 2015–2020 (2016).

67. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

68. J. Koehler Leman, R. Bonneau, A novel domain assembly routine for creating full-length models of membrane proteins from known domain structures. *Biochemistry* **57**, 1939–1944 (2018).

69. P. C. T. Souza, S. J. S. Marrink, Martini 3 - Open Beta-Release (2020); http://cgmartini.nl.

70. Y. Qi, H. I. Ingólfsson, X. Cheng, J. Lee, S. J. Marrink, W. Im, CHARMM-GUI martini maker for coarse-grained simulations with the martini force field. *J. Chem. Theory Comput.* **11**, 4486–4494 (2015).

71. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

72. D. H. de Jong, S. Baoukina, H. I. Ingólfsson, S. J. Marrink, Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **199**, 1–7 (2016).

73. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

74. A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).

75. PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).

76. S. Grudinin, M. Garkavenko, A. Kazennov, Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 449–464 (2017).

77. B. Hammouda, Structure factor for starburst dendrimers. *J. Polym. Sci. Part B Polym. Phys.* **30**, 1387–1390 (1992).

78. J. S. Pedersen, M. C. Gerstenberg, Scattering form factor of block copolymer micelles. *Macromolecules* **29**, 1363–1365 (1996).

79. Y. Wei, A. A. Thyparambil, R. A. Latour, Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim. Biophys. Acta Proteins Proteomics* **1844**, 2331–2337 (2014).

80. R. F. Sommese, S. Sivaramakrishnan, R. L. Baldwin, J. A. Spudich, Helicity of short E-R/K peptides. *Protein Sci.* **19**, 2001–2005 (2010).

81. S. Hansen, BayesApp: A web site for indirect transformation of small-angle scattering data. *J. Appl. Crystallogr.* **45**, 566–567 (2012).

82. E. L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).

83. L. Käll, A. Krogh, E. L. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).

84. L. Käll, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. *Nucleic Acids Res.* **35**, W429-W432 (2007).

85. D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538–544 (2007).

86. UniProt Consortium, UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).

87. W. Zheng, R. B. Best, An extended guinier analysis for intrinsically disordered proteins. *J. Mol. Biol.* **430**, 2540–2553 (2018).

88. J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, T. R. Sosnick, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).

89. J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, K. W. Plaxco, Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491–12496 (2004).

90. J. A. Marsh, J. D. Forman-kay, Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).

91. N. T. Johansen, F. G. Tidemand, T. T. T. N. Nguyen, K. D. Rand, M. C. Pedersen, L. Arleth, Circularized and solubility-enhanced MSPs facilitate simple and high-yield production of stable nanodiscs for studies of membrane proteins in solution. *FEBS J.* **286**, 1734–1751 (2019).

# 9

## MODELING OF FLEXIBLE MEMBRANE-BOUND BIOMOLECULAR COMPLEXES FOR SMALL-ANGLE SCATTERING

---

### CONTRIBUTIONS

This paper describes the model developed for Paper II in great detail. I built the model, calculated the model calculations and wrote the article, with supervision from Lise Arleth and Martin Cramer Pedersen.

# Modeling of flexible membrane-bound biomolecular complexes for solution small-angle scattering

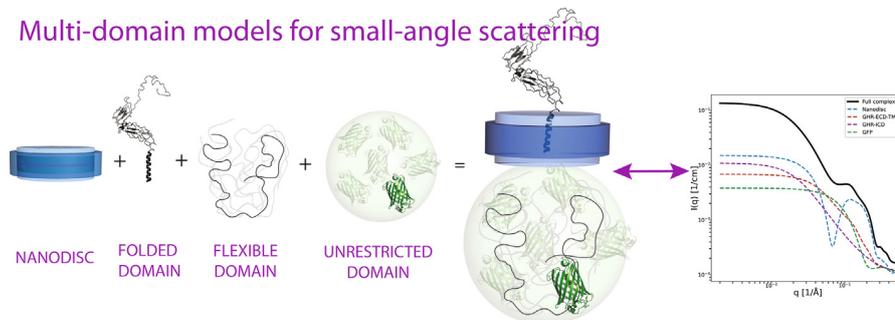Abigail Barclay [a],*, Birthe B. Kragelund [b], Lise Arleth [a], Martin Cramer Pedersen [a]

[a] *Condensed Matter Physics, Niels Bohr Institute, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark*
[b] *Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark*

## GRAPHICAL ABSTRACT



## ARTICLE INFO

## ABSTRACT

Recent advances in protein expression protocols, sample handling, and experimental set up of small-angle scattering experiments have allowed users of the technique to structurally investigate biomolecules of growing complexity and structural disorder. Notable examples include intrinsically disordered proteins, multi-domain proteins and membrane proteins in suitable carrier systems. Here, we outline a modeling scheme for calculating the scattering profiles from such complex samples. This kind of modeling is necessary for structural information to be refined from the corresponding data. The scheme bases itself on a hybrid of classical form factor based modeling and the well-known spherical harmonics-based formulation of small-angle scattering amplitudes. Our framework can account for flexible domains alongside other structurally elaborate components of the molecular system in question. We demonstrate the utility of this modeling scheme through a recent example of a structural model of the growth hormone receptor membrane protein in a phospholipid bilayer nanodisc which is refined against experimental SAXS data. Additionally we investigate how the scattering profiles from the complex would appear under different scattering contrasts. For each contrast situation we discuss what structural information is contained and the related consequences for modeling of the data.

© 2022 Published by Elsevier Inc.

## 1. Introduction

Proteins possess a myriad of different structural characteristics, ranging from well-defined folded structures, to highly disordered conformational ensembles, to networks of both rigid and flexible regions. While many proteins are water-soluble and can exist free

∗ Corresponding author.
   *E-mail addresses:* abigail.barclay@nbi.ku.dk (A. Barclay), bbk@bio.ku.dk (B.B. Kragelund), arleth@nbi.ku.dk (L. Arleth), mcpe@nbi.ku.dk (M.C. Pedersen).

in aqueous environments, others are rich in hydrophobic residues which must be shielded from water in order to be in their native state. The latter relates mainly to membrane proteins. We present some examples in Fig. 1.

The understanding of intrinsically disordered proteins (IDPs), or proteins with large intrinsically disordered regions (IDRs), falls far behind the understanding of well-structured folded proteins [2]. Nonetheless, more than 40% of transmembrane proteins in eukaryotic systems contain IDRs of a significant length [3,4]. Highly flexible proteins like these pose a challenge to classical structural biology techniques as they cannot be crystallised for investigations using traditional X-ray protein crystallography and are imaged poorly with cryo-electron microscopy (cryo-EM). Nuclear magnetic resonance (NMR) spectroscopy is a successful high-resolution technique for studying IDPs and can give information about the small structured subdomains appearing in IDRs [5,6]. NMR is typically not suitable for folded proteins above 35 kDa, although IDPs at this size and larger are more amenable.

Membrane proteins are another family of proteins which are notoriously challenging to study as they require an amphiphillic environment to be soluble and active. To remain soluble, the hydrophobic regions of membrane proteins must be stabilised eg. by detergents or a lipid micelle or other carrier system [7]. Nanodiscs are discoidal particles consisting of a central phospholipid bilayer enclosed by two membrane scaffold proteins (MSPs) [8,9]. Membrane proteins can be loaded into nanodiscs such that the protein is given a native-like environment and structural and functional experiments can be performed [10].

Solution small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) have emerged as complementary methods that can effectively probe the low-resolution structure of disordered protein systems [11,12]. It is straightforward to extract the radius of gyration ($R_g$) from small-angle scattering (SAS) data

[13–15], meaning SAS-methods can effectively contribute to the description of the compactness of disordered chains, a property that relates to its sequence features [72, 73] and relevant to biological function [74].

Further insight into the conformational landscape of disordered proteins can be gained through ensemble methods where large ensembles of theoretical conformations are refined from the SAS data, eg. *EOM* [16,17], *MES* [18], *BE-SAXS* [19]. These methods, however, are reserved to soluble proteins since lipid environments and other carrier particles cannot usually be accounted for. Furthermore, these techniques are limited by the low information content contained in SAXS and SANS data. Molecular dynamics (MD) simulations of IDPs, corroborated by experimental SAS data, are much richer in information in comparison [20,21].

It seems "integrative structural biology" is the way forward for complexes with disordered domains, where MD simulations and high-resolution experimental techniques can be combined and refined against experimental SAS data to obtain a complete description of the structure and dynamics of the system [22,23].

SAXS and SANS have also been applied to investigate the size and shape of membrane protein systems. Classical SAS models represent the particle as simple geometrical objects which allows analytical form factors to be reined from the data [24]. Additional chemical and biophysical information can be included to constrain the solution space of model [25]. This method of modeling has been successful for a variety of carrier systems including phospholipid vesicles [26,27], surfactant micelles [28] and nanodiscs [29,30].

The intricate shape of folded proteins is better suited to bead-based modeling, where each atom, or collection or atoms, is represented by a bead [31]. This means if a structure of the protein is already available, the corresponding scattering profile can be calculated directly. Known protein structures can be included as
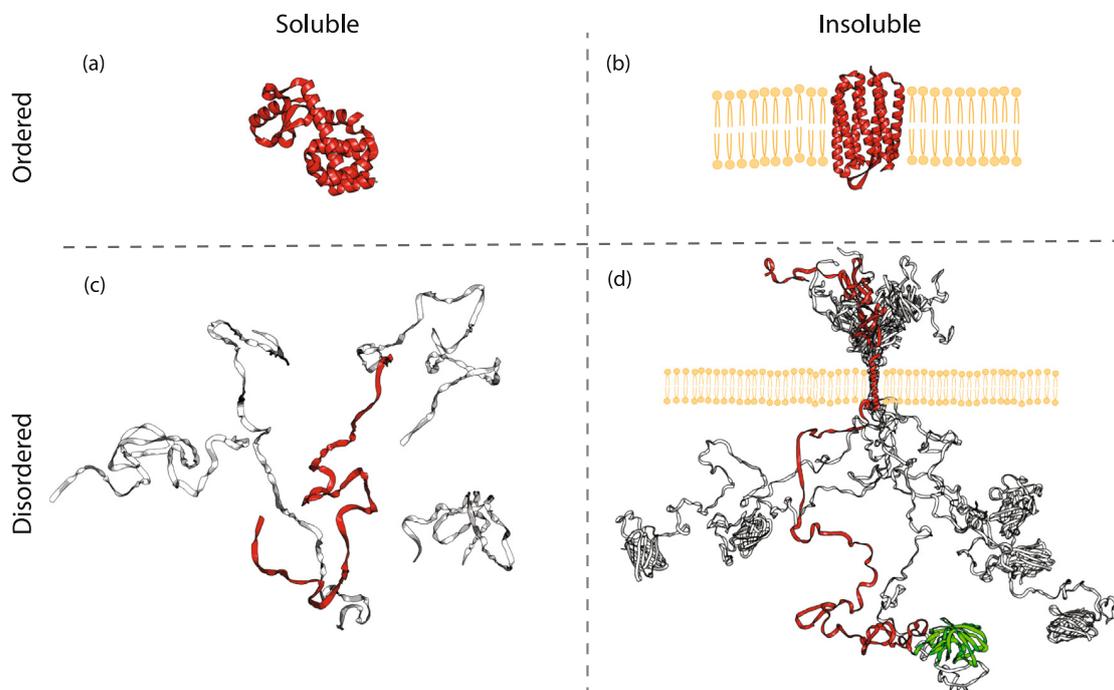


**Fig. 1.** Classes of protein structure. (a) T4 lysozyme (PDB6LZM) which folds into a compact structure in water. (b) Bacteriorhodopsin (PDB1QKP) which folds into a compact structure in a lipid bilayer: a highly-hydrophobic membrane protein composed of seven $\alpha$-helices spanning the length of the membrane. (c) A small ensemble of conformations of $\beta$-synuclein (PED00003) which is natively unfolded throughout its entire length in water. (d) A small ensemble of conformations of the growth hormone receptor, in a lipid bilayer, with a green fluorescent protein attached (Kassem et al. [1]). The growth hormone receptor is a membrane protein which consists of both folded and disordered sections: a disordered N-terminal tail, a folded extracellular domain attached to the transmembrane $\alpha$-helix by a flexible linker, and finally an intracellular domain which is entirely intrinsically disordered. Green fluorescent protein is a well-folded and soluble small protein, which is linked to the construct to facilitate expression.

rigid bodies within complexes of other protein/ lipid subunits [32].

Here, we build upon our previous study on calculating the SAS profile from folded membrane proteins in nanodiscs by combining bead-based modeling for the protein with analytical modeling for the nanodisc [33]. We expand the model to additionally describe disordered protein domains with the form factor for Gaussian random coils. We follow the principle from Pedersen and Gerstenberg [34] to attach the Gaussian random coil model to the rigid body to allow the intrinsically disordered region of a membrane protein to be accounted for alongside its structured regions.

The human growth hormone receptor (GHR) is a 70 kDa transmembrane protein with over 50% disorder and is one of around 40 receptors belonging to the class 1 cytokine receptor family. The family is characterised by a folded extracellular domain (ECD), a single-pass helical transmembrane domain (TMD) and a long intracellular domain (ICD) which is entirely intrinsically disordered [35,36]. The GHR is illustrated in Fig. 1(d).

Furthermore, as a way to ease protein purification and structural experiments, membrane proteins are often fused with marker proteins during expression, such as green fluorescent protein (GFP) [37–39]. These marker proteins are often difficult or impossible to remove before SAS data are collected. It is therefore relevant to develop structural analysis that can handle such an object attached to the end of a long flexible region, and also to investigate how much it obscures the SAS signal from the rest of the complex.

In this work we present a model for the full-length GHR carrying a C-terminal GFP and embedded in a nanodisc. The multiplex nature of this protein system makes it an intriguing example case for those interested in the modeling of small-angle scattering data from biomolecules; the main challenge lies in describing the GHR's large intrinsically disordered region as well as the attached GFP which has no fixed position with respect to the rest of the system but rather diffuses within some volume. The model was recently presented in a combined experimental and MD study by Kassem et al. [1]. This application of our model to experimental SAXS data set is shown in Section 4, but we stress that the modeling scheme may easily be adapted to other membrane proteins of similar geometry.

Additionally, in Section 3.2 we investigate how the protein versus the phospholipid nanodisc contribute to the theoretical scattering profiles under three different contrast conditions: SAXS, SANS and SANS with negligible scattering from the nanodisc. In Section 3.3 we investigate which of the contrasts are effective for studying the degree of expansion of the IDR. We stress the importance of making theoretical predictions ahead preparing samples for SAXS and SANS beamtimes to ensure the data collected will contain information that is relevant to the biological problem in question.

## 2. Theory and Methods

### 2.1. Theory: Small-angle scattering

In a SAS experiment, a sample is irradiated by an incoming beam of monochromatic radiation, and the scattered intensity is recorded on a position-sensitive detector placed behind the sample. From the geometry of the set up, one calculates the scattering angle, $2\theta$, which is combined with the wavelength of the radiation, $\lambda$, to establish the scattering momentum transfer, $q = 4\pi \sin(\theta)/\lambda$. $q$ is usually calculated in units of $\text{Å}^{-1}$ or $\text{nm}^{-1}$. The recorded intensity, $I$, has units of $\text{cm}^{-1}$ and is traditionally presented as $I$ versus $q$.

The theoretical SAS intensity from a dilute dispersion of identical particles in solution can be expressed by:

$$I(q) = n\,\Delta b^2\,P(q) \tag{1}$$

where $n$ is the particle number density, $\Delta b$ is the excess scattering length of the particles which is usually determined as $\Delta b = \Delta\rho\,V$ if $\Delta\rho$ is the excess scattering length density of the particle and $V$ is the particle volume. $P(q)$ is the particle form factor, normalised so $P(q \to 0) = 1$ and averaged over every orientation of the particles. Generally $\left\langle \left|A(\vec{q})\right|^2 \right\rangle_\Omega = \left\langle \left|\Delta b\,F(\vec{q})\right|^2 \right\rangle_\Omega = \Delta b^2\,P(q)$ where $A(\vec{q})$ is known as the scattering amplitude, $F(\vec{q})$ the normalised form factor amplitude, $|\ldots|$ denotes the complex norm and $\langle \ldots \rangle_\Omega$ denotes the orientational average.

As an example, the scattering from a solution of cylindrical particles can be computed using the associated particle form factor [40]:

$$P(q) = \int_0^{\frac{\pi}{2}} \left( \frac{2J_1(qR\sin(\alpha))}{q\sin(\alpha)} \frac{\sin\left(qH\cos\left(\frac{\alpha}{2}\right)\right)}{qH\cos\left(\frac{\alpha}{2}\right)} \right)^2 \sin(\alpha)\,\mathrm{d}\alpha \tag{2}$$

where $R$ and $H$ represent the radius and the height of the cylindrical particles, respectively, $J_1$ is the first Bessel function of the first kind and $\alpha$ is the azimuthal angle.

### 2.2. Theory: The multipole expansion on Bessel functions and spherical harmonics

We build upon previous works from our group detailing a hybrid approach for the structural analysis of small-angle scattering data from protein-nanodisc systems: combining bead-based modeling with analytical modeling [33,41]. Both types of modeling are executed in spherical harmonics to be more computationally-efficient. Fig. 2(b) shows an atomic-level bead model inserted into a continuous nanodisc as a rigid-body. This hybrid approach allows us to incorporate our prior knowledge of the system e.g. folded atomic protein structures, while investigating the lesser known parameters of the system e.g. nanodisc geometry, extendedness of the ICD.

For a structural model composed of a discrete number of $N$ beads representing either atoms or protein residues, the scattering amplitude can be calculated in spherical harmonics as [31]:

$$A(\vec{q}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \mathscr{A}_{lm}(q) Y_{lm}(\alpha, \beta) \tag{3}$$

where $\vec{q} = (q, \alpha, \beta)$ in spherical coordinates, $L$ is the maximum number of spherical harmonics, $Y_{lm}$ is the $lm$'th spherical harmonic and the partial amplitudes are:

$$\mathscr{A}_{lm}(q) = 4\pi i^l \sum_{j=1}^{N} \Delta b_j J_l(qr_j) Y_{lm}^*(\theta_j, \phi_j) \tag{4}$$

where the $j$'th bead is positioned at $\vec{r}_j = (r_j, \theta_j, \phi_j)$ in spherical coordinates and has excess scattering length $\Delta b_j$, $J_l(qr_j)$ is the $l$'th spherical Bessel function and $*$ denotes the complex conjugate. In theory Eqn. 3 holds when $L$ approaches $\infty$, however, we find $L = 17$ to be a satisfactory computational approximation for the applications demonstrated in this manuscript.

A continuous model built from form factor amplitudes can also be expanded into spherical harmonics [33], $\mathscr{B}_{lm}(q)$ like with Eqn. 3, where the spherical harmonics coefficients are given by

$$\mathscr{B}_{lm}(q) = \Delta b \int_0^{\pi} \int_0^{2\pi} F(q, \alpha, \beta) \frac{\exp(-im\beta)}{(2\pi)^{\frac{1}{2}}} \tilde{P}_{lm}(\cos(\alpha)) \sin(\alpha)\,\mathrm{d}\beta\,\mathrm{d}\alpha \tag{5}$$
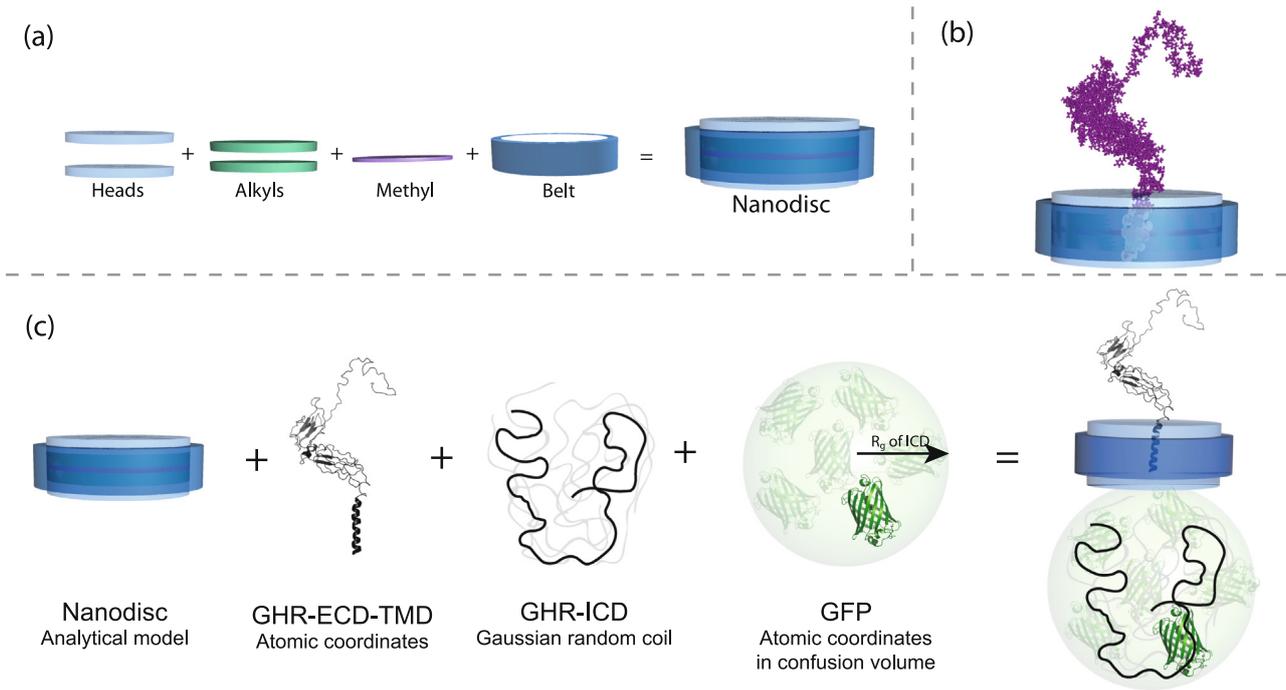
A. Barclay, B.B. Kragelund, L. Arleth et al.

**Fig. 2.** (a) Nanodiscs can be modeled analytically when represented as a collection of discs. Each disc represents a subunit of the nanodisc with varying chemical properties and is modeled using the form factor for elliptical cylindrical particles. (b) Combining bead-based modeling with analytical modeling. Each protein atom is represented by a purple bead. Some volume of the nanodisc is excluded so that there is no overlap with the protein. (c) Construction of the full model with a spherical distribution volume. The scattering amplitude arising from each of the four components are first calculated independently, and then are combined with their phases taken into account to give the total scattering intensity of the GHR-GFP fusion in a nanodisc.

where $\tilde{P}_{lm}(\cos(\alpha))$ are Legendre polynomials in $\cos(\alpha)$. Mohlenkamp [42,43] shows the integration can be implemented on a $2L \times 2L$ grid with sites $(\alpha_j, \beta_k)$, where $\alpha$ ranges from 0 to $\pi$ and $\beta$ ranges from 0 to $2\pi$, as:

$$\mathscr{B}_{lm}(q) = \sum_{j=0}^{2L-1} B_m(\alpha_j) \tilde{P}_{lm}(\cos(\alpha_j)) \sin(\alpha_j) w(\alpha_j) \qquad (6)$$

where the form factor of interest is discretised by:

$$B_m(\alpha_j) = \frac{(2\pi)^{\frac{1}{2}}}{2L} \sum_{k=0}^{2L-1} F(q, \alpha_j, \beta_k) \exp(-im\beta_k) \qquad (7)$$

and Chebychev weights are calculated for each point in $\alpha$:

$$w(\alpha_j) = \frac{2^{\frac{1}{2}}}{L} \sum_{l=0}^{L-1} \frac{1}{2l+1} \sin((2l+1)\alpha_j) \qquad (8)$$

Hence the theoretical scattering intensity from a hybrid model composed of two components can be calculated by taking the square of the absolute sum of the complex amplitudes in spherical harmonics:

$$
\begin{aligned}
I(q) &= n \cdot \sum_{l=0}^{L} \sum_{m=-l}^{l} |\mathscr{A}_{lm}(q) + \mathscr{B}_{lm}(q)|^2 \\
&= n \cdot \sum_{l=0}^{L} \sum_{m=-l}^{l} \underbrace{|\mathscr{A}_{lm}(q)|^2 + |\mathscr{B}_{lm}(q)|^2}_{\text{Auto-correlation terms}} + \\
&\qquad \underbrace{\mathscr{A}_{lm}^*(q)\mathscr{B}_{lm}(q) + \mathscr{A}_{lm}(q)\mathscr{B}_{lm}^*(q)}_{\text{Cross-correlation terms}}
\end{aligned} \qquad (9)
$$

In Eqn. (9), we include the second (somewhat trivial) equality to emphasize the structure of the emerging expression. In this work, it is important to identify the self- and cross-correlation terms, as we shall be correcting them in various manners.

### 2.3. Theory: Semi-analytical model for the small-angle scattering of GHR-GFP embedded in a nanodisc

The GFP-tagged GHR (GHR-GFP) in a nanodisc contains varying degrees of order and disorder throughout the system which require different modeling approaches. Our design consists of separating the system into four distinct components: the extracellular and transmembrane domain of the GHR (GHR-ECD-TMD), the intracellular domain of the GHR (GHR-ICD), the GFP fused to the end of the GHR-ICD (GFP), and the surrounding nanodisc (ND). Our model hence has the architecture:

$$I(q) = n \cdot \left\langle \left| A^{(ND)}\left(\vec{q}\right) + A^{(ECD-TMD)}\left(\vec{q}\right) + A^{(ICD)}\left(\vec{q}\right) + A^{(GFP)}\left(\vec{q}\right) \right|^2 \right\rangle_\Omega \qquad (10)$$

In order to calculate the full theoretical scattering intensity, first the scattering amplitude arising from each component is calculated independently and then are combined, accounting for all auto- and cross-correlation terms and taking the internal phases into account. Scattering amplitudes are calculated using the most appropriate type of modeling for that component. The approach is illustrated in Fig. 2c) and the calculation for each component is explained in detail in the following section.

### 2.3.1. Nanodisc

A well-documented approach is used to account for the scattering amplitude arising from the nanodisc, $A^{(ND)}(\vec{q})$ [29,25]. The nanodisc model describes the phospholipid bilayer as a series of elliptical discs. The discs represent the phospholipids' central methyl-group region, intermediate alkyl chain regions and outer head group regions respectively. Furthermore the bilayer is surrounded by a hollow elliptical disc representing the double stacked membrane scaffold protein (MSP)-belt. The construction of the nanodisc model is illustrated in Fig. 2(a). The scattering arising

from each disc is calculated analytically with the form factor amplitude for elliptical cylinders [40]. $A^{(ND)}(\vec{q})$ can be expanded into spherical harmonics coeficients, $\mathscr{A}_{lm}^{(ND)}(q)$, using Eqn. 6. The center of the nanodisc is the origin of our system, Fig. 2(c).

### 2.3.2. ECD-TMD

For the GHR-ECD-TMD a single set of atomic coordinates can be inserted into the nanodisc as a rigid-body using the same method as outlined in Kynde et al. [33]. Some atoms displace lipids which occupy a corresponding volume, as shown in Fig. 2(b). Each atom's excess scattering length, $\Delta b$, is calculated according to its surroundings, whether that be solvent or some region of the bilayer. The spherical harmonics coefficients, $\mathscr{A}_{lm}^{(ECD-TMD)}(q)$, can then be calculated using Eq. (4).

### 2.3.3. ICD

To include the GHR-ICD attached to the TMD we use the same principle as derived for block copolymer micelles by Pedersen and Gerstenberg [34,44]. Assuming a Gaussian random coil structure, the auto-correlation term for the ICD is given by $I^{(ICD)}(q) = \Delta b_{ICD}^2 \, F_D(q)$ where $F_D(q)$ is the Debye function: [45]

$$F_D(q) = \frac{2(\exp(-x) - 1 + x)}{x^2} \tag{11}$$

where $x = \left(qR_g^{(ICD)}\right)^2$ and $R_g^{(ICD)}$ is the radius of gyration of the chain.

The cross-correlation terms between the GHR-ICD and other components must be calculated using the form factor amplitude corresponding to Gaussian random coils, known as the Hammouda function [46]:

$$\psi(q) = \frac{1 - \exp(-x)}{x} \tag{12}$$

Furthermore, the phase factor is calculated from the physical position of the GHR-ICD in the system. This is mimicked by displacing the centre of mass of the coil straight down from the origin (the centre of the disk) by $\vec{r} = \frac{1}{2}h^{(ND)} + R_g^{(ICD)}$ where $h^{(ND)}$ is the height of the nanodisc bilayer. The phased scattering amplitude of the ICD is hence described with $A^{(ICD)}(\vec{q}) = \Delta b_{ICD} \, \psi(q) \, \exp(-i\vec{q} \cdot \vec{r})$, which can be expanded into spherical harmonics using Eqn. 6, $\mathscr{A}_{lm}^{(ICD)}(q)$. Note that in our final expression, Eqn. 14, $I^{(ICD)}(q)$ does not need to be expanded into spherical harmonics since it is summed outside of the brackets and is inherently orientationally averaged.

### 2.3.4. GFP

The GFP is a folded protein with a well-established structure. The spherical harmonics coefficients of the GFP, $\mathscr{A}_{lm}^{(GFP)}(q)$, can again be calculated directly from its atomic coordinates using Eqn. 4. Since the GFP in our model system is attached to the end of another flexible domain, it has neither a fixed position nor orientation with respect to the rest of the system. The GFP is free to diffuse within some volume which is constrained by the length of the flexible domain. To account for this a special "phase factor" was developed for the analytical expression for scattering arising from the GFP.

Firstly, to decorrelate the orientation of the GFP from the rest of the system, we use the scattering amplitude in the form $\mathscr{A}_{00}^{(GFP)}(q)$: the zeroth-order spherical harmonic coefficient [47]. This is proportional to the orientationally averaged scattering amplitude.

Secondly, following a similar philosophy as that detailed by Pedersen and Gerstenberg [34], we introduce the concept of a "distribution volume": a volume within which a certain scattering density is evenly distributed. The distribution volume can be any 3D

shape for which a form factor amplitude is available. To position the GFP randomly within the distribution volume, in the cross-correlation calculations the scattering amplitude from the GFP is given by the product:

$$\mathscr{A}_{lm}^{(GFP \cdot distribution)}(q) = \mathscr{A}_{00}^{(GFP)} \, \Phi_{lm}(q) \, \exp(-i\vec{q} \cdot \vec{r}) \tag{13}$$

where $\Phi_{lm}(q)$ is the form factor amplitude corresponding to the chosen geometry of the distribution volume expanded into spherical harmonics coefficients and $\vec{r}$ is the displacement of the distribution volume from the origin.

As a simple example we choose to mainly work with a spherical distribution volume placed just below the nanodisc with the radius of the sphere, $R$, equal to the $R_g^{(ICD)}$, and $\vec{r} = \frac{1}{2}h^{(ND)} + R_g^{(ICD)}$. In addition we include results where the sphere is replaced with a thick cylindrical shell, which better mimics the bowl-like distribution of the GFP sampled by MD simulations from Kassem et al. [1]. We arrive at the form factor amplitude for the cylindrical shell by subtracting a smaller disc from a larger one, where the discs are positioned so that the resulting cylinder is closed on the bottom end and open on the top. This version of the model is illustrated in Fig. SI.1 alongside the scattering profiles from both models in Fig. SI.2. The scattering profiles show that this choice of distribution volume has a negligible intensity profile on the output on this occasion.

### 2.3.5. Full expression

We compute the scattering intensity of the full system in spherical harmonics with:

$$I(q) = n \cdot \sum_{l=0}^{L} \sum_{m=-l}^{l} \overbrace{\left| \mathscr{A}_{lm}^{(ND)}(q) + \mathscr{A}_{lm}^{(ECD-TMD)}(q) + \mathscr{A}_{lm}^{(ICD)}(q) + \mathscr{A}_{lm}^{(GFP \cdot distribution)}(q) \right|^2}^{\text{Naïve construction of total intensity}} \\ \underbrace{- \left| \mathscr{A}_{lm}^{(GFP \cdot distribution)}(q) \right|^2 + \left| \mathscr{A}_{lm}^{(GFP)}(q) \right|^2}_{\text{Correction term for GFP auto-correlation}} \underbrace{- \left| \mathscr{A}_{lm}^{(ICD)}(q) \right|^2 + I^{(ICD)}(q)}_{\text{Correction term for ICD auto-correlation}} \tag{14}$$

Eqn. 14 includes some small adjustments to the standard procedure of calculating the total particle intensity as the orientationally averaged, absolute square of the sum of the scattering amplitudes, ie simply as Eqn. 10. The correction term for the ICD ensures the GHR-ICD is described with the correct Gaussian random coil expression in the auto-correlation and cross-correlation terms respectively. The correction term for the GFP ensures the form factor amplitude of the distribution volume is used in the cross-correlation terms to establish a distribution of positions of the GFP without contributing to the shape of the GFP auto-correlation scattering profile.

## 3. Results and discussion

### 3.1. Model calculations

The model was implemented in the framework *WillItFit* [48]. The model incorporates molecular information in order to reduce the number of degrees of freedom. As part of this, the model can be calculated on absolute scale by exploiting the known sample concentration. The scattering lengths and volumes of the phospholipids, MSP and GHR-ICD are pre-set.

The nanodisc fit parameters become: i) the number of lipids per disc, $N_L$, ii) the average area per lipid headgroup, $A_L$, iii) the axis ratio of the disc, $\epsilon$ and iv) the height of the cylinder describing the belt protein, $h_{MSP}$. The flexible domain of the GHR is parameterised by v) the radius of gyration of the GHR-ICD, $R_g^{(ICD)}$. Additionally, a roughness term, ix), is included to correct for the fact that the interfaces are not perfectly smooth [49] as well as a constant background, x) $b$. The partial specific molecular volumes of a lipid,

$v_L$, the MSP, $v_{MSP}$, and the GHR-ICD, $v_{ICD}$, may be allowed to vary within a few percent of their pre-set values.

The model was calculated using a set of reasonable input parameters, Table 1, akin to those we previously refined from SAXS data [1]. The nanodisc constitutes of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) loaded MSP1D1. For the GHR-ECD-TMD, a single set of atomic coordinates were back-mapped from the MD simulation of the full-length GHR in a lipid bilayer [1]. PDB1EMA [50] was used for the GFP atomic coordinates. $v_L$ and $A_L$ were taken from literature [51], as well as $h_{MSP}$ [52]. The volumes of the $v_{MSP}$ and $v_{ICD}$ were estimated from their theoretical mass using the average mass density of proteins 1.35 g cm$^{-3}$ [53].

SAXS and SANS scattering lengths of POPC, MSP1D1 and GHR-ICD according to their molecular compositions are listed in Table SI.1. Neutron scattering lengths can be found in Sears [54]. In the SANS calculations in 100% D$_2$O it was assumed that 100% of the exchangeable H in the disordered GHR-ICD would exchange with the solvent, while 50% of the exchangeable H in MSP1D1 would exchange in line with experimental determination [55]. The software *PSX* [56] was used to simulate H/D exchange in the GHR-ECD-TMD and GFP after one hour with a pH of 7. Furthermore, atomic volumes required for calculations involving the two atomic protein structures were taken from Fraser et al. [57].

### 3.1.1. Model validation and visualisation

In Fig. 3 we plot the theoretical scattering profiles for SAXS and SANS of the GHR-GFP embedded in an ND as calculated from Eqn. (14). We also plot the scattering intensity arising from each component independently in order to visualise the contribution of each component. We stress that the cross-correlation terms are fundamental for calculating the total scattering intensity of the system but are not included Fig. 3 for clarity.

In the SAXS contrast situation, the nanodisc profile displays the characteristic flat Guinier region in the low-$q$ regime, followed by a trough and a broad bump at mid- to high-$q$ [29,58]. These features arise specifically from the negative excess X-ray scattering lengths of lipid alkyl-chains and are not present in the SANS contrast situation. We note that in the SANS the scattering is strongly dominated by the profile of the nanodisc, particularly in the mid-$q$ regime. In SAXS and SANS the forward scattering intensities arising from the three protein components scale appropriately with the length of the protein, with the GHR-ICD containing the most residues (350) followed by the GHR-ECD-TMD (270) and the GFP (240). The profile representing the GHR-ICD is simply the Debye function where the size of the Guinier region is controlled by the $R_g$. The profiles representing the GHR-ECD-TMD and the GFP are both the typical shape for small globular proteins, with the GHR-ECD-TMD signal beginning to decay earlier indicating it occupies a larger volume.

**Table 1**
Parameters used in model the calculations.

| Parameter | Value |
| --- | --- |
| Axis ratio of lipid bilayer patch | 1.4 |
| Area per lipid headgroup [Å$^2$] | 62.7 |
| No. lipids per nanodisc | 120 |
| Height of protein belt [Å] | 25.8 |
| $R_g$ of Gaussian random coil [Å] | 70 |
| Volume of MSP1D1 [Å$^3$] | 27100 |
| Volume of POPC [Å$^3$] | 1260 |
| Volume of GHR-ICD [Å$^3$] | 47300 |
| Roughness [Å] | 3 |
| Background [1/cm] (SAXS) | 0.0001 |
| Background [1/cm] (SANS) | 0.001 |
| Concentration [$\mu$M] | 10 |

### 3.1.2. Impact of the GFP

In Fig. 4 we investigate the repercussions of including a biological marker during SAXS and SANS experiments on the consequential data analysis. We plot the theoretical scattering profiles of the entire system including the GFP alongside the theoretical scattering profiles of the system without the GFP.

As we saw in Fig. 3(b), in SANS experiments the signal from the nanodisc dominates the total scattering profile. Therefore removing the GFP has little impact aside from a relatively small reduction of the forward scattering. In SAXS, however, the signal arising from the GFP has a significant impact on the total scattering profile and contributes to the masking of the minimum in the nanodisc profile at mid-$q$. By removing the GFP, the nanodisc features become more prominent and therefore the profile can be analysed with more certainty. Although we demonstrate here that a distribution volume can be used to model a rigid-body attached to a flexible region, the additional signal may cloak important information about the system which could otherwise be extracted. We advise that, wherever possible, marker proteins such as a GFP are removed from experimental constructs in order to simplify analytical modeling of SAXS data.

### 3.2. Protein/lipid complexes under various scattering contrasts

An advantage of SANS over SAXS is the additional possibility for contrast variation. In SANS experiments of biomolecular complexes the scattering signal from one component can be isolated by suppressing the scattering contribution of the other components by adjusting $\Delta\rho$, either by varying the fraction of D$_2$O versus H$_2$O in the solvent [61] or by chemically switching H for D in certain components of the complex [62].

The breakthrough of stealth nanodiscs, where the nanodisc is deuterated to produce negligible scattering against a 100% D$_2$O background, has the potential to further accelerate structural studies of membrane proteins [63]. The use of stealth nanodiscs in SANS means a coherent signal from the membrane protein can be collected without much cross-contamination from the nanodisc signal [64,65].

Fig. 5 illustrates our GHR-GFP nanodisc complex in three different contrast situations. Fig. 5 is colour coded by scattering length desity (SLD), and also shows the corresponding theoretical scattering profiles. We quantitatively compare the intensities of the scattering profiles arising from the nanodisc and the GHR-GFP independently by plotting their auto-correlation terms.

Fig. 5 (a) shows the X-ray contrast. The GHR-GFP has a slightly larger scattering contribution than the nanodisc. With X-rays the phospholipid tail groups have a negative $\Delta\rho$ which translates into the characteristic features of nanodisc SAXS profiles which are absent in SANS, e.g. the mid-$q$ bump.

Fig. 5 (b) shows the neutron contrast situation in 100% D$_2$O and without deuterium labelling of the nanodisc. The component with the largest contrast compared to the solvent is the lipid tails groups. As can be seen from the plots, the overall scattering signal is therefore hugely dominated by the signal from the nanodisc, and the membrane protein has a very small contribution in comparison, particularly in the mid-$q$ regime. This contrast situation is the least optimal when investigating membrane proteins, whether the proteins are solubilised in nanodiscs or another hydrogenated phospholipid system. It could provide highly useful information, however, if the aspect of interest is effects on the phospholipid structure.

Fig. 5 (c) shows the neutron contrast situation in 100% D$_2$O with the GHR-GFP embedded in a stealth nanodisc. The nanodisc is "invisible" against the solvent and in theory the scattering profile only contains signal from the membrane protein. In practice, although there may be a slight signal from the nanodisc at low-$q$ values,
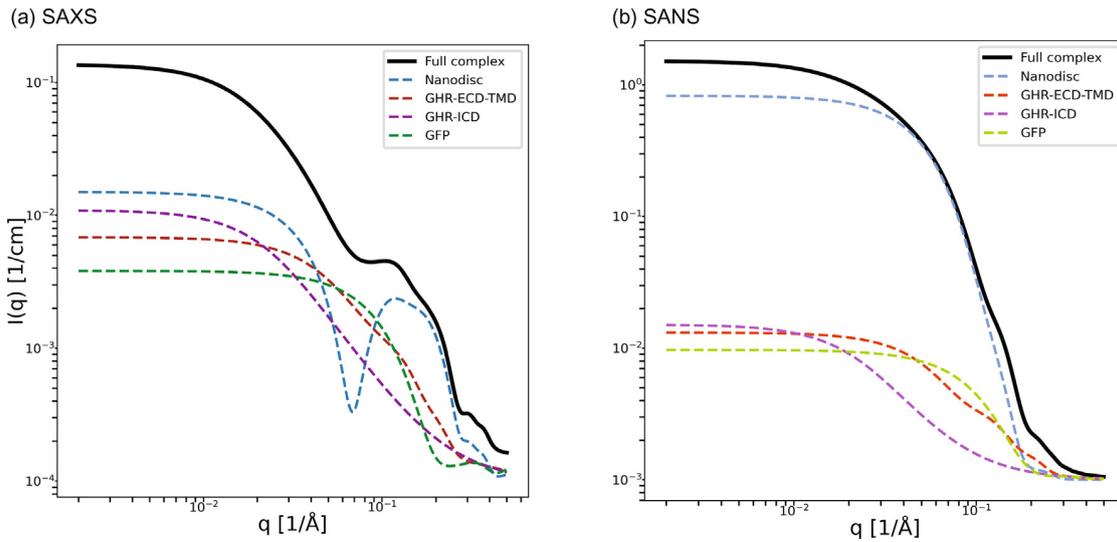
**Fig. 3.** Illustration of the model composition. The black profile shows the total scattering intensity from the complex, Eqn. (14). The colourful profiles represent the auto-correlation terms arising from the four individual components of the model. Blue: nanodisc $\left|A^{(ND)}(q)\right|^2$. Red: GHR-ECD-TMD $\left|A^{(ECD-TMD)}(q)\right|^2$. Purple: GHR-ICD $I^{(ICD)}(q)$. Green: GFP $\left|A^{(GFP)}(q)\right|^2$. (Note the cross-correlation terms are not shown here but they are included in the calculation for the total scattering intensity.)
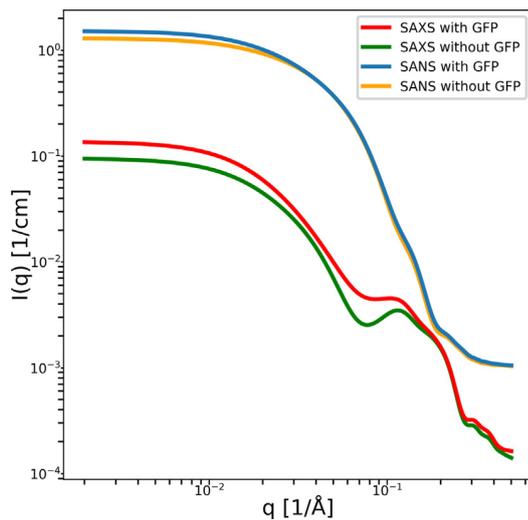


**Fig. 4.** Impact of including a marker protein. Theoretical SAXS and SANS profiles of the full complex (Eqn. 14) with and without the GFP fused to the C-terminus of the GHR.

the signal is negligible and should not impact the subsequent data analysis or modeling [63,64]. Structural descriptors about the membrane protein, such as $R_g$, molecular weight and pair-distance distributions could be retrieved directly. Furthermore, modeling would be much less ambiguous since the nanodisc would not need to be included. Thereby, the number of free parameters is greatly reduced and parameters can be refined from the data with greater certainty and a reduced risk of overfitting. This is therefore the optimal situation for collecting structural information about the GHR-GFP fusion.

Arguably, it could be counterproductive to perform standard SANS on this kind of sample when SAXS is usually available. SAXS is effective for investigating complexes of lipids and proteins together since both components give a strong signal. Furthermore, synchrotron-based SAXS instruments are generally more accessible, have higher flux, require less sample and provide higher qual-

ity data. Based on information theory, SAXS data sets contain quantitatively higher amounts of information compared to SANS data sets for samples such as these, reflecting the different noise levels of the experiments [66]. Increasing the ratio of $H_2O$ to $D_2O$ in SANS experiments increases incoherent background scattering and therefore further decreases the information content of the data set. Certainly confidence in model parameters can be strengthened by coupling SAXS data with high-quality SANS data, though the extent of this often depends on the specific parameter of interest and the chosen contrast situation [66]. Recently Larsen et al. [67] evaluated how much structural information on a flexible multido-main protein can be obtained from a SAXS data set alone and in combination with three additional contrast variation SANS data sets. The study shows that, for their particular system, the inclu-sion of the SANS data had a limited impact on their results. They emphasise the influence of a good SAXS data set, and stress the importance of choosing optimal contrast situations so that the SANS data has low signal-to-noise and carries different structural information than SAXS.

To that end, SANS collected from a stealth nanodisc system (in 100% $D_2O$) is undoubtedly forceful for obtaining structural infor-mation on an embedded membrane protein. Josts et al. [64] demonstrate that a much more complete and precise picture of membrane protein MsbA is gained using the stealth nanodisc com-pared to the information that could be gained from SAXS. On the other hand, preparation of stealth nanodiscs is labour intensive and the samples precious since the phospholipid head and tail groups require specific levels of deuteration respectively.

### 3.3. Investigating the compactness of the IDR

We want to investigate how sensitive our model of the full-length GHR-GFP in a nanodisc is in the three different contrast sit-uations to capturing varying degrees of compactness. We compare theoretical scattering profiles with different values of $R_g^{(ICD)}$ for the Gaussian random coil, Eqn. 14. Calculations are plotted in Fig. 6. The inset in Fig. 6(a) demonstrates the way increasing the $R_g^{(ICD)}$ causes the Debye function to decay earlier and therefore interfere less with the scattering arising from the other components of the system.
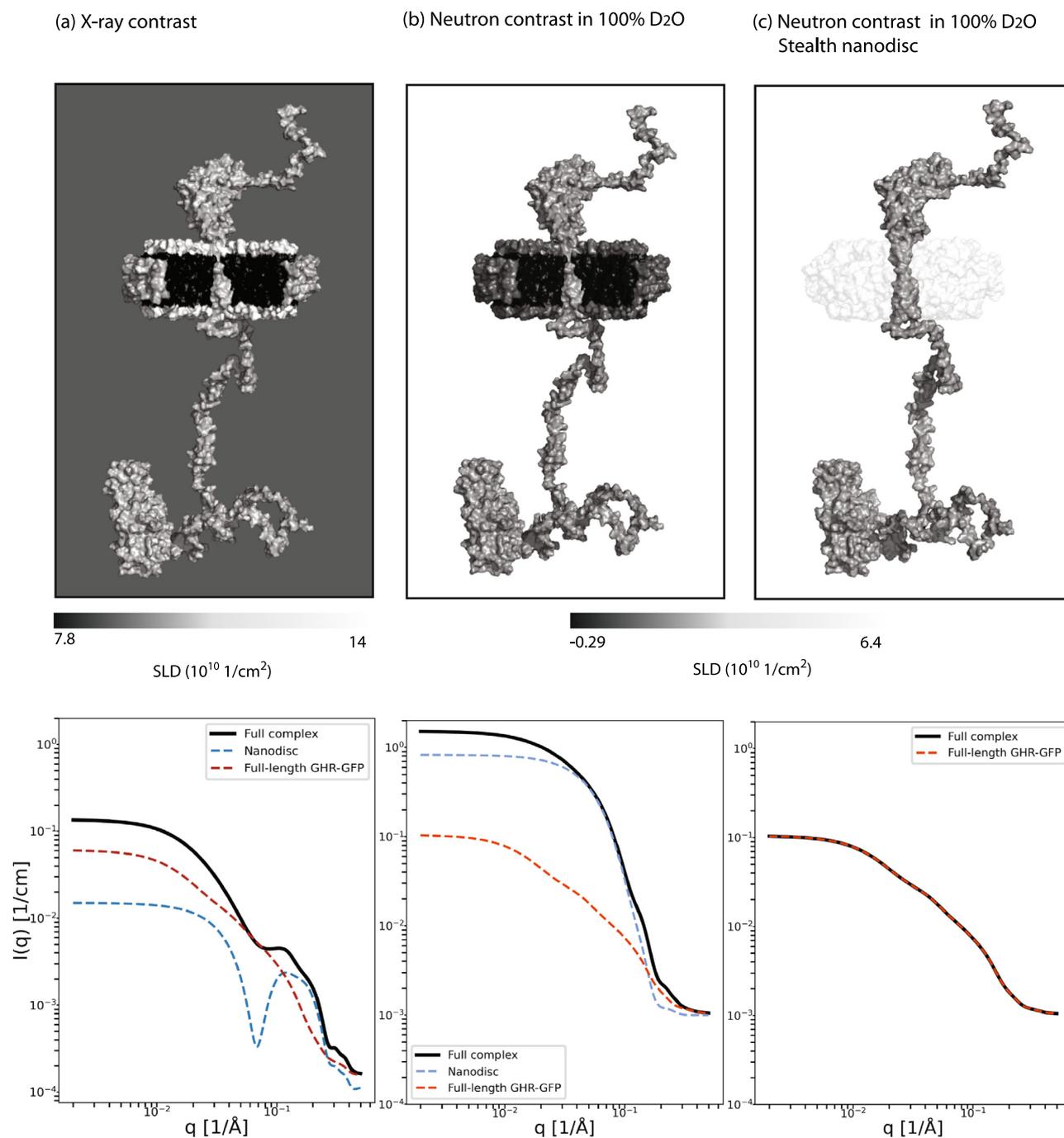
**Fig. 5.** Top: Illustration of the complex in solvent in three different contrast situations. The grey scale corresponds to the SLD of the different subunits: POPC headgroups, POPC tailgroups, MSP, GHR-GFP and solvent. Lighter colours indicate a high SLD and dark colours indicate a low SLD. The protein is represented by a single structure of the full-length GHR-GFP built by Kassem et al. [1]. The nanodisc was generated using the *CHARMM − GUI NanodiscBuilder*[59,60]. The nanodisc has a cut-out so that the phospholipid bilayer is visible. Bottom: Theoretical scattering profiles calculated with the parameters in Table 1. The total combined scattering intensity (Eqn. 14) is plotted in black. The auto-correlation terms of the nanodisc and the GHR-GFP are plotted in red and blue respectively. (Note the cross-correlation terms are not shown here but they are included in our calculations).

Unsurprisingly, SANS with the original nanodiscs, Fig. 6(b), is the least sensitive to varying the $R_g^{(ICD)}$ since the overall scattering profile is hugely dominated by the nanodisc contribution. It would be difficult to extract $R_g^{(ICD)}$ from a real data set with certainty under these experimental conditions.

SANS with stealth nanodiscs is much more sensitive to changes in $R_g^{(ICD)}$. This is where the GHR-ICD makes up the largest proportion of the respective total scattering signal. Perhaps surprisingly, SAXS seems to perform equally well, despite the SAXS signal additionally containing a significant contribution from the nanodisc. In

**Fig. 6.** Theoretical scattering profiles calculated with the parameters in Table 1 but with varying values of the $R_g$ of the Gaussian random coil. The inset is the Debye function in SAXS contrast. Note that the shape of the Debye is identical in SANS, but the forward intensity is different.
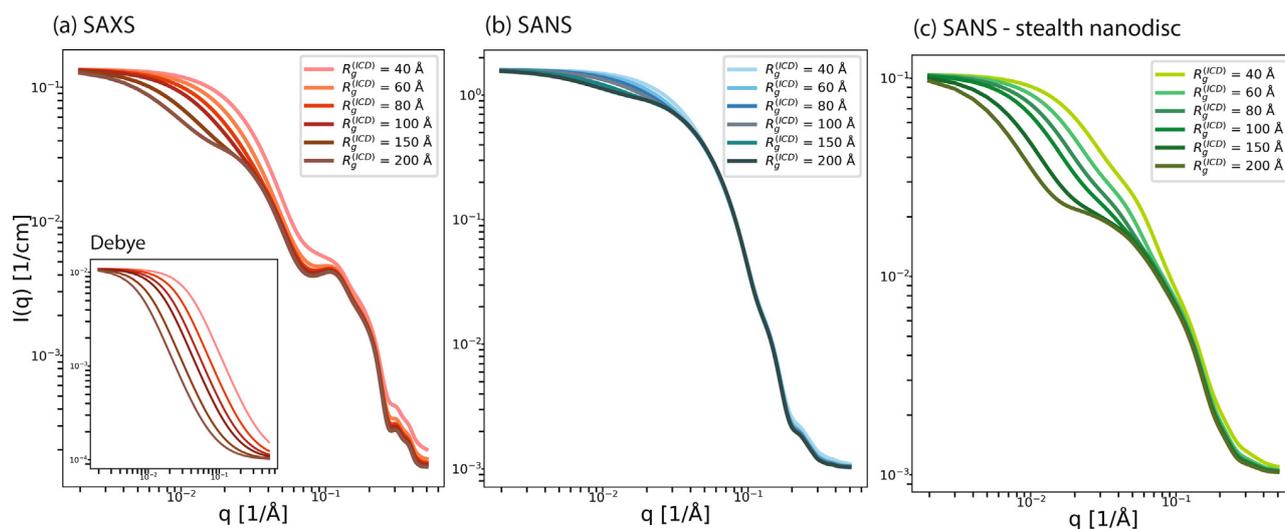
both cases, variations of $R_g^{(ICD)}$ are easily visible in the Guinier range. The nanodisc features do not complicate the SAXS profile until the mid to high-$q$ range.

## 4. Application to experimental SAXS data

The nanodisc-embedded GHR-GFP was studied previously by size-exclusion chromatography inline with SAXS (SEC-SAXS), where particles are separated according to their size to ensure SAXS is collected on a monodisperse sample [1]. Monodispersity is crucial to extract meaningful results from a model refinement since each particle in the sample adds to the resulting scattering profile. Even a small amount of oligomers, aggregates or contaminants can significantly distort the SAXS profile meaning the structural model, which is based on an idealised sample, becomes inconsistent with the data [68].

The averaged SAXS profile from a small uniform fraction of the SEC peak is plotted in Fig. 7 alongside the refinement of our model. Our model describes the data excellently, demonstrating the utility
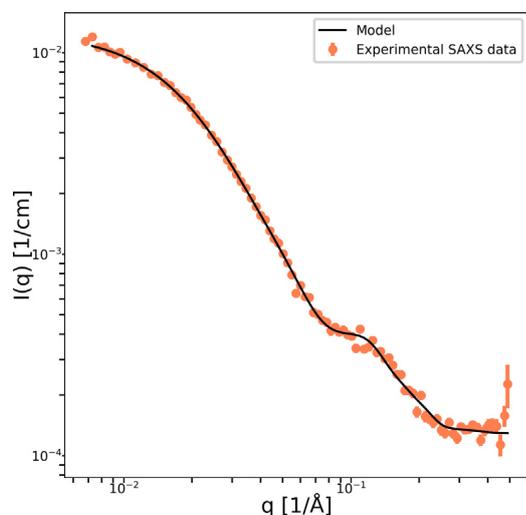
of our approach for real SAS data from membrane protein-nanodisc complexes. Moreover our refined value for the $R_g^{ICD}$ of 73 ± 13 Å is in close agreement with MD simulations of the complex as well as model-free analysis of SAS data from the isolated GHR-ICD, showing the Gaussian random coil can provide reliable descriptions of disordered proteins.

## 5. Conclusions

In this work we present an advanced modeling scheme for obtaining low-resolution descriptions of membrane proteins with large intrinsically disordered regions embedded in a ND from small-angle scattering data. In addition to being able to account for both rigid and flexible domains, we utilise the concept of a 'distribution volume' which allows scattering from domains without a fixed position to be properly accounted for.

Molecular dynamics simulations of membrane proteins in lipid bilayers are at the forefront of investigating the dynamics and function of the protein [69–71,1]. In order to validate MD trajectories with experimental SAS data where the membrane protein is loaded into a nanodisc, however, models must be able to account for the nanodisc also so that the entire experimental structure is refined from the data.

Our strategy consists of splitting the system into distinct components, calculating the scattering amplitude arising from each component independently, and then combining the scattering amplitudes to give the total theoretical scattering intensity. This spherical harmonics-based approach allows one to quickly assess the scattering contributions of each of the components of the investigated complexes. Furthermore, an attractive attribute of our strategy is that users can select amplitudes of components they would like to include and disregard those which are not relevant in order to tune the model to describe their own system.

The recently published results from refinement of our full-length model of the growth hormone receptor against experimental SAXS data testifies to the validity of our model [1]. The model describes these experimental data excellently, as shown in Fig. 7. Notably the refined $R_g$ of the GHR-ICD is in close agreement with other analysis from the same study which corroborates modeling conformations of intrinsic disorder through the simple idealised Gaussian random coil model.

We demonstrated the effect of the GFP attached to the C-terminal of a large membrane protein. The effect of the inclusion



**Fig. 7.** The orange points are experimental SEC-SAXS data from the nanodisc-embedded GHR with GFP. The black curve is the fit by our model to the data. The fit has $\chi^2$ = 1.5. The refined structural parameters can be found in Kassem et al. [1]. The protein concentration is estimated to be 0.91 $\mu$M.

of the GFP is clearly visible with SAXS, where it significantly diminishes distinguishing features in the scattering profile. Glancing at Fig. 4, we recommend where possible removing such irrelevant scattering contributions prior to SAXS experiments to ease data analysis. With SANS this is perhaps less important when the protein signal is regardless masked by the presence of phospholipids.

Finally, we investigated the sensitivity of our model to different levels of compaction/extension of the random coil under different contrast situations. Fig. 6 shows that SANS studies of membrane proteins embedded in the original hydrogenated nanodiscs are quite uninformative since the resulting profile is largely dominated by scattering from the phospholipids. Adopting stealth nanodiscs instead makes it far easier to distinguish between different values of $R_g^{(ICD)}$. Perhaps surprisingly, however, is SAXS is equally as sensitive to $R_g^{(ICD)}$. Still, SANS data from membrane proteins in stealth nanodiscs provides cleaner data and a more complete description of the membrane protein. Furthermore, the data can be modelled with higher certainty due to the reduction in the number of free parameters when the nanodisc is not present. However, this comes at the cost of a more intensive sample preparation.

Generally we try to demonstrate the importance of making theoretical model calculations prior to preparing challenging samples in the lab. It should always be carefully considered whether SANS experiments with contrast variations really adds additional information that cannot be obtained from a simpler SAXS experiment.

## 6. Authors' contributions

**Abigail Barclay:** Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualisation. **Birthe Kragelund:** Conceptualization, Investigation, Funding acquisition. **Lise Arleth:** Conceptualization, Methodology, Resources, Supervision, Funding acquisition. **Martin Cramer Pedersen:** Conceptualization, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jcis.2022.12.024.

## References

[1] N. Kassem, R. Araya-Secchi, K. Bugge, A. Barclay, H. Steinocher, A. Khondker, Y. Wang, A.J. Lenard, J. Bürck, C. Sahin, A.S. Ulrich, M. Landreh, M.C. Pedersen, M. C. Rheinstädter, P.A. Pedersen, K. Lindorff-Larsen, L. Arleth, B.B. Kragelund, Order and disorder—An integrative structure of the full-length human growth hormone receptor, Sci. Adv. 7 (27) (2021) eabh3805, https://doi.org/10.1126/sciadv.abh3805.

[2] V.N. Uversky, Intrinsically disordered proteins and their Mysterious (meta)physics, Front. Phys. 7 (FEB) (2019) 10. doi:10.3389/FPHY.2019.00010/BIBTEX.

[3] A. Tantos, K.H. Han, P. Tompa, Intrinsic disorder in cell signaling and gene transcription, Mol. Cell. Endocrinol. 348 (2) (2012) 457–465, https://doi.org/10.1016/J.MCE.2011.07.015.

[4] J. Bürgi, B. Xue, V.N. Uversky, F.G. Van Der Goot, Intrinsic Disorder in Transmembrane Proteins: Roles in Signaling and Topology Prediction, PLOS ONE 11 (7) (2016) e0158594, https://doi.org/10.1371/JOURNAL.PONE.0158594.

[5] H.J. Dyson, P.E. Wright, Unfolded Proteins and Protein Folding Studied by NMR, Chem. Rev. 104 (8) (2004) 3607–3622, https://doi.org/10.1021/CR030403S.

[6] M.R. Jensen, M. Zweckstetter, J.R. Huang, M. Blackledge, Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy, Chem. Rev. 114 (13) (2014) 6632–6660, https://doi.org/10.1021/CR400688U.

[7] N.T. Johansen, F.G. Tidemand, M.C. Pedersen, L. Arleth, Travel light: Essential packing for membrane proteins with an active lifestyle//// Preprint of manuscript under review - May 2022, Biochimie (May) (8 2022). doi:10.1016/J.BIOCHI.2022.07.014.

[8] T.H. Bayburt, Y.V. Grinkova, S.G. Sligar, Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins, Nano Lett. 2 (8) (2002) 853–856, https://doi.org/10.1021/nl025623k.

[9] I.G. Denisov, Y.V. Grinkova, A.A. Lazarides, S.G. Sligar, Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size, J. Am. Chem. Soc. 126 (11) (2004) 3477–3487, https://doi.org/10.1021/ja0393574.

[10] I.G. Denisov, S.G. Sligar, Nanodiscs in Membrane Biochemistry and Biophysics, Chem. Rev. 117 (6) (2017) 4669–4713, https://doi.org/10.1021/ACS.CHEMREV.6B00690.

[11] P. Bernadó, D.I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering, Mol. BioSyst. 8 (1) (2011) 151–167, https://doi.org/10.1039/C1MB05275F.

[12] T.N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estaña, J. Cortés, N. Sibille, P. Bernadó, Small-angle scattering studies of intrinsically disordered proteins and their complexes, Curr. Opin. Struct. Biol. 42 (2017) 15–23, https://doi.org/10.1016/J.SBI.2016.10.011.

[13] A. Guinier, G. Fournet, Small-Angle Scattering of X-rays, Acta Crystallogr. A 9 (10) (1956) 839, https://doi.org/10.1107/s0365110x56002400.

[14] W. Zheng, R.B. Best, An Extended Guinier Analysis for Intrinsically Disordered Proteins, J. Mol. Biol. 430 (16) (2018) 2540–2553, https://doi.org/10.1016/j.jmb.2018.03.007.

[15] A.G. Kikhney, D.I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins, FEBS Lett. 589 (19) (2015) 2570–2577, https://doi.org/10.1016/J.FEBSLET.2015.08.027.

[16] P. Bernadó, E. Mylonas, M.V. Petoukhov, M. Blackledge, D.I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering, J. Am. Chem. Soc. 129 (17) (2007) 5656–5664, https://doi.org/10.1021/JA069124N/h.

[17] G. Tria, H.D.T. Mertens, M. Kachala, D.I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, IUCrJ 2 (2015) 207–217, https://doi.org/10.1107/S205225251500202X.

[18] M. Pelikan, G.L. Hura, M. Hammel, Structure and flexibility within proteins as identified through small angle X-ray scattering, General physiology and biophysics 28 (2) (2009) 174. doi:10.4149/gpb{\_}2009{\_}02{\_}174.

[19] L.D. Antonov, S. Olsson, W. Boomsma, T. Hamelryck, Bayesian inference of protein ensembles from SAXS data, PCCP 18 (8) (2016) 5832–5838, https://doi.org/10.1039/C5CP04886A.

[20] M.R. Hermann, J.S. Hub, SAXS-Restrained Ensemble Simulations of Intrinsically Disordered Proteins with Commitment to the Principle of Maximum Entropy, J. Chem. Theory Comput. 15 (9) (2019) 5103–5115, https://doi.org/10.1021/ACS.JCTC.9B00338.

[21] U.R. Shrestha, J.C. Smith, L. Petridis, Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations, Communications Biology 2021 4:1 4 (1) (2021) 1–8. doi:10.1038/s42003-021-01759-1.

[22] A.B. Ward, A. Sali, I.A. Wilson, Integrative structural biology, Science 339 (6122) (2013) 913–915, https://doi.org/10.1126/SCIENCE.1228565.

[23] M.P. Rout, A. Sali, Principles for Integrative Structural Biology Studies, Cell 177 (6) (2019) 1384–1403, https://doi.org/10.1016/J.CELL.2019.05.016.

[24] J.S. Pedersen, Neutrons, X-rays and light: scattering methods applied to soft condensed matter–Chapter 16 Modelling of Small-Angle Scattering Data from Colloids and Polymer Systems, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter (2002) 391–420.

[25] N. Skar-Gislinge, L. Arleth, Small-angle scattering from phospholipid nanodiscs: Derivation and refinement of a molecular constrained analytical model form factor, PCCP 13 (8) (2011) 3161–3170, https://doi.org/10.1039/c0cp01074j.

[26] N. Kučerka, M.P. Nieh, J. Katsaras, Small-Angle Scattering from Homogenous and Heterogeneous Lipid Bilayers, Advances in Planar Lipid Bilayers and Liposomes 12 (C) (2010) 201–235. doi:10.1016/B978-0-12-381266-7.00008-0.

[27] H.D. Andersen, C. Wang, L. Arleth, G.H. Peters, P. Westh, Reconciliation of opposing views on membrane-sugar interactions, Proceedings of the National Academy of Sciences of the United States of America 108 (5) (2011) 1874–1878, https://doi.org/10.1073/PNAS.1012516108.

[28] P. Mahapatra, A. Rasheed, P.S. Goyal, J.R. Bellare, Structure of Micelles Calcium Didodecyl Sulfate: A SAXS Study, J Nanomed Nanotechnol 6 (4) (2015) 307, https://doi.org/10.4172/2157-7439.1000307.

[29] N. Skar-Gislinge, J.B. Simonsen, K. Mortensen, R. Feidenhans'l, S.G. Sligar, B. Lindberg Møller, T. Bjørnholm, L. Arleth, Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the

lipids and the protein, Journal of the American Chemical Society 132 (39) (2010) 13713–13722. doi:10.1021/ja1030613.

[30] N.T. Johansen, A. Luchini, F.G. Tidemand, S. Orioli, A. Martel, L. Porcar, L. Arleth, M.C. Pedersen, Structural and Biophysical Properties of Supercharged and Circularized Nanodiscs, Langmuir 37 (22) (2021) 6681–6690, https://doi.org/10.1021/acs.langmuir.1c00560.

[31] D. Svergun, C. Barberato, M.H. Koch, CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Crystallogr. 28 (6) (1995) 768–773, https://doi.org/10.1107/S0021889895007047.

[32] M.V. Petoukhov, D.I. Svergun, Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data, Biophys. J . 89 (2) (2005) 1237–1250, https://doi.org/10.1529/BIOPHYSJ.105.064154.

[33] S.A.R. Kynde, N. Skar-Gislinge, M.C. Pedersen, S.R. Midtgaard, J.B. Simonsen, R. Schweins, K. Mortensen, L. Arleth, Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment, Acta Crystallogr., Sect D: Biol. Crystallogr. 70 (2) (2014) 371–383, https://doi.org/10.1107/S1399004713028344.

[34] J.S. Pedersen, M.C. Gerstenberg, Scattering Form Factor of Block Copolymer Micelles, Macromolecules 29 (4) (1996) 1363–1365, https://doi.org/10.1021/ma9512115.

[35] G.W. Haxholm, L.F. Nikolajsen, J.G. Olsen, J. Fredsted, F.H. Larsen, V. Goffin, S.F. Pedersen, A.J. Brooks, M.J. Waters, B.B. Kragelund, Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes, Biochem. J. 468 (3) (2015) 495–506, https://doi.org/10.1042/BJ20141243.

[36] P. Seiffert, K. Bugge, M. Nygaard, G.W. Haxholm, J.H. Martinsen, M.N. Pedersen, L. Arleth, W. Boomsma, B.B. Kragelund, Orchestration of signaling by structural disorder in class 1 cytokine receptors, Cell Communication and Signaling 18 (1) (2020) 1–30, https://doi.org/10.1186/s12964-020-00626-6.

[37] M. Shiroishi, M. Moriya, T. Ueda, Micro-scale and rapid expression screening of highly expressed and/or stable membrane protein variants in Saccharomyces cerevisiae, Protein Sci. 25 (10) (2016) 1863–1872, https://doi.org/10.1002/PRO.2993.

[38] D. Drew, S. Newstead, Y. Sonoda, H. Kim, G. von Heijne, S. Iwata, GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae, Nature Protocols 2008 3:5 3 (5) (2008) 784–798. doi:10.1038/nprot.2008.44.

[39] S. Newstead, H. Kim, G. Von Heijne, S. Iwata, D. Drew, High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae, Proc. Nat. Acad. Sci. 104 (35) (2007) 13936–13941, https://doi.org/10.1073/PNAS.0704546104.

[40] J.S. Pedersen, Analysis of small-angle scattering data from colloids and polymer solutions: Modeling and least-squares fitting, Adv. Colloid Interface Sci. 70 (1–3) (1997) 171–210, https://doi.org/10.1016/S0001-8686(97)00312-6.

[41] N. Skar-Gislinge, S.A.R. Kynde, I.G. Denisov, X. Ye, I. Lenov, S.G. Sligar, L. Arleth, Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment, Acta Crystallogr., Sect D: Biol. Crystallogr. 71 (2015) 2412–2421, https://doi.org/10.1107/S1399004715018702.

[42] M.J. Mohlenkamp, A Fast Transform for Spherical Harmonics, Journal of Fourier Analysis and Applications 5 (2–3) (1999) 158–184, https://doi.org/10.1007/bf01261607.

[43] M.J. Mohlenkamp, A User's Guide to Spherical Harmonics, 2016.

[44] J.S. Pedersen, Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores, Journal of Applied Crystallography 33 (3 I) (2000) 637–640. doi:10.1107/S0021889899012248.

[45] P. Debye, Zerstreuung von Röntgenstrahlen, Ann. Phys. 351 (6) (1915) 809–823, https://doi.org/10.1002/ANDP.19153510606.

[46] B. Hammouda, Structure factor for starburst dendrimers, J. Polym. Sci., Part B: Polym. Phys. 30 (12) (1992) 1387–1390, https://doi.org/10.1002/POLB.1992.090301209.

[47] R. Høiberg-Nielsen, P. Westh, L.K. Skov, L. Arleth, Interrelationship of steric stabilization and self-crowding of a glycosylated protein, Biophys. J . 97 (5) (2009) 1445–1453, https://doi.org/10.1016/j.bpj.2009.05.045.

[48] M.C. Pedersen, L. Arleth, K. Mortensen, WillItFit: a framework for fitting of constrained models to small-angle scattering data, J. Appl. Crystallogr. 46 (6) (2013) 1894–1898, https://doi.org/10.1107/S0021889813026022.

[49] J. Als-Nielsen, D. McMorrow, Elements of Modern X-ray Physics: Second Edition, John Wiley and Sons, 2011. doi:10.1002/9781119998365.

[50] M. Ormö, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, S.J. Remington, Crystal Structure of the Aequorea victoria Green Fluorescent Protein, Science 273 (5280) (1996) 1392–1395, https://doi.org/10.1126/SCIENCE.273.5280.1392.

[51] N. Kučerka, S. Tristram-Nagle, J.F. Nagle, Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains, J. Membr. Biol. 208 (3) (2006) 193–202, https://doi.org/10.1007/s00232-005-7006-8.

[52] S. Bibow, Y. Polyhach, C. Eichmann, C.N. Chi, J. Kowal, S. Albiez, R.A. McLeod, H. Stahlberg, G. Jeschke, P. Güntert, R. Riek, Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I, Nature Structural and Molecular Biology 24 (2) (2017) 187–193, https://doi.org/10.1038/nsmb.3345.

[53] E. Mylonas, D.I. Svergun, Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering, J. Appl. Crystallogr. 40 (SUPPL. 1) (2007) 245–249, https://doi.org/10.1107/S002188980700252X.

[54] F. Sears, Special Feature Neutron scattering lengths and cross sections, Methods 3 (3) (1992) 26–37.

[55] C.M. Hebling, C.R. Morgan, D.W. Stafford, J.W. Jorgenson, K.D. Rand, J.R. Engen, Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, Anal. Chem. 82 (13) (2010) 5415–5419, https://doi.org/10.1021/ac100962c.

[56] M.C. Pedersen, Y. Wang, F.G. Tidemand, A. Martel, K. Lindorff-Larsen, L. Arleth, PSX: Protein-Solvent Exchange - Software for calculation of deuterium-exchange effects in SANS measurements from protein coordinates, bioRxiv (2019) 1–24. doi:10.1101/609859.

[57] R.D.B. Fraser, T.P. MacRae, E. Suzuki, An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules, J. Appl. Crystallogr. 11 (6) (1978) 693–694, https://doi.org/10.1107/s0021889878014296.

[58] I.G. Denisov, M.A. McLean, A.W. Shaw, Y.V. Grinkova, S.G. Sligar, Thermotropic Phase Transition in Soluble Nanoscale Lipid Bilayers, J. Phys. Chem. B 109 (32) (2005) 15580–15588, https://doi.org/10.1021/jp051385g.

[59] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (11) (2008) 1859–1865, https://doi.org/10.1002/JCC.20945.

[60] Y. Qi, J. Lee, J.B. Klauda, W. Im, CHARMM-GUI Nanodisc Builder for modeling and simulation of various nanodisc systems, J. Comput. Chem. 40 (7) (2019) 893–899, https://doi.org/10.1002/JCC.25773.

[61] B. Jacrot, The study of biological structures by neutron scattering from solution, Rep. Prog. Phys. 39 (10) (1976) 911–953, https://doi.org/10.1088/0034-4885/39/10/001.

[62] C.M. Jeffries, M.A. Graewert, C.E. Blanchet, D.B. Langley, A.E. Whitten, D.I. Svergun, Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments, Nature Protocols 2016 11:11 11 (11) (2016) 2122–2153. doi:10.1038/nprot.2016.113.

[63] S. Maric, N. Skar-Gislinge, S. Midtgaard, M.B. Thygesen, J. Schiller, H. Frielinghaus, M. Moulin, M. Haertlein, V.T. Forsyth, T.G. Pomorski, L. Arleth, Stealth carriers for low-resolution structure determination of membrane proteins in solution, Acta Crystallogr., Sect D: Biol. Crystallogr. 70 (2) (2014) 317–328, https://doi.org/10.1107/S1399004713027466.

[64] I. Josts, J. Nitsche, S. Maric, S. Busch, V.T. Forsyth, Conformational States of ABC Transporter MsbA in a Lipid Environment Investigated by Small-Angle Scattering Using Stealth Carrier Nanodiscs, Structure 26 (8) (2018) 1072–1079, https://doi.org/10.1016/j.str.2018.05.007.

[65] N.T. Johansen, M. Bonaccorsi, T. Bengtsen, A.H. Larsen, F.G. Tidemand, M.C. Pedersen, P. Huda, J. Berndtsson, T. Darwish, N.R. Yepuri, A. Martel, T.G. Pomorski, A. Bertarello, M. Sansom, M. Rapp, R. Crehuet, T. Schubeis, K. Lindorff-Larsen, G. Pintacuda, L. Arleth, Mg2+-dependent conformational equilibria in CorA and an integrated view on transport regulation, eLife 11 (2 2022) doi:10.7554/elife.71887.

[66] M.C. Pedersen, S.L. Hansen, B. Markussen, L. Arleth, K. Mortensen, Quantification of the information in small-angle scattering data, J. Appl. Crystallogr. 47 (6) (2014) 2000–2010, https://doi.org/10.1107/S1600576714024017.

[67] A.H. Larsen, J.S. Pedersen, L. Arleth, Assessment of structure factors for analysis of small-angle scattering data from desired or undesired aggregates, J. Appl. Crystallogr. 53 (2020) 991–1005, https://doi.org/10.1107/S1600576720006500.

[68] J. Pérez, A. Thureau, P. Vachette, SEC-SAXS: Experimental set-up and software developments build up a powerful tool, Academic Press, 2022, https://doi.org/10.1016/bs.mie.2022.08.024.

[69] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, S.J. Marrink, The MARTINI Coarse-Grained Force Field: Extension to Proteins, J. Chem. Theory Comput. 4 (5) (2008) 819–834, https://doi.org/10.1021/CT700324X.

[70] K. Bugge, E. Papaleo, G.W. Haxholm, J.T.S. Hopper, C.V. Robinson, J.G. Olsen, K. Lindorff-Larsen, B.B. Kragelund, A combined computational and structural model of the full-length human prolactin receptor, Nature Communications 7 (2016), https://doi.org/10.1038/ncomms11578.

[71] R. Yeasmin, A. Brewer, L.R. Fine, L. Zhang, Molecular Dynamics Simulations of Human Beta-Defensin Type 3 Crossing Different Lipid Bilayers, ACS Omega 6 (21) (2021) 13926–13939, https://doi.org/10.1021/ACSOMEGA.1C01803/SUPPL_FILE/AO1C01803_SI_001.PDF.

[72] R.K. Das, R.V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, PNAS 110 (33) (2013), https://doi.org/10.1073/pnas.1304749110.

[73] Farag M, Cohen SR, Borcherds WM, Bremer A, Mittag T, Pappu RV. Condensates formed by prion-like low-complexity domains have small-world network structures and interfaces defined by expanded conformations. Nat Commun. 2022 Dec 13;13(1):7722. https://doi.org/10.1038/s41467-022-35370-7.

[74] NS González-Foutel, J Glavina, WM Borcherds, M Safranchik, S Barrera-Vilarmau, A Sagar, A Estaña, A Barozet, NA Garrone, G Fernandez-Ballester, C Blanes-Mira, IE Sánchez, G de Prat-Gay, J Cortés, P Bernadó, RV Pappu, AS Holehouse, GW Daughdrill, LB Chemes, Conformational buffering underlies functional selection in intrinsically disordered protein regions, Nat Struct Mol Biol 29 (8) (2022 Aug) 781–790, https://doi.org/10.1038/s41594-022-00811-w.

# Supporting Information
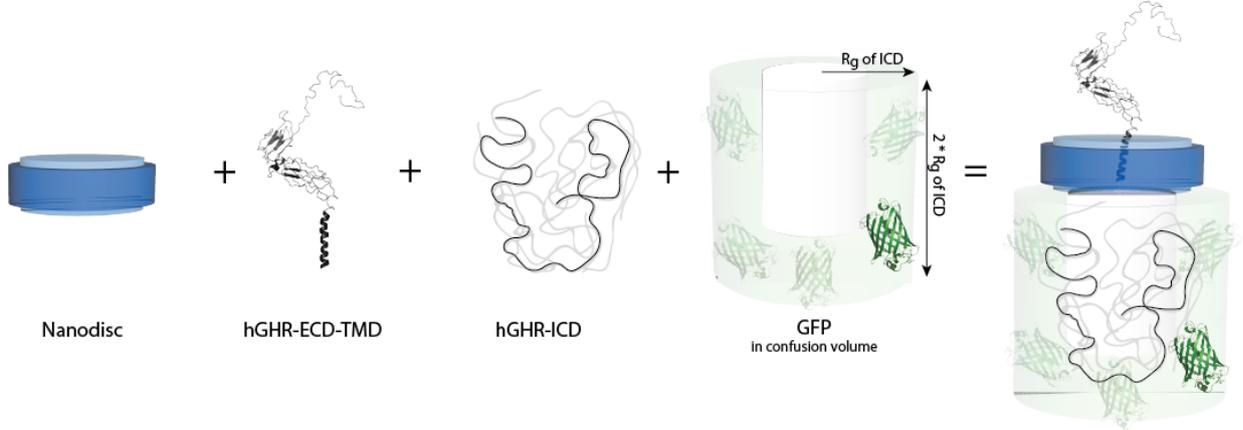


Figure SI.1: Construction of the mode with a cylindrical-shell distribution volume. In line with the dimensions of the sphere, the larger disc has radius $R_{\text{large}} = R_g^{(\text{ICD})}$ and height $h_{\text{large}} = 2 \cdot R_g^{(\text{ICD})}$, and the smaller disc has radius $R_{\text{small}} = 0.5 \cdot R_g^{(\text{ICD})}$ and height $h_{\text{small}} = 1.5 \cdot R_g^{(\text{ICD})}$. The shell therefore has a thickness of $0.5 \cdot R_g^{(\text{ICD})}$ all round.

## SI.1   Scattering lengths

| Component | Chemical composition | X-ray scattering length [cm] | Neutron scattering length [cm] |
|:---:|:---:|:---:|:---:|
| Solvent | $H_2O$ | $2.82 \cdot 10^{-12}$ | - |
| Solvent | $D_2O$ | - | $1.92 \cdot 10^{-12}$ |
| GHR-ICD | $C_{1669}H_{2631}N_{451}O_{559}S_{16}$ | $5.79 \cdot 10^{-9}$ | $1.5 \cdot 10^{-9}$ (100% exch.) |
| MSP1D1 | $C_{971}H_{1561}N_{277}O_{302}S_3$ | $3.32 \cdot 10^{-9}$ | $6.33 \cdot 10^{-10}$ (50% exch.) |
| POPC headgroups | $C_{10}H_{18}NO_8P$ | $4.62 \cdot 10^{-11}$ | $6.00 \cdot 10^{-12}$ |
| POPC tailgroups | $C_{30}H_{58}$ | $6.71 \cdot 10^{-11}$ | $-1.77 \cdot 10^{-12}$ |
| POPC methyl | $C_2H_6$ | $5.08 \cdot 10^{-12}$ | $-9.16 \cdot 10^{-13}$ |

Table SI.1: Chemical compositions, x-ray scattering lengths and neutron scattering lengths for the required components in the modelling.
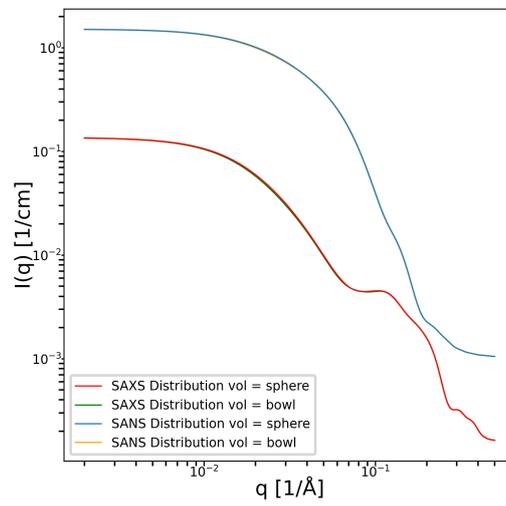
Figure SI.2: Choice of distribution volume. Theoretical SAXS and SANS profiles of the full complex with either a sphere or a cylindrical-shell for the distribution volume.

<div style="text-align: right;">

# 10

</div>

## STRUCTURAL CHARACTERISATION OF α-SYNUCLEIN:MEMBRANE INTERACTIONS AND THE RESULTING AGGREGATION USING SMALL-ANGLE SCATTERING

### CONTRIBUTIONS

In this manuscript SAXS and contrast variation SANS were used to investigate the initial interaction between α-Synuclein and negatively charged lipids, in an attempt to shed light on the subsequent fibrillation process. The data indicates a rearrangement of the lipid structures into discs or rods. A similar transformation has been documented in the literature and so we focus on what SAS can tell us compared to other techniques: fast time scales with stopped-flow experiments, long time scales to capture fibrillation, reversibility of the system and temperature variation.

The project was conceptualised and led by Céline Galvagnion and Alexander Buell, who also supervised the sample preparation and data collection. I performed all of the SAXS and SANS data analysis including model-free analysis and building the analytical models. I made all of the figures and wrote a large part of the manuscript. I was supervised by Lise Arleth and Martine Cramer Pedersen

An old version of this manuscript is available on ChemRXiv but after a productive round of reviews some changes have been implemented.

# Structural characterisation of $\alpha$-synuclein -membrane interactions and the resulting aggregation using small angle scattering

Céline Galvagnion,*,† Abigail Barclay,‡ Katarzyna Makasewicz,¶ Frederik Ravnkilde Marlet,† Martine Moulin,§ Juliette Devos,§ Sara Linse,¶ Anne Martel,§ Lionel Porcar,§ Emma Sparr,‖ Martin Cramer Pedersen,‡ Felix Roosen-Runge,⊥ Lise Arleth,‡ and Alexander K. Buell*,#

†*Department of Drug Design and Pharmacology, University of Copenhagen, 2100 Copenhagen, Denmark*

‡*Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark*

¶*Department of Biochemistry and Structural Biology, Lund University , SE22100 Lund, Sweden*

§*Institut Laue-Langevin, 71 avenue des Martyrs, 38042 Grenoble, France*

‖*Division of Physical Chemistry, Center for Chemistry and Chemical Engineering, Lund University, P.O. Box 124, SE-22100, Lund, Sweden*

⊥*Department of Biomedical Sciences and Biofilms Research Center for Biointerfaces, Malmö University, Malmö, Sweden*

#*Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

E-mail: celine.galvagnion@sund.ku.dk; alebu@dtu.dk

**Abstract**

1

The presence of amyloid fibrils is a hallmark of several neurodegenerative diseases. Some amyloidogenic proteins, such as $\alpha$-synuclein and amyloid $\beta$, can interact with lipids, and this interaction can strongly favour the formation of amyloid fibrils. In particular the primary nucleation step, i.e. the *de novo* formation of amyloid fibrils, has been shown to be accelerated by lipids. However, the exact mechanism of this acceleration is still mostly unclear. Here we use a range of scattering methods, such as dynamic light scattering (DLS) and small angle X-ray and neutron scattering (SAXS and SANS) to obtain structural information on the binding of $\alpha$-synuclein to model membranes formed from negatively charged lipids and their co-assembly into amyloid fibrils. We find that the model membranes do not simply act as a surface that catalyses the nucleation reaction, but that lipid molecules take an active role in the reaction. The binding of $\alpha$-synuclein to the model membranes immediately induces a major structural change in the lipid assembly, which leads to a break-up into small and mostly disc- or rod-like lipid-protein particles. This transition can be reversed by temperature changes or proteolytic protein removal. Incubation of these small lipid-$\alpha$-synuclein particles for several hours, however, yields amyloid fibril formation, whereby the lipids are incorporated into the fibrils.

# Introduction

Protein aggregation into ordered fibrillar structures, amyloid fibrils, is a hallmark of a range of severe disorders, such as Alzheimer's, Parkinson's and the prion diseases.[1] $\alpha$-synuclein is a small highly conserved pre-synaptic protein proposed to be involved in synaptic plasticity and whose aggregation into amyloid fibrils is the hallmark of Parkinson's disease.[2,3] $\alpha$-synuclein is natively disordered in solution,[4,5] but in the presence of lipid bilayers and various surfactants the protein has been shown to adopt an $\alpha$-helical conformation[6] which is proposed to be important for its biological role.[7] It has been known for about two decades that interactions between $\alpha$-synuclein and lipid bilayers influence the kinetics of amyloid

formation of this protein.[8,9]

$\alpha$-synuclein:lipid mixtures can adopt a wide variety of structures whose dimensions and morphologies depend on the lipid-to-protein ratio and the charge, phase state and shape (vesicles vs. supported bilayer) of the lipid system as well as the aggregation state of $\alpha$-synuclein. The observed structures range from an anchored protein shell around a lipid vesicle,[10] to disintegrated vesicles,[11] tubular micelles,[12–15] small lipid-protein nanoparticles[16] and nanodiscs.[17,18] Oligomeric assemblies of $\alpha$-synuclein have also been found to permeabilize lipid bilayers.[19–21] The binding affinity of $\alpha$-synuclein to lipid bilayers is to a large extent determined by the negative charge density of the bilayer,[18] because the positively charged N-terminal region of the protein drives the interaction.[6] At high lipid-to-protein ratios, amyloid fibril formation is often suppressed, because the equilibrium is shifted towards the membrane-bound state, whereas at an excess of protein, amyloid fibril formation can be very significantly accelerated by the presence of model membranes.[22] It is becoming increasingly clear that in such cases, lipid molecules can become incorporated into the growing amyloid fibrils.[23,24] The first atomic resolution structure of $\alpha$-synuclein amyloid fibrils formed in the presence of lipids confirmed the incorporation of lipid molecules into the fibrils.[25] Despite the significant body of work that has been dedicated to $\alpha$-synuclein-lipid interactions, the mechanistic steps of how interactions with lipids enable and accelerate amyloid fibril formation are still unresolved.

In recent years, we have developed a uniquely quantitative $\alpha$-synuclein-lipid system that is particularly suitable to address these questions.[22,24,26] Using a combination of kinetic analysis and systematic variation of the concentrations of lipids and monomeric protein in kinetic experiments of $\alpha$-synuclein aggregation, it has been possible to quantitatively determine the contribution of negatively charged model membranes on $\alpha$-synuclein aggregation. The role of the model membranes consists almost exclusively in an increase of the rate of primary nucleation by three orders of magnitude or more,[22] compared to the aggregation under otherwise identical conditions but in the absence of lipids. This effect is only observed in the presence of an excess of protein, such as to completely saturate the high affinity ($K_D < 1$

$\mu$M) model membranes, while still having unbound protein free in solution. In these experiments, it was observed that the aggregation reaction stops before all the soluble protein is consumed. The thermodynamic stability of amyloid fibrils with respect to the soluble state is high[27,28] and the equilibrium solubility of $\alpha$-synuclein is often less than 1 $\mu$M.[29] However, under quiescent conditions at moderate temperatures in the presence of negatively charged model membranes, the proportion of protein converted into amyloid fibrils is proportional to the initial quantity of added lipid molecules and in some cases several tens of $\mu$M of soluble protein can be present in the plateau phase. This finding suggests that both nucleation and growth of fibrils eventually cease and even that the lipids behave as a reagent that is consumed, rather than a mere catalyst on the surface of which the aggregation processes occur. Our recent solid state nuclear magnetic resonance investigation of the same $\alpha$-synuclein-lipid system indeed showed that the dynamics of the lipid molecules are affected by the aggregation process, supporting the formation of lipid-protein co-aggregates.[24]

In the present study, we aim to elucidate the role of lipids in the interaction with $\alpha$-synuclein, which ultimately leads to the formation of kinetically trapped fibrils.[22] In particular, our goal is to study the structural changes that the model membranes undergo upon the binding of $\alpha$-synuclein and the subsequent protein-lipid co-assembly. The expected changes in lipid arrangement and structure extend down to the nm scale and are hence not easily accessible to direct imaging techniques. Here, we report on the results of a detailed study of different lipid:$\alpha$-synuclein systems by small-angle X-ray and neutron scattering (SAXS and SANS, respectively) techniques that allow the structural characterisation of the system on the relevant length scales. We employ a powerful combination of SANS and SAXS methods. SANS allows the study of lipid and protein components separately through contrast-variation techniques whereas SAXS enables a kinetic investigation down to millisecond time scales through stopped-flow SAXS[30] as well as a larger $q$-range in a single setting. Contrast-variation techniques exploit that $^1$H and $^2$H (D) have very different scattering lengths for neutrons. By adjusting the $H_2O$ to $D_2O$ ratio of the aqueous solvent, individual types or

4

classes of molecules, such as proteins or lipids, can be matched out, i.e. selectively rendered invisible.[31] Our DLS and SAXS/SANS measurements combined with data modelling show that the binding of $\alpha$-synuclein to 1,2-dilauroyl-sn-glycero-3-phospho-L-serine (DLPS) or 1,2-dimyristoyl-sn-glycero-3-phospho-L-serine (DMPS) model membranes leads to a fast (less than 1 ms) and an almost fully reversible break-up into small structures, including nanodiscs and rods. Moreover, our contrast-matching SANS measurements indicate that this fast reversible break-up of the model membranes upon $\alpha$-synuclein binding is followed by lipid-protein co-aggregation. The SANS signature of the lipid is compatible with a systematic incorporation of lipid molecules into the protein fibrils. These results shed light onto the structural details of the lipid-protein structures formed by $\alpha$-synuclein and are relevant to understand the cytotoxicity of the protein via membrane damage and the formation of pathological lipid-protein assemblies *in vivo*.[32]

## Results

### The binding of $\alpha$-synuclein to model membranes leads to a structural re-arrangement into small discs within ms

We first used small angle X-ray scattering to investigate the structural details of the interaction between model membranes, made from either DLPS or DMPS by extrusion, and $\alpha$-synuclein. In particular, we mixed two solutions containing the model membranes and $\alpha$-synuclein, respectively, at temperatures above the melting points of the lipids (i.e. 30°C and 45°C for DLPS and DMPS, respectively) (Figure 1 A,B). The X-ray scattering signatures of DLPS and DMPS model membranes change profoundly upon binding of $\alpha$-synuclein. In particular, we note the loss of the oscillations around $q = 0.02$ $^{-1}$ as well as a significant decrease in the forward scattering intensity for the DLPS sample (Figure 1 A: orange to blue). For both DLPS and DMPS there is a shift of the first minimum in the scattering

5

profile to higher $q$-values upon the binding of $\alpha$-synuclein (Figure 1 A: orange to blue/ B: red to blue). These results suggest that protein is not simply binding to the existing model membranes, but that there are significant structural re-arrangements in terms of average shape and size upon binding of $\alpha$-synuclein.

The apparent structural re-arrangement of the initial lipid structures can be investigated
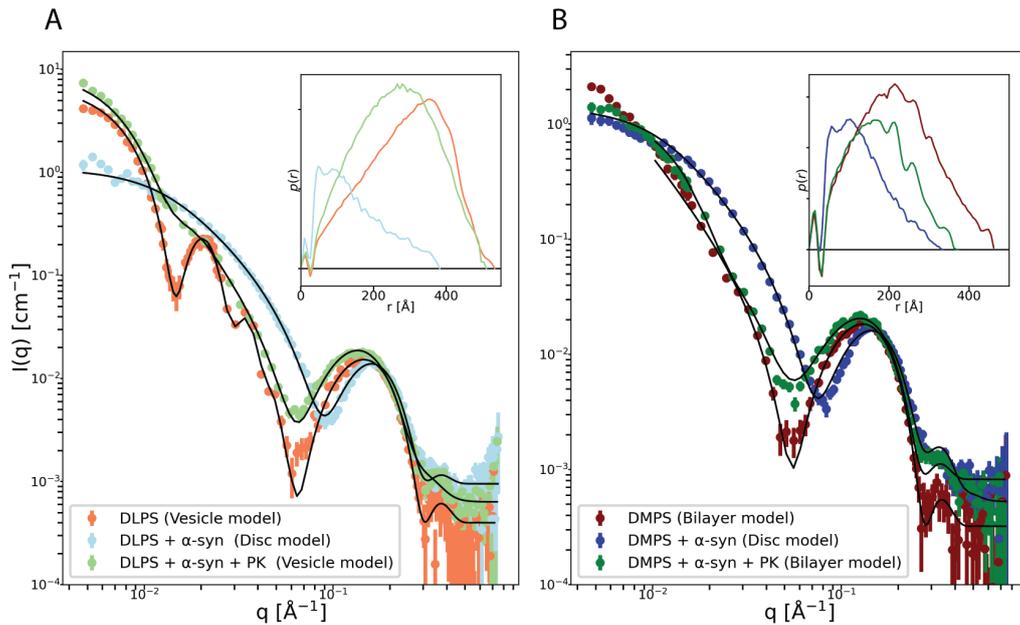


Figure 1: Structural characterisation of the system of $\alpha$-synuclein:model membranes using SAXS. X-ray scattering profiles of 8 mM (A) DLPS and (B) DMPS model membranes incubated in the absence (orange/red) and the presence of 270 $\mu$M $\alpha$-synuclein (blue) as well as after treatment with proteinase K (PK) (green). Model fits to the data are shown in black. The pure lipids were fitted to a vesicle and a bilayer model for DLPS and DMPS (q > 0.01 Å$^{-1}$ (see text)), respectively, whereas the $\alpha$-synuclein:membrane systems were fitted to a disc model. Each PK treatment corresponds to incubation with 15.7 $\mu$M PK for 1 h at 37°C. The data were measured in phosphate buffer pH 6.5 at 30°C (A, DLPS) or 45°C (B, DMPS). The inserts show the p(r)-distribution profiles generated from the SAXS data.

by generating pair-distance distribution profiles[33]. The almost-triangular shape of the $p(r)$-distribution profile from the pure DLPS sample (Figure 1 A inset: orange) is typical of vesicles[34]. The initial fluctuation close to $r = 40$Å represents structural properties of the lipid bilayer and arises due to the presence of lipid headgroups and alkyl chains, which have positive and negative excess X-ray scattering lengths respectively. The $p(r)$-distribution pro-

file from pure DMPS (Figure 1 B inset: red) seems to represent a planar structure which is too large to be fully captured by our $q$-range and we were unable to fit the full SAXS profile to a purely vesicular model. A possible explanation for this behaviour could be that the phase transition temperature of DMPS (i.e. 41°C) is close to room temperature, which is likely to render its behaviour very sensitive to the exact solution conditions at which it was prepared. In the presence of $\alpha$-synuclein, however, the $p(r)$ distribution profiles for both types of lipid (Figure 1 A and B inset: blue) are skewed / shifted to the left with a broad cap and a tail at larger distances, which is indicative of smaller disc-like particles[35,36].

To gain deeper insights into the structural transformations apparent in these data we continue the analysis with a direct model-based approach. Our analytical models represent the lipid and lipid-$\alpha$-synuclein structures as simple geometrical objects using form factors which can be found in the literature[37]. Key information about the overall size and shape of the structures can be refined from the SAS data. The models were implemented in *WillItFit*[38] (see detailed descriptions of the models used in the SI).

Importantly, to constrain the solution-space of the models and ensure the obtained solution is physically meaningful, we incorporate chemical and biophysical information. The scattering lengths and molecular volumes assigned to each part of the model are preset according to which component(s) of the sample they represent (see supplementary table S1).

Additionally, the models are calculated on absolute scale by exploiting the lipid concentration of each sample to give even more robust model calculations. In SAXS, fine tuning the molecular volumes of the various components can correct for small errors on the absolute scale.

A polydisperse three-shell vesicle model (Figure S6), where the three shells represent the outer lipid headgroups, tailgroups, and inner headgroups respectively, provides excellent fits to the SAXS data from pure DLPS (Figure 1 A) while maintaining reasonable structural parameters, listed in Table 1. Size variation is taken into account by assuming a Gaussian distribution of sizes and fitting the mean radius as well as the sigma of the standard devia-

tion of radii. The lipid volume $V_{\text{lipid}}$ stays close to the starting value previously determined in the literature (912 $\text{Å}^{3\ 39}$). The area per lipid headgroup ($A_{\text{head}}$) can be deduced from the model fit results as $V_{\text{lipid}}$ / $t_{\text{lipid}}$ with the thickness of one lipid leaflet $t_{\text{lipid}}$ and lies around 55 $\text{Å}^2$, which is in close agreement with the reports by Szekely *et al.* who found $A_{\text{head}}$ to be 60.5 $\text{Å}^2$ for DLPS in the fluid phase. The obtained mean radius of the DLPS vesicles of 210 Å is reasonable with respect to the extrusion preparation, and shows a small polydispersity, characterized via a relative standard deviation $\sigma_{\text{Radius}}$ of 10%. The polydisperse three-shell vesicle model (Figure S6) did not provide good fits to the low-$q$ SAXS data from pure DMPS (Figure 1 A) as expected given the fact that the $p(r)$-distribution for pure DMPS (Figure 1 A inset: red) lacks the typical liposome shape. To investigate properties of the bilayer we use a planar bilayer model to fit the data for $q$-values above 0.01 $\text{Å}^{-1}$, which corresponds to a length scale of 10 nm. The obtained molecular volume is within 10 % of the previously-determined 978 $\text{Å}^3$ for DMPS[40]. We find DMPS $A_{\text{head}} \sim 57\text{Å}^2$, which we compare to the estimated value of 55.6 $\text{Å}^2$ for DMPS in the fluid phase[26].

Table 1: Structural parameters refined from the SAXS data of DLPS model membranes shown in Figure 1 A using the three-shell vesicle model. The model was calculated on absolute scale. *Parameter was fixed during refinement.
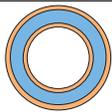
| Parameter | DLPS |
|---|---|
| Scale | 1* |
| Radius [Å] | 211 ± 5 |
| $\sigma_{\text{Radius}}$ | 0.11 ± 0.002 |
| $V_{\text{lipid}}$ [$\text{Å}^3$] | 958 ± 3 |
| $t_{\text{lipid}}$ [Å] | 17.4 ± 0.4 |
| Roughness [Å] | 4.23 ± 0.34 |
| Background [$\text{cm}^{-1}$] | 0.004 ± 0.0003 |
| $\chi^2$ | 2.47 |

Table 2: Structural parameters refined from the SAXS data of DMPS model membranes shown in Figure 1 B using the planar bilayer model. The fit was restricted to q > 0.01 $\text{Å}^{-1}$ where intensity arising from bilayer features dominates.
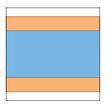
| Parameter | DMPS |
|---|---|
| $V_{\text{lipid}}$ [$\text{Å}^3$] | 1060 ± 3 |
| $t_{\text{lipid}}$ [Å] | 18.5 ± 0.7 |
| $\sigma_{\text{t}}$ | 0.10 ± 0.10 |
| Roughness [Å] | 4.23 ± 0.67 |
| Background [$\text{cm}^{-1}$] | 0.0002 ± 0.0003 |
| $\chi^2$ | 4.88 |

8

Since the $p(r)$-distributions of both the DLPS-$\alpha$-synuclein and the DMPS-$\alpha$-synuclein mixtures hold a close resemblance to that of a disc-like structure we propose a scheme where the binding of $\alpha$-synuclein leads to the deformation of pure model membranes into discoidal bilayer patches, which are stabilized by a belt of $\alpha$-synuclein, in a similar manner as the one that amphipathic membrane scaffold proteins can stabilize lipid nanodiscs[16,41,42] (Figure S9). Our analytical model builds upon our previously developed model for nanodiscs[43,44] where the lipid bilayer is described as a collection of discs. Each disc represents the lipid headgroups, tailgroups and protein belts, respectively. Here we revise the volumes and scattering lengths assigned to each disc, in order to additionally allow the amphipathic N-terminal and hydrophobic regions of $\alpha$-synuclein to be accounted for in the lipid bilayer. The hydrophilic C-terminal tails of $\alpha$-synuclein, which are thought to remain disordered, protrude from the outside of the disc as Gaussian random coils, following the same modelling strategy as presented by Pedersen and Gerstenberg[45,46] (see more details in the SI 'Disc with Gaussian random coils' model). The 'Disc with Gaussian random coils' model provides good fits to the data for both DLPS and DMPS in the presence of $\alpha$-synuclein (Figure 1). The model captures the low-$q$ slopes of the scattering profiles accurately indicating that the overall shape and dimensions are well-described (see fit parameters in Table 3). We assume all of the protein is bound to the lipids since the DLPS and DMPS bilayers are in the fluid phase under the experimental conditions used here, and hence take a molar ratio of 30:1 lipid:protein (L/P) into account[26] when calculating scattering length densities of the discoidal bilayer. Throughout this study, our modelling scheme has difficulties to accurately capture the dimensions of the protein. Therefore the radius of gyration $R_{g\,\alpha-syn}$ of the random coils was fixed to 18 Å, as calculated by Kohn's power law relationship for 44 amino acids,[47] the length of the C-terminal tail of $\alpha$-synuclein that does not interact with the lipid bilayer[6](see SI for more details). Similarly we found the width of the protein belt, $w_{belt}$, did not have much influence on the shape of the model profile and was weakly constrained. Therefore $w_{belt}$ was fixed to the diameter of an $\alpha$-helix, 12Å, under the plausible assumption

Table 3: Structural parameters refined from the SAXS data of DLPS- and DMPS-$\alpha$-synuclein mixtures shown in Figure 1 using the 'Disc with Gaussian random coil' model. *Parameter was fixed during refinement.
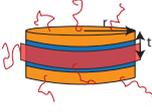
| Parameters | DLPS + $\alpha$-syn | DMPS + $\alpha$-syn |
|---|---|---|
| Scale | 1* | 1* |
| L/P | 30* | 30* |
| $w_{\text{Belt}}$ [Å] | 12* | 12* |
| $R_{\text{g, }\alpha\text{-syn}}$ [Å] | 18* | 18* |
| Radius [Å] | 148 ± 73 | 137 ± 48 |
| $\sigma_{\text{Radius}}$ | 0.28 ± 0.01 | 0.26 ± 0.02 |
| $V_{\text{lipid}}$ [Å$^3$] | 863 ± 43 | 955 ± 30 |
| $V_{\alpha\text{-syn}}$ [Å$^3$] | 21300 ± 1510 | 21300 ± 900 |
| $t_{\text{lipid}}$ [Å] | 17.2 ± 154 | 19.3 ± 1.1 |
| Roughness [Å] | 5.30 ± 1.01 | 4.53 ± 1.07 |
| Background [cm$^{-1}$] | 0.001 ± 0.0003 | 0.0009 ± 0.0003 |
| $\chi^2$ | 2.54 | 3.55 |

that $\alpha$-synuclein forms a helical structure upon binding.[6]

For both the DLPS- and DMPS-$\alpha$-synuclein mixtures, we refine populations of polydisperse discs with mean radii of around 140 Å. The thickness of the bilayer was found to be identical within error to the dimensions refined from the pure lipid data in Table 1. The refined molecular volumes for both DLPS and DMPS are 90% of those found in Table 1 which could potentially be explained by the difference in packing of lipids in vesicles and discs. The refined volume for $\alpha$-synuclein is 21300 Å$^3$ in both cases, which is 20% larger than the volume estimated using the average mass density of proteins as 1.35 cm$^2$ g$^{-1}$[48] (Table S1). We note that we attempted to model these data as a range of different shapes and structures, including vesicles, cylinders and ribbons, but we were not able to match the scattering profile at low-$q$ successfully with such alternative models. We also tried to model the data with variations of the disc model, including bicelles where the protein belt and random coils were not present, but found that our 'Disc with Gaussian random coils' captures all the features of the experimental profile most accurately.

All together these SAXS measurements analysed using a direct model-based approach show that the binding of $\alpha$-synuclein to DMPS and DLPS model membranes leads to the break-up of the vesicles and large bilayer structures into small discs.

## The structural re-arrangement of DLPS and DMPS model membranes into discs upon $\alpha$-synuclein binding occurs on a millisecond timescale

We used stopped-flow SAXS (SF-SAXS) to investigate the time scale of the observed structural re-arrangement of DLPS and DMPS vesicles and bilayer structures into discs upon binding of $\alpha$-synuclein. Figure 2 A,B show the X-ray scattering intensity during the first milliseconds after mixing DLPS or DMPS with buffer (0 ms) or $\alpha$-synuclein (1-21 ms) at 30°C (Figure 2 A,B and S1).

The SAXS profile from DLPS at 0 ms can be modeled as polydisperse vesicles as described in the previous section. For DMPS the data once again indicate very large planar structures and so we used the planar bilayer model to focus on the scattering contribution from the bilayer only. The DLPS-$\alpha$-synuclein and DMPS-$\alpha$-synuclein mixtures at 1 and 21 ms are well-described by the 'Disc with Gaussian random coils' model (Tables 4 and 6). The stopped-flow data could not be converted to absolute scale and so a free scale parameter had to be employed during the model refinement. For consistency, we only fit the scale for the first DLPS and DMPS SAXS profiles and kept the scale fixed in subsequent refinements. Furthermore we observed that the molecular volumes for DLPS and DMPS did not need to be adjusted from their pre-estimated values and hence they were kept fixed throughout the data series to avoid overfitting. Despite the different initial lipid morphologies, the data indicate that the structural re-arrangement of DLPS and DMPS model membranes into discs upon binding of $\alpha$-synuclein takes place very fast within less than 1 ms, the dead-time of the stopped-flow setup,[49] thereby preventing a more precise analysis of the kinetics of the interaction.

In an attempt to rationalize the fast binding kinetics, we can model the binding of monomeric $\alpha$-synuclein to the surface of DLPS lipid vesicles as a process of ligand binding to a fully absorbing sphere, i.e. every diffusive encounter yields a binding event. We can calculate

11

the diffusion-limited rate at steady state, according to $\frac{dn}{dt} = 4\pi DRc$,[50] where n is the number of bound molecules, D is the diffusion coefficient of the protein ($8.2 \cdot 10^{-11}$ $m^2 s^{-1}$ for $\alpha$-synuclein[51]), R is the radius of the vesicle and c is $\alpha$-synuclein concentration. At c = 70 $\mu$M, we obtain $\frac{dn}{dt} \approx 6.5 \cdot 10^5$ $s^{-1}$. A vesicle with a radius of 15 nm contains of the order of 10,000 lipid molecules, if we consider a surface area per lipid of 55.6 $\mathring{A}^2$, which translates into approximately 300 molecules of $\alpha$-synuclein per vesicle at full saturation.[26] At a constant rate of binding, the vesicle would therefore be saturated after 0.5 ms. In reality the rate of binding will slow down, as more and more binding sites are occupied. However, this simple estimate shows that the observed rate of interaction between $\alpha$-synuclein and DLPS vesicles at 30°C is probably close to the diffusion limit, despite the fact that the lipid vesicles and the $\alpha$-synuclein both carry a net negative charge, which has the potential to reduce the maximal possible rate of binding due to electrostatic repulsion. However, it is possible that the binding reaction is even electrostatically enhanced[52] once the $\alpha$-synuclein molecule is close enough to the negatively charged vesicle, because it interacts with the lipids through its N-terminal region, which carries a net positive charge.

Table 4: Structural parameters refined from the SAXS data shown in Figure 2 A (blue) using the three-shell vesicle model. *Parameter was fixed during refinement. We note that it was not possible to convert this data set to absolute scaling units and the model therefore requires a free Scale parameter far from 1.

| | Parameters | DLPS (0ms) |
|---|---|---|
| | Scale | $0.056 \pm 0.008$ |
| | Radius [$\mathring{A}$] | $259 \pm 16$ |
| | $\sigma_{Radius}$ | $0.17 \pm 0.06$ |
| | $v_{lipid}$ [$\mathring{A}^3$] | 912* |
| | $t_{lipid}$ [$\mathring{A}$] | $20.2 \pm 1.8$ |
| | Background [$cm^{-1}$] | $0.0003 \pm 8e\text{-}5$ |
| | $\chi^2$ | 2.70 |
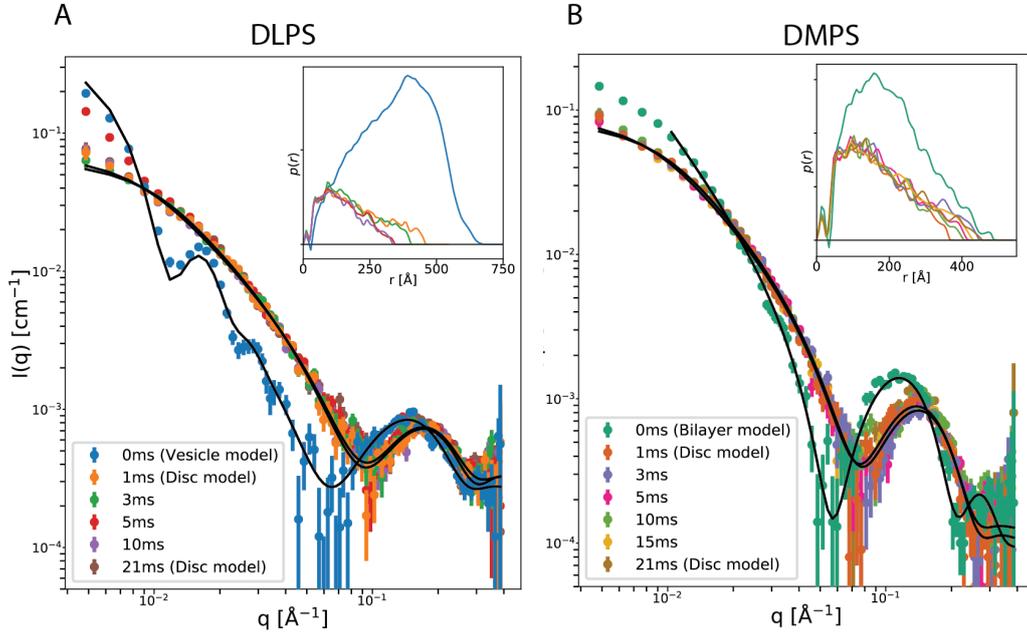
Figure 2: Characterisation of the kinetics and structural details of $\alpha$-synuclein:membrane interactions using stopped-flow SAXS. X-ray scattering function of a solution containing either (A) DLPS or (B) DMPS model membranes of 2 mM lipids and 70 $\mu$M $\alpha$-synuclein, measured at different time points after mixing. The solution conditions were 20 mM sodium phosphate buffer at pH 6.5 and 30°C. Insets show the $p(r)$ distribution profiles generated from the data.

Table 5: Structural parameters refined from the SAXS data of DMPS model membranes shown in Figure 2 B (green) using the bilayer model. The fit was restricted to q > 0.01 Å$^{-1}$ where intensity arising from bilayer features dominates. *Parameter was fixed during refinement

| | Parameter | DMPS (0ms) |
|---|---|---|
| | $V_{\text{lipid}}$ [Å$^3$] | 978* |
| | $t_{\text{lipid}}$ [Å] | 24.2 ± 1.5 |
| | $\sigma_{\text{t}}$ | 0.11 ± 0.1 |
| | Background [cm$^{-1}$] | 8e-5 ± 8e-5 |
| | $\chi^2$ | 1.55 |

13

Table 6: Structural parameters refined from the SAXS data shown in Figure 2 using the Disc with Gaussian random coil model. *Parameters fixed during refinement. For consistency the Scale is fixed at the value obtained in Table 4 For DLPS and Table 4 for DMPS.

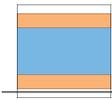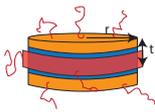| Parameters | DLPS + $\alpha$-syn (1ms) | DLPS + $\alpha$-syn (21ms) | DMPS + $\alpha$-syn (1ms) | DMPS + $\alpha$-syn (21ms) |
|---|---|---|---|---|
| Scale | 0.056* | 0.056* | 0.056* | 0.056* |
| $L/P$ | 30* | 30* | 30* | 30* |
| $w_{belt}$ [Å] | 12* | 12* | 12* | 12* |
| $R_{g\ \alpha-syn}$ [Å] | 18* | 18* | 18* | 18* |
| Radius [Å] | 190 $\pm$ 118 | 170 $\pm$ 76 | 194 $\pm$ 103 | 169 $\pm$ 75 |
| $\sigma_{Radius}$ | 0.30 $\pm$ 0.02 | 0.29 $\pm$ 0.02 | 0.29 $\pm$ 0.02 | 0.28 $\pm$ 0.02 |
| $v_{lipid}$ [Å$^3$] | 912* | 912* | 978* | 978* |
| $v_{\alpha-syn}$ [Å$^3$] | 19000 $\pm$ 640 | 18800 $\pm$ 470 | 19000 $\pm$ 500 | 18800 $\pm$ 480 |
| $t_{lipid}$ [Å] | 16.6 $\pm$ 1.9 | 20.6 $\pm$ 1.83 | 21.2 $\pm$ 1.9 | 20.9 $\pm$ 1.8 |
| Roughness [Å] | 4.30 $\pm$ 2.24 | 4.73 $\pm$ 2.36 | 4.37 $\pm$ 2.65 | 3.89 $\pm$ 2.95 |
| Background [cm$^{-1}$] | 0.0003 $\pm$ 0.0002 | 0.0001 $\pm$ 0.0001 | 0.0001 $\pm$ 0.0001 | 0.0001 $\pm$ 0.0001 |
| $\chi^2$ | 1.18 | 1.46 | 0.85 | 1.42 |

# The structural re-arrangement of DLPS and DMPS vesicles into discs upon $\alpha$-synuclein binding is reversible

Next, we investigated whether the structural changes that the model membranes undergo upon $\alpha$-synuclein binding are reversible. To this end, it was necessary to analyze the structures of the model membranes after removing the protein. The high affinity between $\alpha$-synuclein and DLPS or DMPS membranes renders it difficult to completely remove the protein on a short time scale. We therefore designed an experimental protocol, whereby $\alpha$-synuclein was digested with proteinase K (PK). To test the efficacy of this treatment, we performed differential scanning calorimetry (DSC) measurements of the $\alpha$-synuclein-lipid mixture before and after incubation with PK while exploiting the known effect of $\alpha$-synuclein binding on the melting temperature (T$_m$) of the DMPS bilayer, i.e. T$_{m,\ DMPS} \sim 40°$C whereas T$_{m,\ DMPS:\alpha\text{-synuclein}} \sim 25°$C.[26] The T$_m$ of the $\alpha$-synuclein:DMPS system was found to increase from $\sim 25°$C before PK treatment to $\sim 41°$C after PK-treatment (Figure S2 A), suggesting that the incubation of $\alpha$-synunclein-DMPS mixtures with PK leads to the virtually complete removal of the protein from the surface of the membranes. The mechanism of this displacement is likely to be the digestion of the free monomeric $\alpha$-synuclein, which is highly susceptible to PK digestion due to its intrinsically disordered nature, and the resulting displacement of the binding equilibrium caused by the decreased lipid-affinity of the proteolytic

fragments. Circular dichroism (CD) spectra of $\alpha$-synuclein-DMPS mixtures before and after addition of PK confirmed that the secondary structure of the protein changed from mainly $\alpha$-helical to random-coil upon PK treatment (Figure S2 B), reflecting again the digestion of $\alpha$-synuclein into short, unstructured peptides.
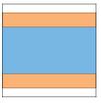
Having established a method to remove the bound $\alpha$-synuclein *in situ*, we then acquired the SAXS data of the PK treated protein-lipid mixtures for DLPS and DMPS and found that they bear a close resemblance with those of the lipid structures before protein addition (Figure 1 A,B). Indeed, the $p(r)$-distribution profiles of the PK treated $\alpha$-synuclein:DLPS mixtures appear to show restored vesicles with dimensions much larger than the disc-like particles, though the $p(r)$-distribution and SAXS profile suggests that the recovered population has increased heterogeneity compared to the situation before protein addition [34] (Figure 1 A: green). The corresponding SAXS data are well-described by the polydisperse three-shell vesicle and bilayer model, for DLPS and DMPS, respectively, with structural parameters similar to those of intact lipid structures before protein addition (Table 7 and 1). In order to achieve a satisfying fit, the DLPS vesicles formed after PK treatment of the mixtures require a polydisperse distribution of thicknesses of the lipid shell as well as the radius, whereas the initial intact vesicles only required a distribution of radii. For the PK treated $\alpha$-synuclein:DMPS mixtures, the $p(r)$-distribution profile has a similar shape to the one from pure DMPS, but the particles do not appear to revert to their original maximum dimensions (Figure 1 B : green). The structural parameters of the bilayer, on the other hand, are close to those before protein addition (Table 8 and 1). The thickness of the bilayer is identical but appears more polydisperse. The molecular volume per lipid appears very slightly lower which could suggest some minor residual disturbance in the lipid packing.

The treatment of the vesicles with PK does not in itself affect their SAXS profiles and $p(r)$ distribution profiles (Figure S3). When we refine the bilayer model results from the data collected from DMPS with PK, we obtain a very similar description of the bilayer as from the pure DMPS structures (Table 8 and 1).Moreover, the SAXS profile of DMPS model

Table 7: Structural parameters refined from the SAXS data of PK treated protein-DLPS mixtures in Figure 1 A using the Three-shell vesicle model. The model was calculated on absolute scale. *Parameter fixed during refinement.

| | Parameters | DLPS + $\alpha$-synuclein + PK |
|---|---|---|
| | Scale | 1* |
| | Radius [Å] | 197 ± 15 |
| | $\sigma_{\text{Radius}}$ | 0.30 ± 0.04 |
| | $v_{\text{lipid}}$ [Å$^3$] | 941 ± 1 |
| | $t_{\text{lipid}}$ [Å] | 17.7 ± 0.5 |
| | $\sigma_t$ | 0.19 ± 0.05 |
| | Roughness [Å] | 3.66 ± 0.36 |
| | Background [cm$^{-1}$] | 0.0006 ± 0.0004 |
| | $\chi^2$ | 4.32 |

Table 8: Structural parameters refined from the SAXS data of PK treated protein-DMPS mixtures (Figure 1 B), and mixtures of DMPS and PK (Figure S3) using the planar model. The fits are restricted to q > 0.01 Å$^{-1}$ where intensity arising from bilayer features dominates. *Parameter was fixed during refinement.

| | Parameter | DMPS + $\alpha$-synuclein + PK | DMPS + PK |
|---|---|---|---|
| | $V_{\text{lipid}}$ [Å$^3$] | 1050 ± 3 | 1060 ± 2 |
| | $t_{\text{lipid}}$ [Å] | 18.2 ± 0.7 | 20.0 ± 0.5 |
| | $\sigma_t$ | 0.21 ± 0.06 | 0.07 ± 0.07 |
| | Roughness [Å] | 3.72 ± 0.65 | 4.15 ± 0.48 |
| | Background [cm$^{-1}$] | 0.0006 ± 0.0003 | 0.0001 ± 0.0003 |
| | $\chi^2$ | 7.91 | 2.13 |

membranes in the presence of PK-pre-digested $\alpha$-synuclein, is characterized by an increased underlying intensity at mid- and high-$q$. The position of the minimum and bump (at $q \sim$ 0.04 to 0.1 Å$^{-1}$) is unchanged compared with the data of pure DMPS (Figure S3: purple and red). This observation is consistent with conserved lipid structures with $\alpha$-synuclein in a disordered conformation both in solution and decorating the lipids. From the $p(r)$-distributions it is evident that the overall planar structure is hardly affected by this short peptide-membrane interaction.

Taken together, these results indicate that the structural changes to small discs induced by the binding of $\alpha$-synuclein to DLPS and DMPS model membranes can be reverted by removal of the protein. For DLPS the reformed vesicles have a similar size to that of intact vesicles before $\alpha$-synuclein binding but with a higher polydispersity compared to the pure vesicle samples. For DMPS the overall structure is recovered but the structures do not grow back to their original dimensions.

We further investigated the reversibility of the binding of $\alpha$-synuclein to DMPS model mem-

branes using temperature-ramp experiments. $\alpha$-synuclein binding to DMPS model membranes was found to be more favourable when the membrane was in the fluid phase than in the gel phase,[26] with the stoichiometry of lipid-binding differing by approximately one order of magnitude between the two phase states of the lipid. Therefore, when DMPS discs saturated with protein are cooled from above to below the melting temperature, $\sim$90% of the protein can be expected to detach and the DMPS lipids to re-organise themselves into their initial structures, if the interaction is reversible.

We measured SAXS data of DMPS model membranes in the absence and the presence of $\alpha$-synuclein at temperatures ranging from 14 °C to 48 °C (Figure 3 and S4), temperatures at which the DMPS bilayer is in the gel phase and fluid phase, respectively, irrespective if the protein is bound or not.[26]

The increase of temperature from 14 °C to 30 °C led to a decrease of the intensity at low-$q$, a change of the slope in the initial decay of the profiles and a shift of the position of the first intensity minimum to higher-$q$ for the SAXS profile of $\alpha$-synuclein:DMPS mixture (Figure 3A). If the $\alpha$-synuclein:DMPS mixtures conserved the initial DMPS structure, the forward scattering intensity would be expected to steadily increase with the amount of $\alpha$-synulcein bound. However, the reduction in forward scattering points towards the spontaneous break-up of the DMPS large planar structures into smaller particles that contain a lower number of lipids. Such changes were not observed for the pure DMPS model membranes when the temperature was increased from 14°C to 49°C (Figure S4). Furthermore, the low-$q$ range of the SAXS data from pure DMPS in Figure S4 is steeper than what is observed in Figure 3 even at the lowest temperatures, indicating that $\alpha$-synuclein is prone to disrupting the DMPS structures even under conditions with low binding stoichiometry. The scattering profiles of the first five temperatures in Figure 3 (14 °C to 18 °C) of $\alpha$-synuclein:DMPS mixture lie approximately on top of each other. The corresponding $p(r)$-distribution profiles indicate very large particles with a $D_{max}$ of at least 600 Å but which are not characteristic of any homogeneous, well-defined lipid vesicle population. Rather the $p(r)$-distribution

17
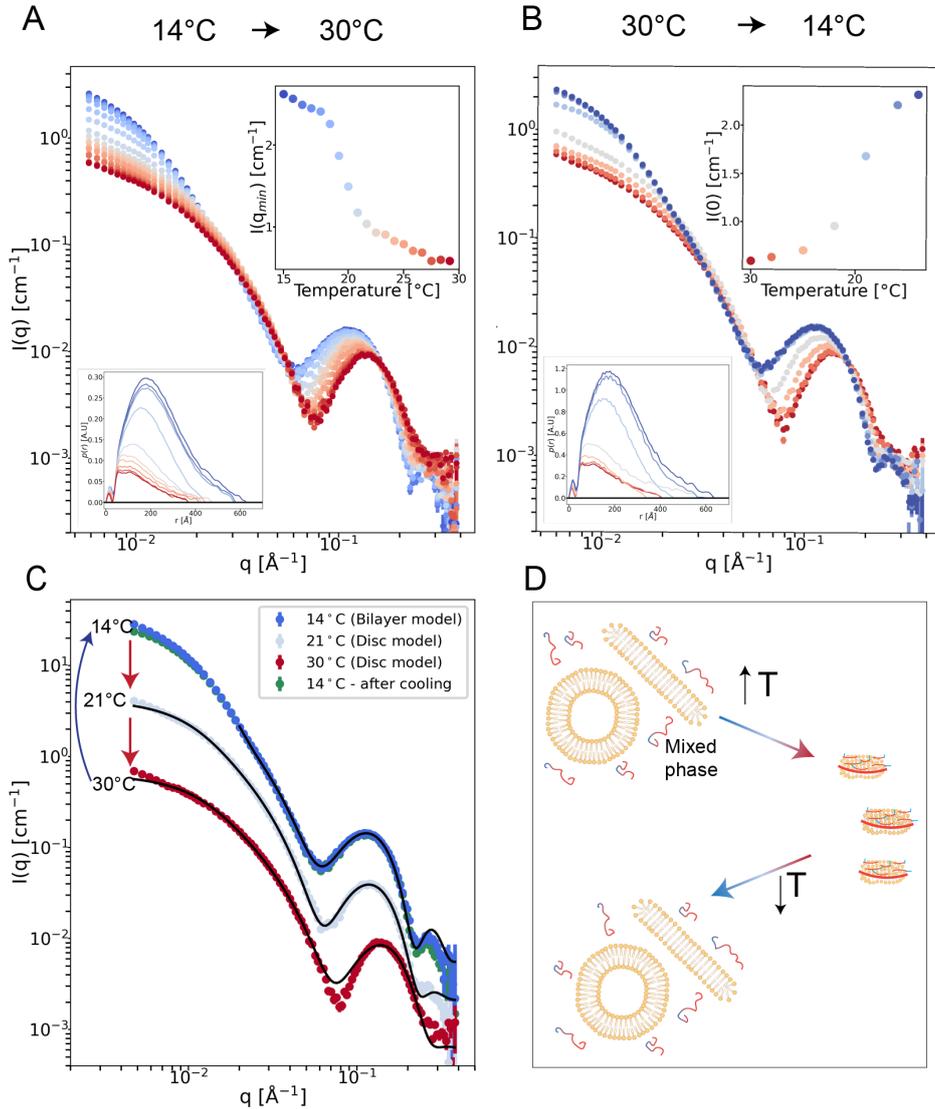
Figure 3: Structural characterisation of the $\alpha$-synuclein:DMPS system at increasing and decreasing temperature using SAXS. X-ray scattering and normalized $p(r)$-distribution profiles of (A) 3 mM DMPS in the presence of 100 $\mu$M $\alpha$-synuclein heated from 14 °C to 30 °C and (B) 3 mM DMPS in the presence of 100 $\mu$M $\alpha$-synuclein cooled from 30 °C to 14 °C. Insets: the first data point (q = 0.0048 Å$^{-1}$) of each SAXS profile as a function of temperature in order to highlight the transition temperature of the system at $\sim$ 20 °C. (C) Structural modelling investigating the reversible structural reorganisation of DMPS:$\alpha$-synuclein from large planar structures to discs with increasing temperature and hence binding of $\alpha$-synuclein. For ease of viewing, the bottom scattering profile (30°C) remains on absolute scale but the middle profile (21° C) is scaled by 2 and the two topmost profiles 14°C are scaled by 4. The refined structural parameters are listed in Tables 9 and 10. (D) Schematic representation of the system.

18

suggests large planar or nearly planar structures[33]. Between 19 °C and 20 °C, however, a clear transition can be observed, corresponding to the melting of $\alpha$-synuclein bound DMPS bilayers. The temperature-dependence of the profiles levels off again around 25°C suggesting the DMPS:$\alpha$-synuclein co-structures may have reached a new (meta-)stable structural state. At these temperatures (T above 25°C), $p(r)$-distribution profiles resemble that of disk-like particles[35,36] with a prominent flat maximum at $r = 50$ to 150 Å and a tail at long distances (Figure 3A). The disc-like particles appear much smaller than the initial aggregates. Figure 3B shows that the structural changes that DMPS model membranes undergo upon binding of $\alpha$-synuclein can be reversed by cooling the system back down to 14 °C. As the temperature was decreased from 30 °C to 14 °C, we observed the transition of $\alpha$-synuclein-bound DMPS bilayer around 19°C and the shape and intensity of the SAXS profile after cooling was restored to that before heating. The $p(r)$-distribution profiles of the DMPS:$\alpha$-synuclein co-structures below the melting temperature for the heating (Figure 3A) and cooling (Figure 3B) measurements are very similar, suggesting that the structural changes of the DMPS model membranes upon $\alpha$-synuclein binding can be reversed upon detachment/removal of the protein induced by cooling. This is highlighted in Figure 3C where both SAXS profiles collected at 14 °C (one before heating and one after cooling) are plotted on top of each other and show remarkably little deviation. We again attempted to refine structural models against data collected at various temperatures (Figure 3C). At 14 °C, the low-$q$ data is again not compatible with a vesicle model. We therefore focus our modeling on the range $q = 0.02$ Å$^{-1}$ and upwards, in order to investigate properties of the bilayer rather than the whole structure. We again use the planar bilayer model, but with an altered scattering length density to reflect the previously determined stoichiometry at this temperature (i.e. 300 DMPS lipids per $\alpha$-synuclein monomer).[26] Additionally, we use the form factor for Gaussian random coils to model the remaining free $\alpha$-synuclein which have a non-negligable contribution to the scattering intensity. We find our description of the bilayer to be in line with the data, refining a molecular volume for DMPS of 987 Å$^3$ which is in close agreement

the pre-estimated value of 978 Å$^3$[40], as well as observing an increase in the thickness of the lipid leaflets compared to DMPS lipid structures without $\alpha$-synuclein. The refined $R_g$ of the unbound $\alpha$-synuclein of 31 Å is in close agreement with Kohn's estimation of 37 Å for 140 residues,[47] and with our SANS data from d-$\alpha$-synuclein alone (Figure S12). We find that the initial state of DMPS is fully restored upon cooling-induced dissociation of $\alpha$-synuclein, despite the fact that the initial state appears to be a heterogeneous mixed lipid phase rather than a well-defined single species.

Table 9: Structural parameters refined from the data in Figure 3C, using the planar bilayer model to represent the DMPS bilayer and a Gaussian random coil model to represent the free $\alpha$-synuclein in solution. The fit was restricted to q > 0.02 Å$^{-1}$ where intensity arising from bilayer features dominates The model was calculated on absolute scale.

| Parameter | 14 °C | 14 °C after cooling |
|---|---|---|
| $V_{\text{lipid}}$ [Å$^3$] | 987 ± 2 | 987 ± 1 |
| $t_{\text{lipid}}$ [Å] | 23.1 ± 0.2 | 23.1 ± 0.2 |
| $R_g$ [Å] | 30.6 ± 3.8 | 31.0 ± 3.4 |
| Roughness [Å] | 5.65 ± 0.27 | 5.65 ± 0.23 |
| Background [cm$^{-1}$] | 0.0005 ± 0.0003 | 0.0003 ± 0.0003 |
| $\chi^2$ | 3.94 | 6.27 |
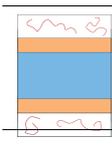
For temperatures above 21°C, i.e. just above the transition, the low-$q$ slope of X-ray scattering profiles of DMPS:$\alpha$-synuclein mixtures becomes more gentle and the 'Disc with Gaussian random coils' model, introduced above and described in the SI, can be employed. We fix the volume of DMPS to 987 Å$^3$ in line with the results reported in Table 9. At 21 °C, the model achieves an excellent fit (Figure 3C). A ratio of lipid to bound-$\alpha$-synuclein of 36 ± 7 is found, corresponding to around 90% of $\alpha$-synuclein being bound which is possibly overestimated for 21 °C. At 30 °C the fit is satisfactory, although it overshoots the minimum at $q$=0.1, possibly suggesting the model is too simple to explain the DMPS:$\alpha$-synuclein mixtures in this high binding regime. At 30 °C we assume that ~100% of $\alpha$-synuclein is bound to the lipid bilayer and therefore the molar ratio is fixed to 30:1 in line with the experimental concentrations. The model fit results show polydisperse discs with an average radius and bilayer thickness of 170 Å and 50 Å, respectively at 21 °C, which are reduced to 142 Å and 38 Å at 30 °C (Table 10).

20

Table 10: Structural parameters refined from the data in Figure 3C using the Disc with Gaussian random coils model. *Parameters fixed during refinement.
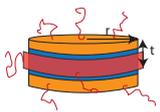


| Parameter | 21°C | 30°C |
|---|---|---|
| Scale | 1* | 1* |
| $L/P$ | $33.2 \pm 7.6$ | 30* |
| $w_{Belt}$ [Å] | 12* | 12* |
| $R_{g,\ \alpha\text{-syn}}$ [Å] | 18* | 18* |
| Radius [Å] | $170 \pm 17$ | $140 \pm 16$ |
| $\sigma_{Radius}$ | $0.27 \pm 0.007$ | $0.28 \pm 0.007$ |
| $v_{lipid}$ [Å$^3$] | 987* | 987* |
| $v_{\alpha\text{-syn}}$ [Å$^3$] | $19600 \pm 200$ | $19600 \pm 100$ |
| $t_{lipid}$ [Å] | $22.7 \pm 0.36$ | $19.0 \pm 0.4$ |
| Roughness [Å] | $6.28 \pm 0.47$ | $5.83 \pm 0.48$ |
| Background [cm$^{-1}$] | $0.0007 \pm 0.0002$ | $0.0006 \pm 0.0002$ |
| $\chi^2$ | 3.68 | 9.94 |

Taken together, these results show that the structural changes from vesicles/large planar structures to smaller discs that, respectively, DLPS and DMPS undergo upon the initial binding of $\alpha$-synuclein can be largely reversed, either by degradation-induced protein displacement (demonstrated for DMPS and DLPS) or by temperature-induced partial displacement of the protein from the membrane (demonstrated for DMPS).

## Rod-like structures can also assemble from mixtures of DLPS or DMPS with $\alpha$-synuclein

In order to obtain further structural information about the lipid and the protein upon binding, we collected SANS data from deuterated $\alpha$-synuclein and DLPS or DMPS model membranes under protein- and lipid-matching conditions. A full discussion is in the supplementary results. Akin to the SAXS data presented above, the SANS data clearly report a structural reorganisation of the initial lipid structures once $\alpha$-synuclein is added. However, upon binding of $\alpha$-synuclein to the lipids we observe a change in morphology to small rod- or ribbon-like particles rather than discs.

We believe that these two types of co-structures, discs and rods, have similar thermodynamic stability[53] (see also Figure 6) and that a very small variation in the sample conditions could move the system from one state to the other. The qualitative difference between the SAXS and SANS data is observed for both DLPS and DMPS, and hence independent

21

of whether the initial state is a well-defined vesicle population or larger bilayer structures. This difference could stem from an isotope effect caused by the deuterium labeling of the $\alpha$-synuclein and using $D_2O$ in the solvent in SANS. Self-assembly of lipids and amphiphillic molecules into specific structures is driven by the hydrophobic effect, i.e. ultimately the solvent water, and therefore the substitution of H to D could have a measurable effect on the equilibrium structures of systems with near-isoenergetic states.

It is well accepted that moving from $H_2O$ to $D_2O$ can shift the cooperative melting temperatures of lipid phases by several degrees Celsius due to stronger hydrophobic interactions[54–56]. While $D_2O$ solvent has been reported to have a stabilising effect on folded proteins[57–59], deuteration of nonexchangeable protons destabilize folded protein[60,61] and can have an effect on their dynamics[61,62]. These phenomena are not widely studied or understood, especially for disordered proteins which may be particularly sensitive to protein-solvent interactions[63] and H-D substitutions. In $H_2O$ with hydrogenated $\alpha$-synuclein, we observe that the binding of $\alpha$-synuclein induces a spontaneous break-down of vesicles or large planar bilayers into small discs and mainly assembles around the outside of the discs to shield the hydrophobic tail groups from the solvent and stabilise the structure. Self-assembly into discs has been observed for many types of surfactants including small-molecule amphiphiles[64], membrane scaffold proteins[41], amphiphilic polymers[65] and specifically $\alpha$-synuclein[18,42]. With the deuterated version of $\alpha$-synuclein in $D_2O$, we speculate that a change in hydrophobicity of the protein side chains changes the interactions with the lipid head and tail groups, such that the protein can coat the length of the rod-like structure and hence stabilise them.

We present Figure 4 as an overview and summary of the results from our SAXS and SANS experiments. Throughout our study we observe that DLPS forms exemplar vesicles (Figures 1A, 2A and S12A. We observe that DMPS, on the other hand, forms large polydisperse planar structures (Figures 1B, 2B, 3, S12B and S13). We believe this could be a mixed phase which includes vesicles, large bicelles and other types of micelles. Our SAXS data show the original lipid structures of both DMPS and DLPS are transformed into smaller

disc-like particles when $\alpha$-synuclein binds, and that this process is reversible (Figures 1, 2 and 3). The SANS data, on the other hand, indicate break down into small rods (Figures 5, S12 and S13).
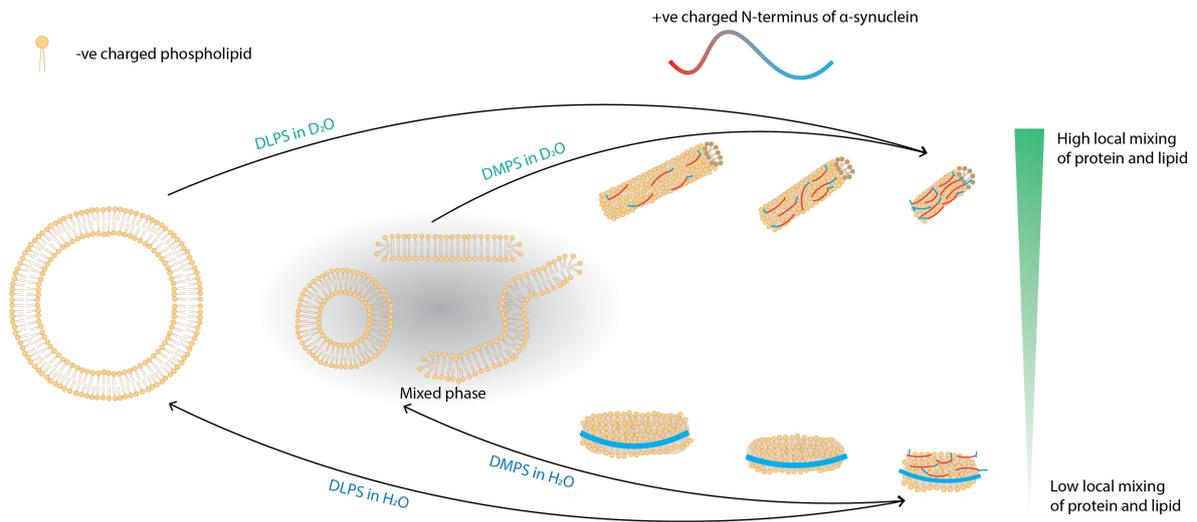


Figure 4: Overview of the different types of structures and structural re-arrangements of DLPS and DMPS in the absence and presence of $\alpha$-synuclein. DLPS re-arranges from vesicles into disks ($H_2O$) and rods ($D_2O$) upon binding of $\alpha$-synuclein. DMPS re-arranges from large bilayer structures into disks ($H_2O$) and rods ($D_2O$) upon binding of $\alpha$-synuclein. The possible origin of the observed isotope effects is discussed in the text.

## $\alpha$-synuclein-lipid structures elongate during the process of amyloid formation

Finally, we used Small Angle Neutron Scattering (SANS) measurements under protein and lipid contrast-matching conditions to study the individual structural contribution of the protein and lipid molecules to the process of amyloid fibril formation. Indeed, this technique allows the acquisition of the neutron scattering intensity of a mixture containing model membranes and $\alpha$-synuclein under various $D_2O$:$H_2O$ contents to match out either the signal of the protein or that of the lipids. Our SANS measurements of deuterated $\alpha$-synuclein (d-$\alpha$-synuclein) and DMPS or DLPS model membranes in $D_2O$ concentrations ranging from 0 to 100% show that the forward scattering of the protein and the lipids are matched out in the

23

presence of 100% and 18% $D_2O$, respectively (Figure S5).

Having established the reversibility of the rapid structural re-arrangement of the model membranes upon $\alpha$-synuclein binding using SAXS (Figures 1 - 3) and SANS (Figures S13, S15) measurements, as well as temperature-dependent dynamic light scattering (DLS) experiments (Figure S16 and S17), we investigated in more detail the subsequent aggregation process that occurs on a much slower time scale of hours to days.[22,26] Due to the better defined vesicular nature of DLPS, as well as better compatibility of the fast DLPS-induced aggregation of $\alpha$-synuclein (compared to DMPS) with the time scale of SANS measurements, we used exclusively DLPS in these kinetic experiments. We prepared several samples under identical conditions, except for the ratios of $D_2O$ to $H_2O$ (100% $D_2O$ (Figure 5A) and 18% $D_2O$ (Figure 5B)) and the absence or presence of the fluorescent dye Thioflavin-T (ThT), which reports on the formation of amyloid fibrils. The sample with ThT allowed us to follow the formation of fibrils in real time by monitoring the increase in fluorescence intensity over time in a plate reader (Figure 5C, black curve). While the binding of $\alpha$-synuclein to DLPS vesicles initially leads to a strong decrease of the scattering function at low $q$ values under protein matching conditions (Figure S12 B), the intensity increases at low $q$ as a consequence of the slow aggregation process (Figure 5A). Between 14 and 20 hours, the scattering profiles lie on top of each other suggesting the particles have evolved to a stable state. Under lipid-matching conditions, on the other hand, the scattering intensity at low $q$ values increases rapidly upon binding of $\alpha$-synuclein to the lipid vesicles and continues to increase during the slower aggregation process (Figure 5 B). Both the change in lipid and protein scattering signal occur on a time scale very similar to that of the accompanying ThT experiments (Figure 5C). These results show that the aggregation of $\alpha$-synuclein is accompanied by a large scale rearrangement of the lipids, strongly supporting the hypothesis that lipid molecules are incorporated into the protein-rich aggregates.[24]

We were able to model the kinetic neutron scattering profiles in 100% $D_2O$ as the emergence of cylinders with increasing dimensions, where the fitting parameters were the radius

24

$R$, length $L$ and scale. At early time points the cylinders presumably represent the rod-like mixtures of DLPS and $\alpha$-synuclein which were also observed in data in Figures S12 and S13. At late time points the cylinders presumably represent lipids distributed along the length of fibrils of $\alpha$-synuclein. The cylinders were assigned a "bulk" scattering length density corresponding to that of pure DLPS. A free scale parameter was required to correct the scattering length density of the lipid particle as more d-$\alpha$-synuclein becomes incorporated and decreases the overall contrast, as well as ambiguity in calculating the number density of scatterers. At $t = 2$ min, already the data indicate rods with $R$=18 Å and $L$=210 Å. Between 1 and 4 hours, there is a significant upturn at low-$q$ in the data which cannot be captured by the single-state cylindrical form factor. We hypothesize that this upturn present in the data collected between 2 min and 4 hours indicates the presence of one or more populations of much larger structures; potentially a population of fibril-like structures which have matured on a shorter time-scale. Therefore the cylindrical model fits to these data capture the smaller population of scatterers. From 6 hours onward, the upturn is no longer present and the model can be employed to describe the entire scattering profiles. The refined parameters indicate long cylinders of at least 700 Å. The data at low-$q$ do not contain a Guinier region or any initial plateau and hence the maximum dimension of the cylinders is outside the size range accessible by the experimental setup. For this reason, when refining the length of these cylinders we obtain a very large uncertainty of around 50%. The associated fit parameters can be found in Table 11.

The data collected in 18% $D_2O$ displays other types of features and reflects the complicated underlying structure of fibrils of $\alpha$-synuclein. There is a drop in forward scattering between 5 minutes and 1 hour (Figure S12 B) indicating that there is some initial transformation in the arrangement of $\alpha$-synuclein within the lipid co-structure before the fibril formation process begins. The scattering profiles collected after 1, 2, 4 and 6 hours fall closely on top of each other showing the protein structure remains largely unchanged and there is a lag in the aggregation process. The scattering profiles at these early time points

contain a strong Guinier range and do not indicate the presence of a sub-population of very large particles, which are obvious in the data collected in 100% $D_2O$ (Figure 5), suggesting that these large intermediate structures must be composed mainly of lipids. Between 6 and 20 hours there is a systematic increase in the forward scattering of the data indicating that the volume of the $\alpha$-synuclein-associated particles is growing. After 18 hours the Guinier range becomes completely diminished indicating the protein also forms structures which are outside the size range accessible by the experimental setup.

Although the low-$q$ data in Figure 5 B resembles cylinders, the data points above $q = 0.02$ Å$^{-1}$ are much more indicative of a flexible structure, which can be explained as the C-terminus of $\alpha$-synuclein which is not likely to become part of the fibril core.
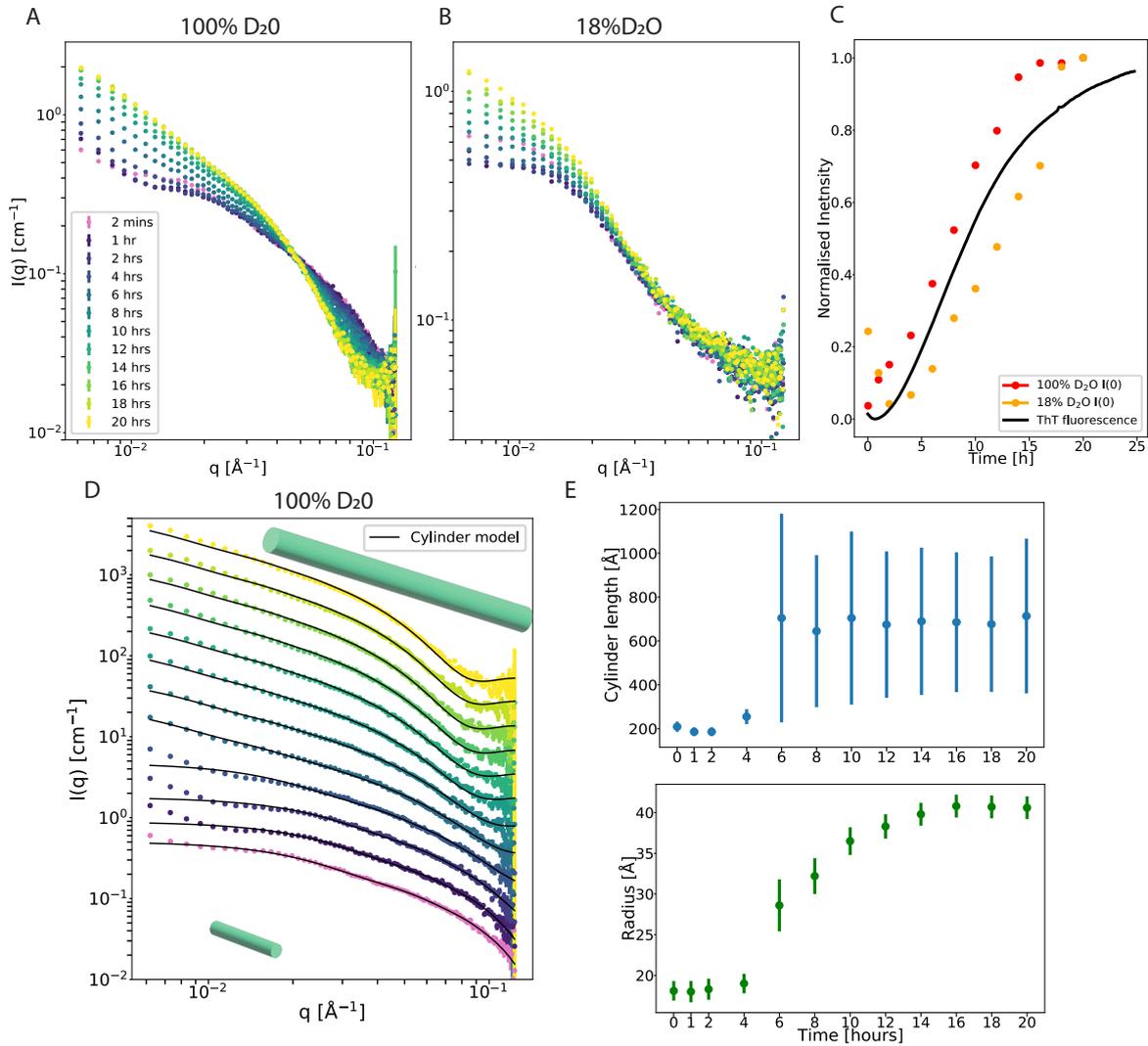
Figure 5: The morphology of DLPS vesicles and $\alpha$-synuclein changes within the same time scale as that of amyloid formation, as monitored by Thioflavin-T fluorescence. Neutron scattering function of the reaction mixture 2 mM DLPS + 200 $\mu$M $\alpha$-synuclein measured at 30 °C over time in buffer conditions corresponding to the contrast matching of the protein, 100% $D_2O$ (A) or the vesicles, 18% $D_2O$ (B). (C) Normalized change in the ThT fluorescence and the forward scattering of the reaction mixture under both contrast conditions. (D) Cylindrical model fits refined against each scattering profile collected under the contrast matching of the protein, 100% $D_2O$. (E) Refined fit parameters from the data fits in (D), cylindrical length and radius as a function of time, indicating significant growth in both dimensions, particularly in the length dimension.

All together, our SANS measurements of the $\alpha$-synuclein-DLPS mixtures show that both the lipid and protein molecules undergo structural re-arrangements characteristics of cylinder

27

Table 11: Structural parameters refined from the SANS data in Figure 5 D using a simple cylinder model model.

|  | | Scale | Radius [Å] | Length [Å] | Background [cm$^{-1}$] | $\chi^2$ |
|---|---|---|---|---|---|---|
|  | 2 mins | 0.69 ± 0.08 | 18.1 ± 1.2 | 209 ± 23 | - | 1.43 |
|  | 1 h | 0.70 ± 0.08 | 18.0 ± 1.3 | 186 ± 20 | - | 5.43 |
|  | 2 h | 0.67 ± 0.08 | 18.3 ± 1.3 | 186 ± 20 | - | 7.41 |
|  | 4 h | 0.61 ± 0.06 | 19.0 ± 1.2 | 255 ± 34 | - | 8.71 |
|  | 6 h | 0.26 ± 0.07 | 28.6 ± 3.2 | 705 ± 476 | 0.023 ± 0.008 | 2.45 |
|  | 8 h | 0.24 ± 0.04 | 32.2 ± 2.2 | 646 ± 350 | 0.025 ± 0.006 | 1.69 |
|  | 10 h | 0.21 ± 0.03 | 36.5 ± 1.7 | 705 ± 395 | 0.026 ± 0.005 | 2.96 |
|  | 12 h | 0.22 ± 0.02 | 38.3 ± 1.5 | 675 ± 334 | 0.025 ± 0.004 | 4.80 |
|  | 14 h | 0.22 ± 0.02 | 39.8 ± 1.4 | 690 ± 336 | 0.024 ± 0.004 | 6.68 |
|  | 16 h | 0.22 ± 0.02 | 40.8 ± 1.4 | 686 ± 319 | 0.024 ± 0.004 | 7.37 |
|  | 18 h | 0.22 ± 0.02 | 40.7 ± 1.4 | 677 ± 309 | 0.024 ± 0.004 | 8.58 |
|  | 20 h | 0.22 ± 0.02 | 40.6± 1.4 | 714 ± 353 | 0.023 ± 0.004 | 9.38 |

growth during the same time scale as that of the formation of amyloid fibrils, demonstrating their co-assemblies.
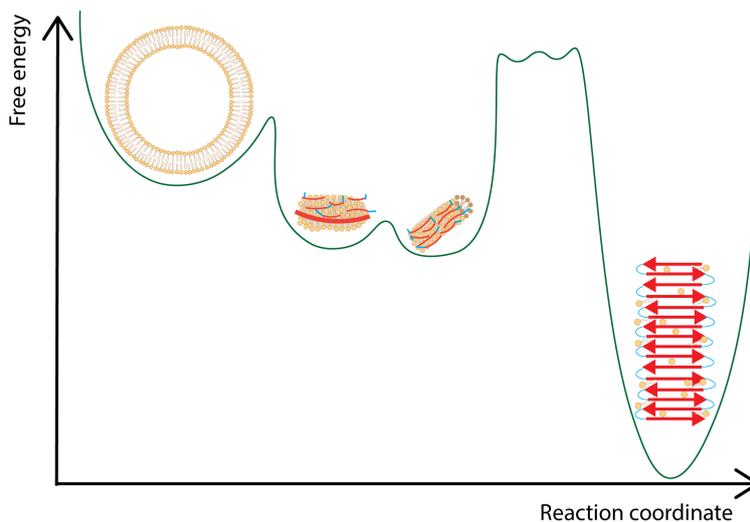


Figure 6: Proposed free-energy surface for the lipid-$\alpha$-synuclein system, here illustrated for the case of DLPS. Upon addition of $\alpha$-synuclein to DLPS vesicles, nearly isoenergetic disks and rods form (time scale ms), followed by a slow conversion in amyloid fibrils (time scale days).

# Discussion

Protein-lipid interactions have long been recognized as a key process involved in the initiation and / or modulation of the formation of amyloid fibrils by peptides and proteins associated

with neurodegeneration, such as $\alpha$-synuclein, amyloid $\beta$ or IAPP.[22,66–68] These proteins have high affinities towards certain types of lipid membranes, depending on the solution conditions and the physical and chemical properties of the lipids, such as charge, length of the acyl chain, degree of unsaturation and phase state of the lipid bilayer.[26] These interactions, as well as their consequences for amyloid fibril formation, have been studied in detail. The relative concentrations of lipids and peptides have emerged as a key control parameter that defines in many cases whether lipids accelerate or slow down amyloid fibril formation .[22,66,68] Detectable (by ThT fluorescence) amounts of amyloid fibrils usually only form if there is a significant excess of monomeric peptide in solution over the amount that saturates the lipid membranes,[22] even though it has also been demonstrated by microscopy that fibrils can form on the membrane at exceedingly low protein concentrations.[69,70] The interactions with lipid membranes are often part of the physiological functions of the amyloidogenic peptide, as is in particular also the case for $\alpha$-synuclein.[7,71] A finely tuned balance appears to exist between benign and detrimental lipid-protein interactions and it has been proposed that the onset of pathology and disease may be caused if the system gets out of balance, e.g. due to age-related chemical modifications of protein and lipids, as well as changes in protein concentration.[66] The mechanism by which lipid-binding can facilitate and induce amyloid fibril formation is still not fully understood for any such system. In the simplest view, the binding of the protein to the lipid bilayer increases the local protein concentration and/or changes protein conformation, both of which can favour aggregation. However, in the framework of this simple mechanism, the lipid bilayer acts as a mere catalyst and, assuming that the aggregates can detach from the bilayer, a small amount of lipid vesicles should be able to induce the amyloid fibril formation of a large excess of protein. However, for $\alpha$-synuclein, this is not observed. On the contrary, the quantity of aggregates formed is proportional to the initial concentration of lipids, and if a small amount of lipid vesicles is added at the beginning, the aggregation reaction stops before all the monomeric protein has been converted into aggregates in the absence of secondary processes.[22,26] This behaviour can only

29

be rationalized if it is assumed that the lipids are a reagent and are being consumed from the membrane as the amyloid aggregates form.

The results of our extensive scattering measurements on $\alpha$-synuclein-lipid systems show that the binding of the protein to lipid vesicles or large planar bilayers leads to a major structural re-arrangement into small particles, that is followed by the incorporation of lipid molecules into the aggregates, and strongly support this model. The main contribution and novelty of the present study is the ability to clearly distinguish between the very rapid (ms time scale) structural change induced by the binding of the protein to the lipid vesicles or large planar bilayers and the structural changes associated with the amyloid formation that occur on a time scale of hours. We were unable to resolve the kinetics of initial binding by stopped-flow SAXS; suggesting that the binding reaction occurs in a few ms, and is essentially diffusion-limited. DLS, SAXS, and SANS experiments show that at temperatures below the melting transition, the protein coats the large planar bilayers, which remain largely intact. At temperatures above the melting transition (ca. 21-23°C for DMPS), this interaction immediately leads to a major structural change of the large planar bilayers, i.e. a break-up into smaller particles. Interestingly, this break-up is largely reversible when the protein is removed, either by cooling down or by enzymatic digestion of the protein. In the SAXS experiments, these smaller particles could be well-described by a model for disc-like particles, reminiscent of $\alpha$-synuclein-stabilized nanodiscs,[42] whereas the SANS data were more consistent with the formation of rod-like structures. This apparent discrepancy can be explained by the fact that these structures are probably very similar in energy, and mixtures of different shapes have indeed previously been observed in a very similar system.[53] Whether a given mixture tends more towards the formation of discs or rods presumably depends on subtle differences, such as for example differences in the solvent (purely hydrogenated water for SAXS, partly or fully deuterated water in the case of SANS). However, both techniques agree in that the resulting structures have significantly lower volume and hence scattering intensity, compared to the original vesicles and planar bilayers. The lipid systems that were characterized in de-

30

tail in this study (DLPS and DMPS), and which were demonstrated to undergo a strong structural disruption upon interaction with $\alpha$-synuclein, are both known to efficiently induce amyloid fibril formation.[22,26] It is worth noting that the amyloid fibril formation induced by these lipids yields highly reproducible kinetic data, in particular also for DMPS,[22] where we find the lipid to initially form heterogeneous bilayer structures, rather than pure vesicles. However, upon contact with $\alpha$-synuclein, homogeneous disk-like structures are formed, which removes most of the initial heterogeneity and explains the well-behaved amyloid fibril formation. In the case of DLPS, the SANS measurements show that the lipids are being incorporated into the fibrils, which is in agreement with a range of previous studies.[23–25,72–74] It is interesting to speculate that there is a close connection between the ability of the protein to disrupt the lipid structures and the lipids to induce amyloid fibril formation. Such a hypothesis is supported by the observation that lipids that form vesicles that are not disrupted (e.g. POPG) by $\alpha$-synuclein binding are also not efficiently inducing amyloid fibril formation.[10] Therefore the simple adsorption onto the lipid bilayer, despite increasing the local concentration of $\alpha$-synuclein on the vesicle surface, does not seem to be the main cause of the observed acceleration of amyloid fibril formation upon lipid binding. Rather, the formation of intricately mixed protein-lipid nanostructures, discs or rods, in which the lipid molecules or membranes might mediate favourable protein-protein interactions, seem to be the most efficient inducer of amyloid fibril formation. This model provides a mechanistic framework that can be used to analyse the effect of lipid modifications that are suspected to favour deleterious protein-protein interactions leading to amyloid fibril formation.

## Conclusion

In conclusion, we have used a combination of time resolved small angle scattering measurements (SAXS and SANS), dynamic light scattering and differential scanning calorimetry in order to investigate the reversibility and structural consequences of the binding of $\alpha$-synuclein

31

to negatively charged lipid membranes (DMPS and DLPS). We find that the initial binding of the protein to the lipid vesicles or planar bilayer structures is extremely rapid (faster than 1 ms), easily reversible and leads to a major structural transition and break-up of intact vesicles and larger bilayer structures into significantly smaller disc- and rod-shaped lipid-protein co-assemblies. The resulting subsequent formation of fibrillar aggregates of $\alpha$-synuclein leads to yet another substantial structural change in the lipids that occur on the same time scale as that of the protein aggregation and hence confirms that lipid-protein co-aggregates are formed.

# Experimental methods

## Protein and lipid samples

Unlabelled $\alpha$-synuclein was purified as described previously.[22] Matchout-deuterated human $\alpha$-synuclein was produced in the Deuteration Laboratory of the Institut Laue-Langevin (ILL D-Lab, Grenoble, France). To obtain d-$\alpha$-synuclein, a transposition reaction was first carried out on the original plasmid in order to modify the resistance selection marker from ampicillin to kanamycin. The Tn5 transposon insertion kit (EZ-Tn5TM <KAN-2> Insertion Kit) from Epicentre® Biotechnologies/Illumina® was used for this transposition reaction. New kanamycin-resistant plasmid containing cDNA coding for human $\alpha$-synuclein was transformed into One ShotTMBL21(DE3) E. coli cells (Invitrogen). Deuterated $\alpha$-synuclein was over-expressed in E.coli strain BL21(DE3) adapted to growth in deuterated minimal medium.[75] A 1.7 L (final volume) deuterated high cell-density fed-batch fermenter culture was carried out at 30°C. Feeding with glycerol was started at an $OD_{600}$ value of about 4.2. Expression of d-$\alpha$-synuclein was induced at an $OD_{600}$ of 19 by addition of IPTG (1mM final concentration). Cells were harvested at an $OD_{600}$ of 23 yielding 74 g wet weight of deuterated cell paste. Lipid vesicles were prepared via sonication as described previously .[22]

## Circular Dichroism (CD) spectroscopy

CD samples were prepared by incubating 267 $\mu$M $\alpha$-synuclein in the presence of 8 mM DMPS in 20 mM sodium phosphate buffer (pH 6.5). Far-UV CD spectra were recorded using a JASCO J-810 spectropolarimeter equipped with a Peltier thermally controlled cuvette holder at 30°C. The spectra were measured before and after incubation of the sample with 15.7 $\mu$M proteinase-K for 1 hours. The samples were diluted 10 times before measurement.

## Differential Scanning Calorimetry

The DSC samples were prepared by incubating 0.86 mM DMPS in the absence or presence of 28.6 $\mu$M $\alpha$-synuclein in 20mM sodium phosphate buffer pH 6.5. DSC thermograms of these samples untreated or treated with 1.7 $\mu$M proteinase-K for 1 h at 20 °C were acquired using a Microcal VP-DSC calorimeter (Malvern Instruments) with a scanning rate of 1°C.min$^{-1}$ from 5 to 65°C. Protein and lipid samples were degassed for 20 min at room temperature before mixing and acquisition of the DSC thermograms. The DSC thermograms reported in this article were corrected by subtracting the thermogram of the phosphate buffer and normalized with respect to the lipid concentration.

## Small Angle X-ray Scattering

### Measurements at ESRF in Grenoble

SAXS experiments were performed at the beamline ID02 at ESRF, Grenoble. SAXS signals were measured at a distance of 2 m. For temperature series, standard temperature-controlled capillary holder at ID02 (capillary diameter 1.5 mm) was used. A fresh solution spot was used for each temperature to avoid effects of radiation damage. 3 mM DMPS was incubated in the absence or presence of 50 or 100 $\mu$M $\alpha$-synuclein in phosphate buffer pH 6.5 at temperatures ranging from 15 to 65°C. For kinetic series with rapid mixing, a commercial stopped-flow setup (Biologic, SFM 400) adapted for use at ID02 with a special X-ray head

33

with a measurement capillary (1.5 mm diameter) was used.[76] The dead time was found to range between 2-3 ms. The samples were prepared by mixing DMPS solution with either buffer or $\alpha$-synuclein solutions to reach the final concentrations specified in the paper. Effects of radiation damage was checked beforehand in static samples, and the acquisition times were adapted accordingly. The modelling/fitting of the resulting SAXS data are described in details in the Supplementary Information.

### Measurements with the laboratory X-ray source

SAXS measurements were performed in static capillaries mounted on a SAXSLAB Ganesha 300XL equipped with a temperature regulator Julabo CF41 water bath. 8 mM DMPS was incubated in the absence or presence of 270 $\mu$M $\alpha$-synuclein in phosphate buffer pH 6.5 and the SAXS function was recorded before and after proteinase-K treatment. The sample was and the SAXS function was measured at a q range of 0.003-0.73 Å.

## Small Angle Neutron Scattering

SANS measurements were performed on D22 instrument of Institut Laue Langevin. The samples were in a 1 mm thick suprasil quartz cuvette (Hellma QS 100-1-40), mounted on a 22-position sample rack temperature-regulated by two water baths, enabling a rapid switching between 20, 30 and 45°C. 8 mM DMPS was incubated in the absence or presence of 270 $\mu$M $\alpha$-synuclein in 20mM phosphate buffer pH 6.5. For static measurements, the SANS function was recorded at a q range of 0.006-0.635 Å using two collimation:sample-detector distance sets: one for small angles (11.2m:11.2m) and one for large angles (1.4m:2m). Kinetics measurements were only performed at small angle configuration. The collimation was rectangular of 40 mm width and 55 mm height, the sample aperture was of 7mm width and 10 mm height and the wavelength was 6 A +/- 10%. Data were reduced using Grasp software, including blocked beam and empty cell background subtraction, sample thickness normalisation, and absolute intensity scaling using direct beam intensity measurement. Data

are permanently curated at ILL.[77,78]

**Aggregation measurements**

Samples were prepared by mixing $\alpha$-synuclein, DLPS or DMPS solutions at the protein-to-lipid ratios indicated in the manuscript together with Th-T ($50\mu$M) in 20 mM sodium phosphate buffer (pH 6.5). The kinetics of amyloid formation were measured at $30^{circ}$C under quiescent conditions in Corning 96 well plates with half-area (black/clear bottom polystyrene) non-binding surfaces using a BMG plate reader.

## Dynamic light scattering

The temperature-ramp experiments were performed with a Prometheus Panta (Nanotemper, Munich), which is a microcapillary-based instrument that combines differential scanning fluorimetry (DSF) and dynamic light scattering (DLS) measurements. In the present study we used solely the DLS functionality. The samples were prepared with protein and lipid solutions and buffer that was cooled to below the starting temperature of the temperature ramp ($15°$C). We measured pure lipids (DMPS and DLPS) at 1 mM, as well as in the presence of 9.5 and 95 $\mu$M $\alpha$-synuclein. The temperature was varied from 15-35°C and back to 15°C, at two different scan rates (0.5 and 5°C/min) and we plot the cumulant radius from the DLS measurements, as well as the absolute light scattering intensity, as a function of temperature.

## Acknowledgement

## Supporting Information Available

The following files are available free of charge.

- Supplementary Figures

  Figure S1: Additional repeats of the stopped-flow SAXS measurements of the binding between DMPS or DLPS model membranes and $\alpha$-synuclein

  Figure S2: Differential scanning calorimetry and circular dichroism measurements of the $\alpha$-synuclein:DMPS system before and after proteinase-K treatment

  Figure S3: X-ray scattering control experiments for the PK-treatment of the DMPS:$\alpha$-synuclein system

  Figure S4: X-ray scattering profiles of DMPS model membranes at different temperatures

  Figure S5: Determination of the contrast match points of the protein and DMPS and DLPS model membranes at different temperatures

- Supplementary Methods - Models for the analysis of the SAS data

  Figure S6: Schematic of the three-shell vesicle model

  Figure S7: Schematic of the local planar bilayer model

  Figure S8: Schematic of the local planar bilayer model with Gaussian random coils

  Figure S9: Schematic of the disc model with Gaussian random coils

  Figure S10: Schematic of the core-shell ribbon model

# References

(1) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884–890.

(2) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. Alpha-synuclein in Lewy bodies. *Nature* **1997**, *388*, 839–840.

(3) Mahul-Mellier, A.-L.; Burtscher, J.; Maharjan, N.; Weerens, L.; Croisier, M.; Kuttler, F.; Leleu, M.; Knott, G. W.; Lashuel, H. A. The process of Lewy body formation, rather than simply $\alpha$-synuclein fibrillization, is one of the major drivers of neurodegeneration. *Proceedings of the National Academy of Sciences* **2020**, *117*, 4971–4982.

37

(4) Burré, J.; Vivona, S.; Diao, J.; Sharma, M.; Brunger, A. T.; Südhof, T. C. Properties of native brain $\alpha$-synuclein. *Nature* **2013**, *498*, E4–6; discussion E6–7.

(5) Theillet, F.-X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuiver, M.; Verzini, S.; Lorenz, D.; van Rossum, M.; Goldfarb, D.; Selenko, P. Structural disorder of monomeric $\alpha$-synuclein persists in mammalian cells. *Nature* **2016**, *530*, 45–50.

(6) Fusco, G.; Simone, A. D.; Gopinath, T.; Vostrikov, V.; Vendruscolo, M.; Dobson, C. M.; Veglia, G. Direct observation of the three regions in $\alpha$-synuclein that determine its membrane-bound behaviour. *Nat Commun* **2014**, *5*, 3827.

(7) Bendor, J.; Logan, T.; Edwards, R. H. The Function of $\alpha$-Synuclein. *Neuron* **2013**, *79*, 1044–1066.

(8) He-Jin Lee, C. C.; Lee, S.-J. Membrane-bound $\alpha$-Synuclein Has a High Aggregation Propensity and the Ability to Seed the Aggregation of the Cytosolic Form. *J Biol Chem* **2002**, *277*, 671–678.

(9) Zhu, M.; Li, J.; Fink, A. L. The Association of Alpha-Synuclein With Membranes Affects Bilayer Structure, Stability, and Fibril Formation. *J. Bio. Chem.* **2003**, *278*, 40186–40197.

(10) Cholak, E.; Bucciarelli, S.; Bugge, K.; Vestergaard, N. T. J. B.; Arleth, L.; Kragelund, B. B.; Langkilde, A. E. Distinct $\alpha$-Synuclein:Lipid Co-Structure Complexes Affect Amyloid Nucleation Through Fibril Mimetic Behavior. *Biochemistry* **2019**, *58*, 5052–5065.

(11) Chaudhary, H.; Stefanovic, A. N. D.; Subramaniam, V.; Claessens, M. M. A. E. Membrane interactions and fibrillization of $\alpha$-synuclein play an essential role in membrane disruption. *FEBS Lett* **2014**, *588*, 4457–4463.

(12) Varkey, J.; Isas, J. M.; Mizuno, N.; Jensen, M. B.; Bhatia, V. K.; Jao, C. C.; Petrlova, J.; Voss, J. C.; Stamou, D. G.; Steven, A. C.; Langen, R. Membrane curvature induction and tubulation are common features of synucleins and apolipoproteins. *Journal of Biological Chemistry* **2010**, *285*, 32486–32493.

(13) Pandey, A. P.; Haque, F.; Rochet, J.-C.; Hovis, J. S. α-Synuclein-induced tubule formation in lipid bilayers. *The Journal of Physical Chemistry B* **2011**, *115*, 5886–5893.

(14) Mizuno, N.; Varkey, J.; Kegulian, N. C.; Hegde, B. G.; Cheng, N.; Langen, R.; Steven, A. C. Remodeling of lipid vesicles into cylindrical micelles by α-synuclein in an extended α-helical conformation. *Journal of Biological Chemistry* **2012**, *287*, 29301–29311.

(15) Jiang, Z.; D.Flynn, J.; Jr, W. E. T.; Gawrisch, K.; Lee, J. C. Stimulation of α-synuclein amyloid formation by phosphatidylglycerol micellar tubules. *BBA- Biomembranes* **2018**, *1860*, 1840–1847.

(16) Varkey, J.; Mizuno, N.; Hegde, B. G.; Cheng, N.; Steven, A. C.; Langen, R. α-Synuclein oligomers with broken helical conformation form lipoprotein nanoparticles. *Journal of Biological Chemistry* **2013**, *288*, 17620–17630.

(17) Eichmann, C.; Campioni, S.; Kowal, J.; Maslennikov, I.; Gerez, J.; Liu, X.; Verasdonck, J.; Nespovitaya, N.; Choe, S.; Meier, B. H.; Picotti, P.; Rizo, J.; Stahlberg, H.; Riek, R. Preparation and Characterization of Stable alpha-Synuclein Lipoprotein Particles. *J Biol Chem* **2016**, *291*, 8516–8527.

(18) Viennet, T.; Wördehoff, M. M.; Uluca, B.; Poojari, C.; Shaykhalishahi, H.; Willbold, D.; Strodel, B.; Heise, H.; Buell, A. K.; Hoyer, W.; Etzkorn, M. Structural insights from lipid-bilayer nanodiscs link α-Synuclein membrane-binding modes to amyloid fibril formation. *Communications biology* **2018**, *1*, 44.

(19) Stöckl, M.; Claessens, M. M. A. E.; Subramaniam, V. Kinetic measurements give new insights into lipid membranepermeabilization by $\alpha$-synuclein oligomers. *Mol. BioSyst.* **2012**, *8*, 338–345.

(20) Lorenzen, N.; Nielsen, S. B.; Buell, A. K.; Kaspersen, J. D.; Arosio, P.; Vad, B. S.; Paslawski, W.; Christiansen, G.; Valnickova-Hansen, Z.; Andreasen, M.; Enghild, J. J.; Pedersen, J. S.; Dobson, C. M.; Knowles, T. P. J.; Otzen, D. E. The Role of Stable $\alpha$-Synuclein Oligomers in the Molecular Events Underlying Amyloid Formation. *J Am Chem Soc* **2014**, *136*, 3859–3868.

(21) Ghio, S.; Camilleri, A.; Caruana, M.; Ruf, V. C.; Schmidt, F.; Leonov, A.; Ryazanov, S.; Griesinger, C.; Cauchi, R. J.; Kamp, F.; Giese, A.; Vassallo, N. Cardiolipin Promotes Pore-Forming Activity of Alpha-Synuclein Oligomers in Mitochondrial Membranes. *ACS Chem Neurosc* **2019**, *10*, 3815–3829.

(22) Galvagnion, C.; Buell, A. K.; Meisl, G.; Michaels, T. C. T.; Vendruscolo, M.; Knowles, T. P. J.; Dobson, C. M. Lipid vesicles trigger $\alpha$-synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol* **2015**, *11*, 229–234.

(23) Chaudhary, H.; Subramaniam, V.; Claessens, M. M. A. E. Direct Visualization of Model Membrane Remodeling by $\alpha$-Synuclein Fibrillization. *ChemPhysChem* **2017**, *18*, 1620–1626.

(24) Galvagnion, C.; Topgaard, D.; Makasewicz, K.; Buell, A. K.; Linse, S.; Sparr, E.; Dobson, C. M. Lipid Dynamics and Phase Transition within $\alpha$-Synuclein Amyloid Fibrils. *J Phys Chem Lett* **2019**, *10*, 7872–7877.

(25) Frieg, B.; Antonschmidt, L.; Dienemann, C.; Geraets, J. A.; Najbauer, E. E.; Matthes, D.; de Groot, B. L.; Andreas, L. B.; Becker, S.; Griesinger, C.; F, S. G. The 3D structure of lipidic fibrils of $\alpha$-synuclein. *Nature Communications* **2022**, *13*, 6810.

(26) Galvagnion, C.; Brown, J. W. P.; Ouberai, M. M.; Flagmeier, P.; Vendruscolo, M.; Buell, A. K.; Sparr, E.; Dobson, C. M. Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of $\alpha$-synuclein. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 7065–7070.

(27) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* **2011**, *133*, 14160–14163.

(28) van Gils, J. H. M.; van Dijk, E.; Peduzzo, A.; Hofmann, A.; Vettore, N.; Schützmann, M. P.; Groth, G.; Mouhib, H.; Otzen, D. E.; Buell, A. K.; Abeln, S. The hydrophobic effect characterises the thermodynamic signature of amyloid fibril growth. *PLOS Computational Biology* **2020**, *16*, e1007767.

(29) Agerschou, E. D. et al. An engineered monomer binding-protein for $\alpha$-synuclein efficiently inhibits the proliferation of amyloid fibrils. *eLife* **2019**, *8*, e46112.

(30) Panine, P.; Finet, S.; Weiss, T.; Narayanan, T. Probing fast kinetics in complex fluids by combined rapid mixing and small-angle X-ray scattering. *Adv Coll Interf Sc* **2006**, *127*, 9–18.

(31) Haertlein, M.; Moulin, M.; Devos, J. M.; Laux, V.; Dunne, O.; Forsyth, V. T. *Methods in enzymology*; Elsevier, 2016; Vol. 566; pp 113–157.

(32) SH, S. et al. Lewy pathology in Parkinson's disease consists of crowded organelles and lipid membranes. *Nat Neurosci. Jul;22(7):1099-1109. doi: 10.1038/s41593-019-0423-2. Epub 2019 Jun 24. PMID: 31235907.* **2019**, *22*, 1099–1109.

(33) Glatter, O. A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr.* **1977**, *10*, 415–421.

41

(34) Glatter, O. In *Scatt. Methods their Appl. Colloid Interface Sci.*; Glatter, O., Ed.; Elsevier, 2018; pp 33–74.

(35) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **2004**, *126*, 3477–3487.

(36) Johansen, N. T.; Luchini, A.; Tidemand, F. G.; Orioli, S.; Martel, A.; Porcar, L.; Arleth, L.; Pedersen, M. C. Structural and Biophysical Properties of Supercharged and Circularized Nanodiscs. *Langmuir* **2021**, *37*, 6681–6690.

(37) Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: Modeling and least-squares fitting. *Adv. Colloid Interface Sci.* **1997**, *70*, 171–210.

(38) Pedersen, M. C.; Arleth, L.; Mortensen, K. WillItFit: a framework for fitting of constrained models to small-angle scattering data. *Journal of Applied Crystallography* **2013**, *46*, 1894–1898.

(39) Armen, R. S.; Uitto, O. D.; Feller, S. E. Phospholipid component volumes: Determination and application to bilayer structure calculations. *Biophys. J.* **1998**, *75*, 734–744.

(40) Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F. Structure and Fluctuations of Charged Phosphatidylserine Bilayers in the Absence of Salt. *Biophysical Journal* **2004**, *86*, 1574–1586.

(41) Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters* **2002**, *2*, 853–856.

(42) Eichmann, C.; Bibow, S.; Riek, R. $\alpha$-Synuclein lipoprotein nanoparticles. *Nanotechnol Rev* **2017**, *6*, 105–110.

(43) Skar-Gislinge, N.; Simonsen, J. B.; Mortensen, K.; Feidenhans'l, R.; Sligar, S. G.; Lindberg Møller, B.; Bjørnholm, T.; Arleth, L. Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the lipids and the protein. *Journal of the American Chemical Society* **2010**, *132*, 13713–13722.

(44) Skar-Gislinge, N.; Arleth, L. Small-angle scattering from phospholipid nanodiscs: Derivation and refinement of a molecular constrained analytical model form factor. *Physical Chemistry Chemical Physics* **2011**, *13*, 3161–3170.

(45) Pedersen, J. S.; Gerstenberg, M. C. Scattering form factor of block copolymer micelles. *Macromolecules* **1996**, *29*, 1363–1365.

(46) Pedersen, J. S. Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores. *Journal of Applied Crystallography* **2000**, *33*, 637–640.

(47) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R., et al. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences* **2004**, *101*, 12491–12496.

(48) Mylonas, E.; Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of applied crystallography* **2007**, *40*, s245–s249.

(49) Lund, R.; Willner, L.; Richter, D.; Lindner, P.; Narayanan, T. Kinetic Pathway of the Cylinder-to-Sphere Transition in BlockCopolymer Micelles Observed in Situ by Time-Resolved Neutron and Synchrotron Scattering. *ACS Macro Lett* **2013**, *2*, 1082–1087.

(50) von Smoluchowski, M. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift für Physikalische Chemie* **1917**, *92*, 129.

(51) Wolff, M.; Mittag, J. J.; Herling, T. W.; Genst, E. D.; Dobson, C. M.; Knowles, T. P. J.; Braun, D.; Buell, A. K. Quantitative thermophoretic study of disease-related protein aggregates. *Scientific reports* **2016**, *6*, 22829.

(52) Vijayakumar, M.; Wong, K.-Y.; Schreiber, G.; Fersht, A.; Szabo, A.; Zhou, H.-X. Electrostatic Enhancement of Diffusion-controlled Protein-Protein Association: Comparison of Theory and Experiment on Barnase and Barstar. *J Mol Biol* **1998**, *278,*, 1015–1024.

(53) Hoover, B. M.; Shen, Z.; Gahan, C. G.; Lynn, D. M.; Van Lehn, R. C.; Murphy, R. M. Membrane remodeling and stimulation of aggregation following $\alpha$-synuclein adsorption to phosphotidylserine vesicles. *The Journal of Physical Chemistry B* **2021**, *125*, 1582–1594.

(54) Chen, C. H. Interactions of lipid vesicles with solvent in heavy and light water. *J. Phys. Chem.* **1982**, *86*, 3559–3562.

(55) Emerson, M. F.; Holtzer, A. Hydrophobic bond in micellar systems. Effects of various additives on the stability of micelles of sodium dodecyl sulfate and of n-dodecyltrimethylammonium bromide. *J. Phys. Chem.* **1967**, *71*, 3320–3330.

(56) Cinelli, S.; Onori, G.; Santucci, A. Solvent isotope effects on the phase-transition properties of lipid bilayers. *Colloids Surfaces B Biointerfaces* **2001**, *20*, 297–302.

(57) Maybury, R. H.; KATZ, J. J. Protein denaturation in heavy water. *Nature* **1956**, *177*, 629–630.

(58) Oas, T. G.; Toone, E. J. Thermodynamic solvent isotope effects and molecular hydrophobicity. *Advances in Biophysical Chemistry* **1997**, *6*, 1–52.

(59) Makhatadze, G. I.; Clore, G. M.; Gronenborn, A. M. Solvent isotope effect and protein stability. *Nature structural biology* **1995**, *2*, 852–855.

(60) Hattori, A.; Crespi, H. L.; Katz, J. J. Effect of side-chain deuteration on protein stability. *Biochemistry* **1965**, *4*, 1213–1225.

(61) Nichols, P. J.; Falconer, I.; Griffin, A.; Mant, C.; Hodges, R.; McKnight, C. J.; Vögeli, B.; Vugmeyster, L. Deuteration of nonexchangeable protons on proteins affects their thermal stability, side-chain dynamics, and hydrophobicity. *Protein Sci.* **2020**, *29*, 1641–1654.

(62) Meilleur, F.; Contzen, J.; Myles, D. A.; Jung, C. Structural stability and dynamics of hydrogenated and perdeuterated cytochrome P450cam (CYP101). *Biochemistry* **2004**, *43*, 8744–8753.

(63) Tempra, C.; Chamorro, V. C.; Jungwirth, P. Effects of Water Deuteration on Thermodynamic and Structural Properties of Proteins and Biomembranes. *J. Phys. Chem. B* **2023**, *127*, 1138–1143.

(64) Mahler, F.; Meister, A.; Vargas, C.; Durand, G.; Keller, S. Self-Assembly of Protein-Containing Lipid-Bilayer Nanodiscs from Small-Molecule Amphiphiles. *Small* **2021**, *17*, 2103603.

(65) Vargas, C.; Arenas, R. C.; Frotscher, E.; Keller, S. Nanoparticle self-assembly in mixtures of phospholipids with styrene/maleic acid copolymers or fluorinated surfactants. *Nanoscale* **2015**, *7*, 20685–20696.

(66) Galvagnion, C. The Role of Lipids Interacting with $\alpha$-Synuclein in the Pathogenesis of Parkinson's Disease. *Journal of Parkinson's Disease* **2017**, *7*, 433–450.

(67) Zhang, X.; St. Clair, J. R.; London, E.; Raleigh, D. P. Islet amyloid polypeptide membrane interactions: effects of membrane composition. *Biochemistry* **2017**, *56*, 376–390.

(68) Habchi, J.; Chia, S.; Galvagnion, C.; Michaels, T. C. T.; Bellaiche, M. M. J.; Ruggeri, F. S.; Sanguanini, M.; Idini, I.; Kumita, J. R.; Sparr, E.; Linse, S.; Dobson, C. M.;

Knowles, T. P. J.; Vendruscolo, M. Cholesterol catalyses Aβ42 aggregation through a heterogeneous nucleation pathway in the presence of lipid membranes. *Nature chemistry* **2018**,

(69) Rabe, M.; Soragni, A.; Reynolds, N. P.; Verdes, D.; Liverani, E.; Riek, R.; Seeger, S. On-surface aggregation of α-synuclein at nanomolar concentrations results in two distinct growth mechanisms. *ACS Chem Neurosci* **2013**, *4*, 408–417.

(70) Banerjee, S.; Hashemi, M.; Zagorski, K.; Lyubchenko, Y. L. Cholesterol in membranes facilitates aggregation of amyloid β protein at physiologically relevant concentrations. *ACS Chemical Neuroscience* **2021**, *12*, 506–516.

(71) Lautenschläger, J.; Stephens, A. D.; Fusco, G.; Ströhl, F.; Curry, N.; Zacharopoulou, M.; Michel, C. H.; Laine, R.; Nespovitaya, N.; Fantham, M., et al. C-terminal calcium binding of α-synuclein modulates synaptic vesicle interaction. *Nature communications* **2018**, *9*, 1–13.

(72) Hellstrand, E.; Nowacka, A.; Topgaard, D.; Linse, S.; Sparr, E. Membrane lipid co-aggregation with α-synuclein fibrils. *PLoS One* **2013**, *8*, e77235.

(73) van Maarschalkerweerd and Valeria Vetri and Annette Eva Langkilde and Vito Foderà and Bente Vestergaard, A. Protein/Lipid Coaggregates are Formed During α-Synuclein-Induced Disruption of Lipid Bilayers. *Biomacromolecules* **2014**, *15*, 3643–3654.

(74) Fridolf, S.; Pham, Q. D.; Pallbo, J.; Bernfur, K.; Linse, S.; Topgaard, D.; Sparr, E. Ganglioside GM3 stimulates lipid-protein co-assembly in α-synuclein amyloid formation. *Biophys Chem* **2022**, *293*, 106934.

(75) Haertlein, M.; Moulin, M.; Devos, J. M.; Laux, V.; Dunne, O.; Trevor Forsyth, V. In *Methods in Enzymology*; Kelman, Z., Ed.; Isotope Labeling of Biomolecules - Applications; Academic Press, 2016; Vol. 566; pp 113–157.

46

(76) Narayanan, T.; Gummel, J.; Gradzielski, M. In *Probing the Self-Assembly of Unilamellar Vesicles Using Time-Resolved SAXS*; Iglič, A., Kulkarni, C. V., Eds.; Advances in Planar Lipid Bilayers and Liposomes; Academic Press, 2014; Vol. 20; pp 171–196.

(77) Galvagnion, C.; Buell, A. K.; Dobson, C.; Martel, A.; Roosen-Runge, F. Real-time monitoring of the change in lipid vesicle structure upon amyloid fibril formation. 2015; Institut Laue-Langevin (ILL), doi:10.5291/ILL-DATA.8-02-754.

(78) Galvagnion, C.; Buell, A. K.; Dobson, C.; Martel, A.; Porcar, L.; Roosen-Runge, F. Structural investigation of early protein-lipid coaggregates. 2015; Institut Laue-Langevin (ILL), doi:10.5291/ILL-DATA.8-03-874.

47

1

# Structural characterisation of

# $\alpha$-synuclein-membrane interactions and the

# resulting aggregation using small angle

# scattering - Supplementary materials

Céline Galvagnion,*,† Abigail Barclay,‡ Katarzyna Makasewicz,¶ Frederik
Ravnkilde Marlet,† Martine Moulin,§ Juliette Devos,§ Sara Linse,¶ Anne Martel,§
Lionel Porcar,§ Emma Sparr,‖ Martin Cramer Pedersen,‡ Felix Roosen-Runge,⊥
Lise Arleth,‡ and Alexander K. Buell*,#

†*Department of Drug Design and Pharmacology, University of Copenhagen, 2100
Copenhagen, Denmark*
‡*Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark*
¶*Department of Biochemistry and Structural Biology, Lund University , SE22100 Lund,
Sweden*
§*Institut Laue-Langevin, 71 avenue des Martyrs, 38042 Grenoble, France*
‖*Division of Physical Chemistry, Center for Chemistry and Chemical Engineering, Lund
University, P.O. Box 124, SE-22100, Lund, Sweden*
⊥*Department of Biomedical Sciences and Biofilms Research Center for Biointerfaces,
Malmö University, Malmö, Sweden*
#*Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800
Kgs. Lyngby, Denmark*

E-mail: celine.galvagnion@sund.ku.dk; alebu@dtu.de

2

# Supplementary Figures



Figure S1: Additional repeats of the measurements shown in Figure 2 A-C. Change in the X-ray scattering function with time when 2 mM DLPS (A,B) or DMPS (C) was mixed with 70 $\mu$M $\alpha$-synuclein using a stopped-flow set-up.
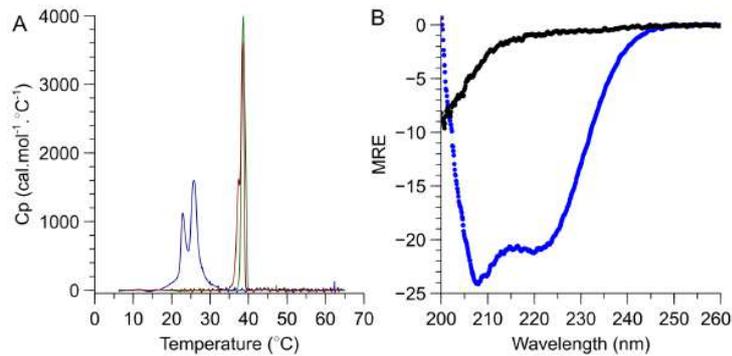


Figure S2: Differential scanning calorimetry and circular dichroism measurements of the $\alpha$-synuclein:DMPS system before and after proteinase-K treatment. A. Differential scanning calorimetry scans of 0.86 mM DMPS in the absence (red) or the presence of 28.6 $\mu$M $\alpha$-synuclein before (blue) and after (green) treatment with 1.7 $\mu$M PK for 1h at 20°C. B. Mean Residue Ellipticity spectrum of $\alpha$-synuclein and DMPS measured at 37 °C before (blue) and after PK treatment (black). The samples were prepared by incubating 8 mM DMPS, 267 $\mu$M $\alpha$-synuclein $\pm$ 15.7 $\mu$M PK for 1h and then diluted 10 times for the CD measurements.
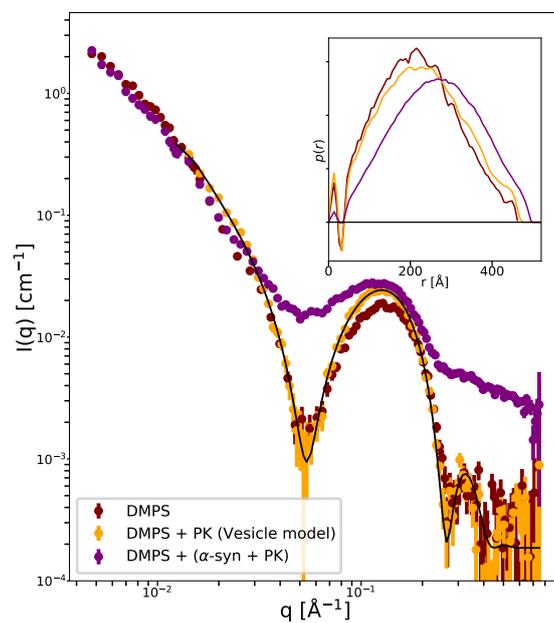
Figure S3: X-ray scattering control experiments for the PK-treatment of the DMPS:$\alpha$-synuclein system. X-ray scattering profiles of 8 mM DMPS vesicles before (red) and after treatment with pK (orange), and in the presence of 270 $\mu$M $\alpha$-synuclein pre-treated with pK (purple). The data were measured in phosphate buffer pH 6.5 at 37°C. The inserts show the p(r)-distribution profiles generated from the SAXS data.
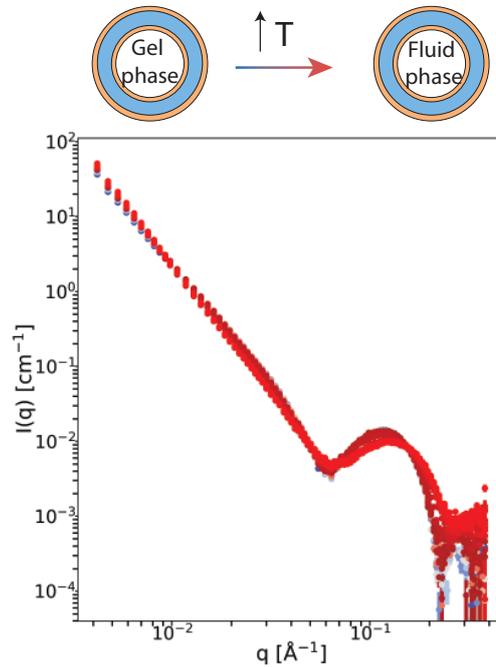
Figure S4: X-ray scattering profiles of DMPS model membranes at different temperatures. X-ray scattering profiles of 3 mM DMPS heated from 14 °C - 49 °C,
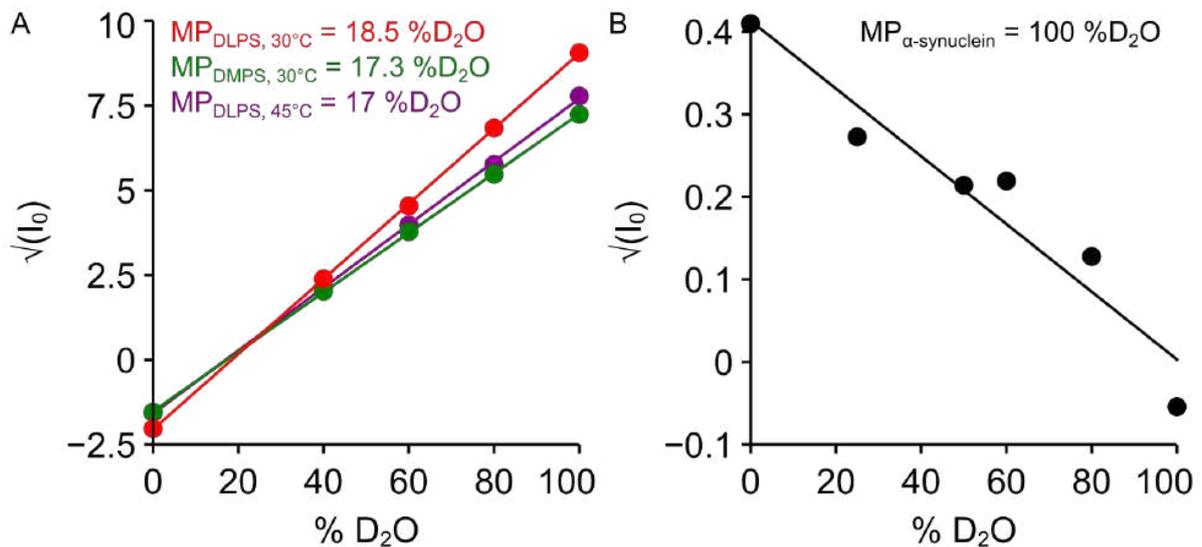


Figure S5: Determination of the contrast match points of the protein and DMPS and DLPS model membranes at different temperatures. A,B. Change in the square root of the intensity of the neutron scattering function of model membranes made with DLPS or DMPS (3mM) (A) or $\alpha$-synuclein (200 $\mu$M) (B) at different temperatures. The samples were measured in 20 mM phosphate buffer pH 6.5.

# Supplementary Methods - Models for the analysis of the SAS data

## Three-shell vesicle

The theoretical scattering form factor for polydisperse three-shell bilayer vesicles can be calculated as a spherical structure with a core representing the solvent surrounded by three shells. Polydispersity is taken into account by assuming a Gaussian distribution of sizes and fitting the sigma of the standard deviation of radii and thicknesses. The innermost and outermost shells have the relative scattering length density corresponding to the lipid hydrophilic headgroups, while the centre shell corresponds to the lipid hydrophobic tailgroups. The model is parameterised by i) the radius to the centre of the shell,s $r$, ii) the thickness of a single lipid, $t$, iii) the volume of a single lipid which is only allowed to vary within a few percent of the estimated nominal volume, $\nu_L$, iv) the sigma of the standard deviation of radii, $\sigma_r$, v) the sigma of the standard deviation of lipid thickness. The inner and outer leaflets are assumed to have the same thickness, $t$. $t$ is divided into $t_{\text{heads}}$ and $t_{\text{tails}}$ in proportion with their respective volumes. By dividing the volume of the vesicle by the volume of a single lipid, the average number of lipids per vesicle can be calculated and used to convert lipid concentration in molar into particle number density. In this way the model is calculated on absolute scale.
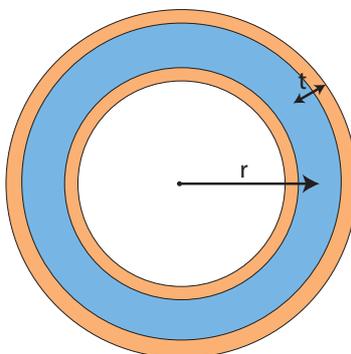


Figure S6: Schematic of the three-shell vesicle model where lipid headgroups are represented in orange and lipid tails are represented in blue. Not to scale.

6

## Planar bilayer structure

The theoretical scattering form factor for lamellar phases[1,2] can be used to model "infinitely large" planar bilayer structures. The form factor of the local membrane structure can be seen at high-$q$, approximately above $q = 0.01$ Å$^{-1}$. The model can be implemented with either a uniform scattering length density or with distinct headgroup and tailgroup regions.

The model is parameterised by i) the volume of single lipid, $V_L$, ii) the thickness of a single lipid, $t$, iii) the sigma of the standard deviation of thicknesses, $\sigma_t$.
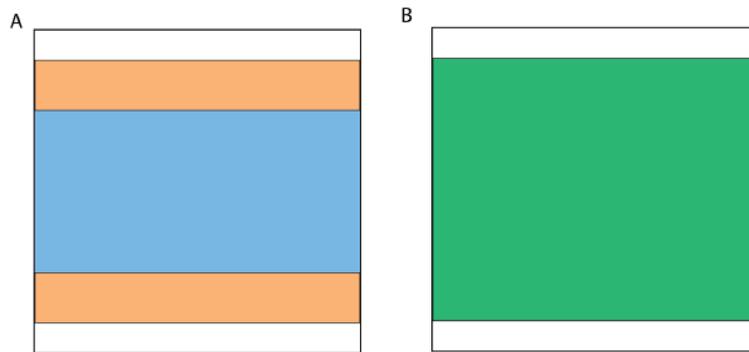


Figure S7: Schematic of the local planar bilayer model which focuses on the membrane structure. The absolute size of the particle cannot be resolved. A The orange and blue sections represent the lipid headgroups and tailgroups respectively. B Uniform contrast representing a simple plane.

## Planar bilayer structure with Gaussian random coils

With this model, $\alpha$-synuclein which is free in solution can additionally be accounted for alongside planar bilayer structures. The protein is described as Gaussian random coils. The model is calculated as

$$P(q) = P(q)_{\text{planar}} + n_{\text{protein}} \cdot P(q)_{\text{GRC}} \tag{1}$$

where $P(q)_{\text{planar}}$ is the form factor for lamella phases, $n_{\text{protein}}$ is the particle number density of the protein and $P(q)_{\text{GRC}}$ is the Debye form factor.[3] The model has an additional parameter, iv) the radius of gyration of the Gaussian random coils, $R_g$.

7
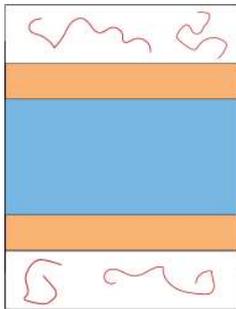
Figure S8: Schematic of the local planar bilayer model Gaussian random coils.

## Discs decorated with Gaussian random coils

This model is akin to the model which has been used frequently to describe phospholipid nanodiscs[4,5] with a slight modification to account for the bound $\alpha$-synuclein inspired by the SAXS model for $\alpha$-synuclein-decorated vesicles presented in Cholak *et al*[6]. The geometric structure of the lipid bilayer is described as a collection of cylinders, representing the lipid headgroups and tails. The tail structure is surrounded by a hollow cylinder in this case representing a belt of $\alpha$-synuclein. Polydispersity is taken into account by assuming a Gaussian distribution of radii. The model additionally accounts for scattering arising from $\alpha$-synuclein embedded in the lipid bilayer by altering the scattering length densities assigned to each cylinder using Eqn 2. Gaussian random coils protrude from the outside of the disc using the same method derived by Pedersen and Gerstenberg[7,8] and used by Arleth et al for their model of PEG-covered micelles[9].

The model is parameterised by i) the radius of the disk, $r$, ii) the sigma of the standard deviation of radii, $\sigma_r$, iii) the thickness of a single lipid $t$, iv) the volume of a single lipid $V_L$, v) the volume of $\alpha$-synuclein, $V_{\alpha\text{-syn}}$, vi) the width of the protein belt, $w_{\text{belt}}$, vii) the radius of gyration of the Gaussian random coils, $R_{g\,\alpha\text{-syn}}$, and viii) the molar ratio of lipid and bound protein, $L/P$.

The average number of lipids per disc, $N_L$ is calculated as $V_{\text{disc}}$ / ($V_L$ + L/P * $V_{\alpha\text{-syn:bilayer}}$) where $V_{\text{disc}}$ is the volume of the disc excluding the protein belt and $V_{\alpha\text{-syn:bilayer}}$ is the volume of the fraction of $\alpha$-synuclein embedded in the bilayer. The number of random coils per disc

8

is $N_L/(\text{L/P})$. Furthermore $N_L$ is then used to convert lipid concentration in molar to particle number density which can used to implement the model on absolute scale.
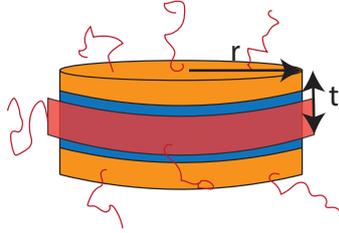


Figure S9: Schematic of the disc model with Gaussian random coils. The orange and blue shells represent lipid headgroups and tailgroups, respectively. The red hollow cylinder surrounding the tails represents a belt of α-synuclein and the red unfolded structures represent the protein extending from the surface of the disc. Not to scale.

The scattering length density for e.g. the outer discs representing lipid headgroups as well as some α-syn is calculated as:

$$\rho = \frac{(L/P \cdot b_{\text{heads}}) + (f_{\alpha\text{-syn, heads}} \cdot b_{\alpha\text{-syn}})}{(L/P \cdot V_{\text{heads}}) + (f_{\alpha\text{-syn, heads}} \cdot V_{\alpha\text{-syn}})} \tag{2}$$

where $b$ are the scattering lengths, $V$ are the molecular volumes and $f_{\alpha\text{-syn, heads}}$ is the fraction of the α-synuclein monomer that is embedded in the lipid heads.

## Core-shell ribbon

This model calculates the form factor for a rectangular core-shell structure. The scattering length densities are again calculated according to Eqn 2. The outside shell has a thickness, i) $t_{\text{shell}}$, representing the lipid headgroups. The core represents the lipid tailgroups. The model has three free parameters describing the three lengths of the particle; ii) width, $A$, iii) height, $B$ and iv) length, $C$. We also allow a Gaussian distribution of lengths in the width v) $\sigma_A$.
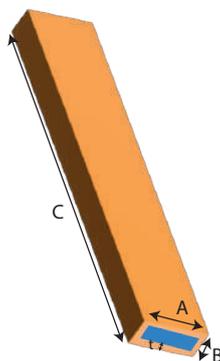
Figure S10: Schematic of the core-shell ribbon model where the orange shell represents the lipid headgroups mixed with some mount of $\alpha$-synuclein and the blue core represents the lipid tailgroups mixed with some $\alpha$-synuclein.

## Core-shell cylinder decorated with Gaussian random coils

This model follows the same philosophy as our 'Disc with Gaussian ranom coils' model, with a smaller inner cylinder representing the lipid tailgroups and a larger outer cylinder representing the lipid headgroups. Again a fraction of $\alpha$-synuclein can be incorporated into the particle by altering the scattering length densities of the cylinders using Eqn. 2. Gaussian random coils decorate the lengths of the cylinder. The free parameters of the model are therefore i) the length of the cylinder, $L$, ii) the radius of the inner cylinder, $R$, iii) the thickness of the shell, $t_{\mathrm{shell}}$, iv) the radius of gyration of the protruding Gaussian random coils, $R_g$, and v) the molar ratio of lipids and bound protein $L/P$.
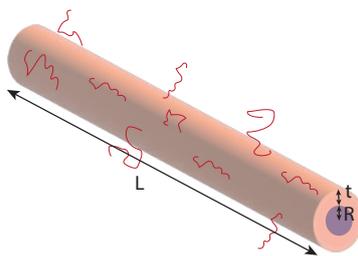


Figure S11: Schematic of the core-shell cylinder with Gaussian random coils. The orange shell represents the lipid headgroups mixed with some mount of $\alpha$-synuclein and the blue core represents the lipid tailgroups mixed with some $\alpha$-synuclein. The red represent the protein extending from the surface of the structure.

10

## Model implementation

All of the form factors used are available in the literature[10]. The models were implemented in $WillItFit$[11]. The models are implemented on absolute scale by exploiting the experimental lipid concentrations, as well as calculating scattering lengths of each model component through its corresponding molecular composition, as listed in Table S1. Molecular volumes are forced to remain within ten percent of the pre-estimated values. Aside from those previously described, three further free parameters are associated with each model: a scaling factor, $S$, a constant background contribution, $b$, and a term accounting for interface roughness in the samples, $R$.[12] Therefore our models are calculated following:

$$I(q) = S \cdot R \cdot n \cdot \langle P(q) \rangle_\Omega + b \tag{3}$$

where $n$ is the particle number density of the sample and $\langle P(q) \rangle_\Omega$ is the form factor term including the usual $\Delta\rho^2 V^2$ terms and averaged over all possible orientations.

The fraction of each molecule of $\alpha$-synuclein inserted into the layer of lipid tails of the outer leaflet of the bilayer was fixed to 10% in line with the previously determined 14 residues.[13] 30% of the protein is assumed to be in the Gaussian random coil formation to match the portion of the C-terminal thought to remain intrinsically disordered post-binding,[14] which leaves 60% to be incorporated into the lipid headgroups of the outer leaflet. The dimensions of the protein proved to be difficult to determine with our modelling scheme and hence the radius of gyration, $R_{g\ \alpha\text{-syn}}$, of the C-terminal part that does not interact with the lipid bilayer, was fixed to 18Å throughout the study, as calculated by Kohn's power law relationship for 44 amino acids.[15] For the "Disc with Gaussian random coils" model $w_{belt}$ was fixed to the diameter of an $\alpha$-helix, 12Å.

Table S1: Chemical compositions, X-ray scattering lengths and neutron scattering lengths for the components required in the modelling.
∗Calculated using an average mass density of proteins 1.35 cm$^2$ g$^{-1}$.[16]
†.[17]
‡[18].

| Component | Chemical composition | X-ray scattering length [cm] | Neutron scattering length [cm] | Estimated volume [Å$^3$] |
|---|---|---|---|---|
| Solvent | $H_2O$ | $2.82 \cdot 10^{-12}$ | - | 30 |
| Solvent | $D_2O$ | - | $1.92 \cdot 10^{-12}$ | 30 |
| $\alpha$-synuclein | $C_{627}H_{1012}N_{166}O_{216}S_4$ | $2.18 \cdot 10^{-9}$ | - | 17800* |
| PS headgroups | $C_8H_{11}NO_{10}PNa$ | $4.85 \cdot 10^{-11}$ | $8.82 \cdot 10^{-12}$ | 244† |
| DM tailgroups | $C_{26}H_{54}$ | $5.92 \cdot 10^{-11}$ | $-2.91 \cdot 10^{-12}$ | 734† |
| DMPS total | $C_{34}H_{65}NO_{10}PNa$ | $2.08 \cdot 10^{-10}$ | $5.91 \cdot 10^{-12}$ | 978† |
| DL tailgroups | $C_{22}H_{46}$ | $5.02 \cdot 10^{-11}$ | $-2.58 \cdot 10^{-12}$ | 619 ‡ |
| DLPS total | $C_{30}H_{57}NO_{10}PNa$ | $9.87 \cdot 10^{-11}$ | $6.24 \cdot 10^{-12}$ | 912 |
| d-$\alpha$-synuclein | $C_{627}D_{782}H_{230}N_{166}O_{216}S_4$ | - | $1.13 \cdot 10^{-9}$ | 17800* |

## Volumes and scattering lengths

## Data processing

Pair-distance ($p(r)$) distributions were obtained by the Indirect Fourier Transform method using the online program *BayesApp* available at https://genapp.rocks.[19,20] SAXS data were re-binned to lie evenly on a log-scale.

# Supplementary Results - Analyses of the SANS and DLS data

## The break up of DMPS and DLPS vesicles and planar bilayer structures into small particles upon $\alpha$-synuclein binding is confirmed by contrast-matching SANS measurements

Similar to the SAXS data presented in the main text, the SANS data clearly report a structural reorganisation of the initial lipid structures once $\alpha$-synuclein is added. Under protein-matching conditions, i.e. 100% $D_2O$, we observed that the forward scattering intensity of the lipids decreased by more than an order of magnitude upon binding of the protein (Figure S12A and B) which can be interpreted as a break down into particles of decreased volume. The low-$q$ trends change from a slope of $q^{-2}$, indicative of planar structures such as vesicles or discs, to a gentler slope of $q^{-1}$, indicative of rod-like particles. The $p(r)$-distribution for pure DLPS (insets of Figure S12 A: purple) appears characteristic of well-formed liposomes[21], whereas for DMPS the $p(r)$-distribution appear to indicate polydisperse globular particles but the underlying structure is less clear. This mirrors our observations from our SAXS experiments (Figures 1, 2). The $p(r)$-distributions from the lipid:protein mixtures (Figure S12A and B insets: green) however, diverge from the SAXS since they are clearly characteristic of rod-like particles, featuring a sharp peak around $r = 50$ Å.

The polydisperse Three-shell vesicle model, described above, provides an excellent fit to the data from pure DLPS (Figure S12 A). The model fit requires a free scale parameter, increasing the theoretical scattering intensity by a factor 3, indicating there may be some errors with the concentration calculation. SANS is less sensitive to small changes in molecular volume and hence the volumes of the lipid headgroups, lipid tailgroups and $\alpha$-synuclein are fixed at their pre-estimated values (Table S1) throughout the SANS analysis. We note, however, that akin to the SAXS modeling in Figure S4, we were unable to capture the full
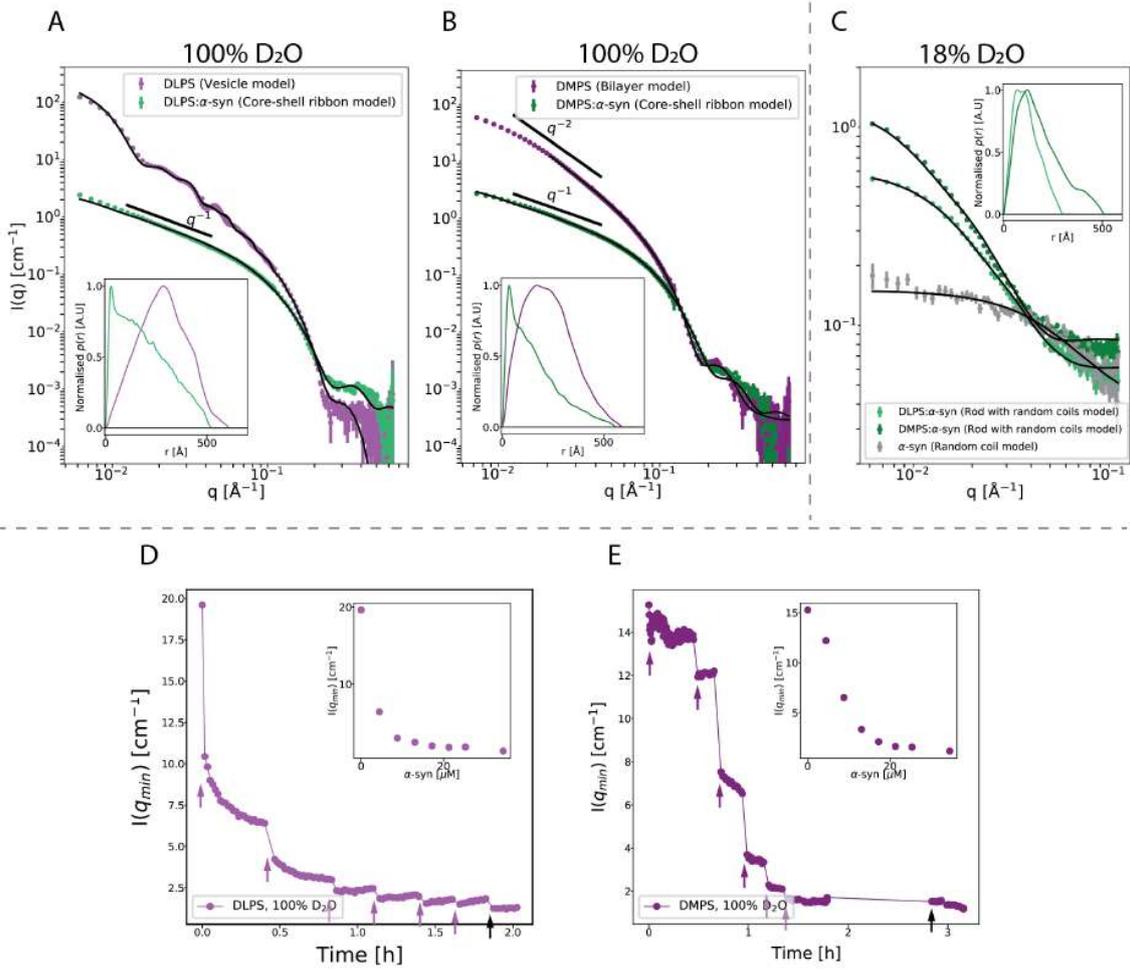
Figure S12: Structural characterisation of DMPS and DLPS model membranes upon binding of $\alpha$-synuclein using SANS. (A,B) Neutron scattering function of 3 mM DLPS or DMPS meodel membranes (purples) and mixtures of model membranes and 100 $\mu$M deuterated $\alpha$-synuclein (greens) measured at 30°C under protein contrast matching conditions, 100% $D_2O$. (C) Neutron scattering function of 100 $\mu$M $\alpha$-synuclein (grey) and mixtures of DLPS or DMPS model membranes and $\alpha$-synuclein (greens) measured under lipid matching conditions, 18% $D_2O$. The insets show the $p(r)$-distribution profiles generated from the SANS data. (D), (E). Change in the intensity at low q of the neutron scattering function ($I_0$), of DMPS (D) or DLPS (E) model membranes (1 mM lipid) measured under protein contrast matching conditions after addition of 4.4 $\mu$M (purple arrows) or 10 $\mu$M (black arrows) $\alpha$-synuclein. Insets: Change in the fraction of protein bound to DMPS (D) or DLPS (E) model membranes with increasing $\alpha$-synuclein calculated from change in SLD. The solution conditions in these experiments were phosphate buffer pH 6.5 and 30 °C.

$q$-range of SANS data from pure DMPS as either vesicles or discs. We therefore again focus our modeling on the range $q = 0.01$ Å$^{-1}$ and upwards, and the planar bilayer model to

14

investigate properties of the bilayer rather than the whole structure. We speculate that the data contain a heterogeneous population of lipid aggregates but with a conserved bilayer structure. The results are presented in Tables S2 and S3.

We use a core-shell ribbon model to describe the lipid:$\alpha$-synuclein co-structures as elongated particles with a rectangular cross-sectional area, as described in the section above. Since $\alpha$-synuclein is matched out under these experimental conditions we do not need to include the protruding Gaussian random coils in the model; however, the protein is implicitly included in calculations of the scattering length density of both shells, which will vary depending on whether $\alpha$-synuclein is bound or not. The model provides excellent fits to the data (Figure S12 A and B) and suggests the formation of rods with lengths between at least 500 and 650 Å and a distribution of widths. Although it is most probable that there is a distribution of sizes of particles, polydispersity in the length of the ribbon is not visible in our data. The ribbon model captures the local intensity maximum at $q = 0.2$ Å$^{-1}$ well, suggesting it provides a good description of the bilayer structure which seems to remain present during the reorganisation of the lipids. The refined parameters imply slight asymmetry of the lipid structures, yielding 24 Å and 29 Å for the average cross-sectional width for DLPS and DMPS, respectively, compared to an average height of 39 Å for both types of lipids. We note that we also tried to fit this data with a cylindrical model which appears the same at low-$q$ values but performed much worse at high-$q$.

Table S2: Structural parameters refined from the SANS data in Figure S12 A (purple) using the Three-shell vesicle model. The model was calculated on absolute scale.

| | Parameter | DLPS |
|---|---|---|
| | Scale | $3.37 \pm 0.03$ |
| | Radius [Å] | $201 \pm 19$ |
| | $\sigma_{\text{Radius}}$ | $0.11 \pm 0.001$ |
| | $t_{\text{lipid}}$ [Å] | $15.7 \pm 0.1$ |
| | $\sigma_{\text{t}}$ | $0.11 \pm 0.03$ |
| | Background [cm$^{-1}$] | $0.00007 \pm 0.0003$ |
| | $\chi^2$ | 211 |

Table S3: Structural parameters refined from the SANS data in Figure S12 B (purple) using the Planar bilayer model. The fit was restricted to q > 0.015 Å$^{-1}$ where intensity arising from bilayer features dominates. The model was calculated on absolute scale.
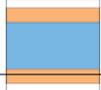
| Parameter | DMPS |
|---|---|
| Scale | 3.64 ± 0.05 |
| $t_{\text{lipid}}$ [Å] | 20.4 ± 0.2 |
| $\sigma_{\text{t}}$ | 0.16 ± 0.03 |
| Background [cm$^{-1}$] | 0.00003 ± 0.0004 |
| $\chi^2$ | 9.88 |

Table S4: Structural parameters refined from the SANS data in Figure S12 (green) using a core-shell ribbon model. *Parameters fixed during refinement.
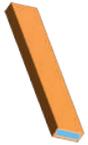
| Parameter | DLPS + d-$\alpha$-synuclein | DMPS + d-$\alpha$-syn |
|---|---|---|
| Scale | 1.00* | 1.00* |
| $L/P$ | 30* | 30* |
| Shell thickness [Å] | 2.37 ± 2.21 | 5.78 ± 3.02 |
| Width [Å] | 39.0 ± 4.7 | 39.2 ± 3.6 |
| Height [Å] | 24.4 ± 2.2 | 28.8 ± 2.7 |
| Length [Å] | 659 ± 160 | 539 ± 98 |
| $\sigma_{\text{width}}$ | 0.33 ± 0.12 | 0.33 ± 1e-11 |
| Background [cm$^{-1}$] | 0.0002 ± 0.0004 | 0.0003 ± 0.0004 |
| $\chi^2$ | 22.7 | 8.84 |

Table S5: Structural parameters refined from the SANS data in Figure S12 C (green) using a core-shell rod with random coils model. *Parameters fixed during refinement.

| Parameter | DLPS + d-$\alpha$-synuclein | DMPS + d-$\alpha$-synuclein |
|---|---|---|
| Scale | 1.00* | 1.00* |
| $L/P$ | 30* | 30* |
| Shell thickness [Å] | 4.67 ± 0.19 | 5.40 ± 0.17 |
| Radius [Å] | 26.7 ± 2.9 | 34.4 ± 2.1 |
| Length [Å] | 278 ± 74 | 465 ± 162 |
| $R_{\text{g}, \alpha\text{-syn}}$ | 18* | 18* |
| Background [cm$^{-1}$] | 0.069 ± 0.007 | 0.092 ± 0.006 |
| $\chi^2$ | 1.71 | 5.70 |
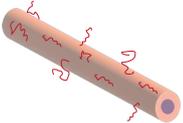
The neutron scattering data from deuterated $\alpha$-synuclein (d-$\alpha$-synuclein) alone in 18% D$_2$O can be well described with a simple Gaussian random coil model where the only free parameter is the radius of gyration (Figure S12 C). We refine an $R_g$ of 30.6 ± 6.9 Å in good agreement with Kohn's estimation of 37 Å for 140 residues.[15] After addition of DMPS or DLPS model membranes to d-$\alpha$-synuclein, we observed an increase in the scattering intensity at low-q values (q ∼ 0.06 Å$^{-1}$) suggesting that the apparent size of the protein particles increases upon binding to the lipids (Figure S12 C). The $p(r)$-distribution of the data from mixtures of DMPS and d-$\alpha$-synuclein in 18% D$_2$O resemble large elongated structures but with a broader peak around r=150 Å as compared to the 100% D$_2$O data. This indicates a

16

larger or less well-defined cross-section which could be explained by a corona of $\alpha$-synuclein tails extending radially from the surface of the rod-like lipid particles. $p(r)$-distributions from DLPS:d-$\alpha$-synuclein in 18% $D_2O$ seem to represent shorter particles, also with a broader distribution of cross-sectional areas.

In order to model the lipid:d-$\alpha$-synuclein structures contained in the 18% $D_2O$ data, the contribution from the flexible part of protein must be taken into account. We built an analytical model representing a core-shell rod decorated with Gaussian random coils. The core and shell of the rods are assigned scattering length densities of lipid tailgroups and headgroups mixed with d-$\alpha$-synuclein, respectively. In order to reduce the number of free parameters and avoid overfitting the model to these noisier data, we choose a monodisperse cylindrical rod rather than a ribbon. The differences in scattering profiles arising from cylindrical rods versus rectangular rods would be mostly indistinguishable within the limited $q$-range of the data in Figure S12 C.

This model provides excellent fits to the data presented in Figure S12 C. The refined parameters are listed in Table S5. The refined values for the shell thickness and cylinder inner radius are in agreement with the dimensions refined from the data collected in 100% $D_2O$ (Table S4). The average cylinder lengths, however, are shorter in comparison at 278 Åfor DLPS-d-$\alpha$-synuclein. This could reflect some clustering or ordered arrangement of $\alpha$-synuclein on the surface of the lipid particles.

In order to investigate the interaction in more detail, we performed a titration of the model membranes with d-$\alpha$-synuclein under protein matching conditions, i.e. 100% $D_2O$ (Figure S12 D,E), and we found that the decrease in forward scattering intensity was gradual with more added protein leading to a further decrease of the scattering intensity. Interestingly the observed decrease in signal intensity as a function of $\alpha$-synuclein concentration was steeper than what would be predicted if the fraction of maximal binding was linearly related to the amount of bound protein (insets in Figure S12D and E). To understand this behaviour, we need to consider the cooperativity in protein binding to the membrane.[22] Two extreme

scenarios can be envisioned: when protein molecules are introduced into an excess of lipids, they can either be distributed equally among the available lipid vesicles, or they can saturate some vesicles while leaving others unoccupied. The first scenario corresponds to the total absence of cooperativity, the second to full cooperativity. It has recently been reported that $\alpha$-synuclein binding to lipids can be highly cooperative, under conditions where the vesicles remain largely intact.[22] The results of the titration experiments, in which we measured how the SANS scattering signal at low q-values decreased as a function of added protein, suggest however, that the interaction is less cooperative in our system. Given that the individual vesicles contribute independently to the scattering signal, their successive individual disruption under conditions of full cooperativity should lead to a linear decrease in signal intensity as a function of protein concentration. However, we observe a stronger than linear/proportional disruptive effect, suggesting that vesicles are only partly saturated, and that partial saturation of a vesicle can lead to a higher than proportional loss in signal intensity. In other words, if a given vesicle is to 50% covered in protein, its scattering intensity will have dropped to below 50% of the initial value.

Different degrees of cooperativity in different $\alpha$-synuclein lipid systems can be explained by the relative strength of the protein-lipid and protein-protein interactions, which may be shifted in the current study due to the use of different lipid systems and different buffer conditions.

Next, we investigated the structural change of the initial lipid structures due to $\alpha$-synuclein binding by acquiring SANS data of the DMPS:$\alpha$-synuclein mixtures at different temperatures ranging from 15 to 30 °C under both contrast-matching conditions (Figure S13 A, B). Under protein-matching conditions, i.e. 100% $D_2O$, we observed a decrease in the intensity at low $q$ in the scattering profile of the mixture $\alpha$-synuclein-DMPS with increasing temperature (Figure S13A). This transition was found to occur at a temperature of around 23 °C, a value matching that of the melting transition observed in the SAXS (Figure 3 A, inset) and DSC measurements[23] (Figure S2 A). Such a change in the SANS data was not observed

18
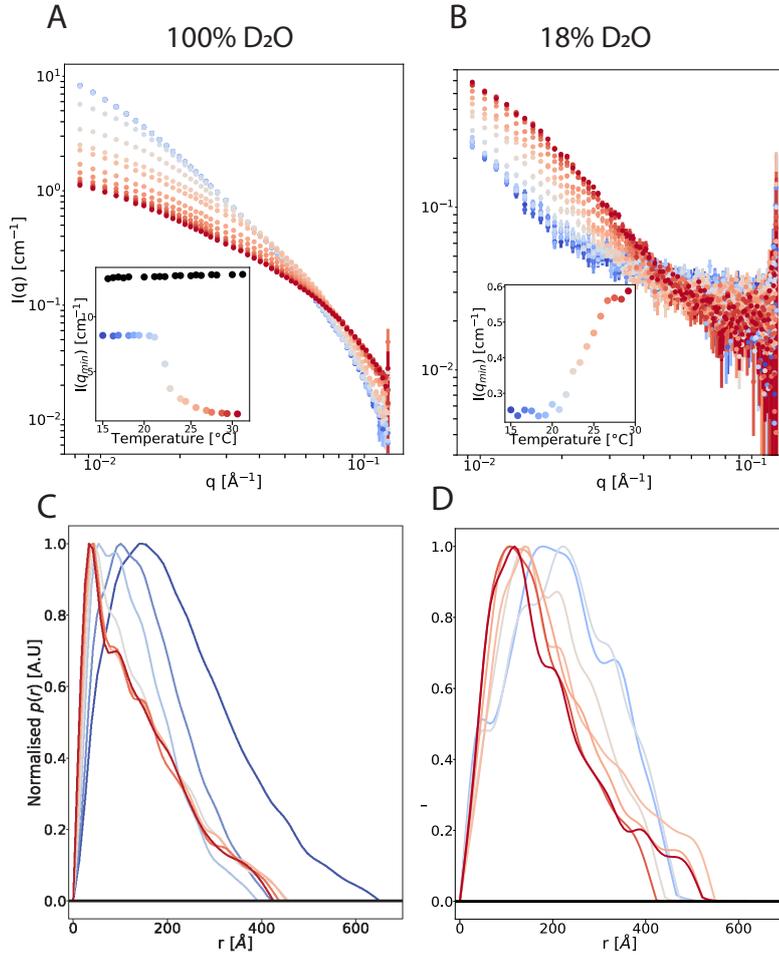
Figure S13: Structural characterisation of the $\alpha$-synuclein:DMPS system at increasing temperature using SANS. Neutron scattering function of 3 mM DMPS and 100 $\mu$M $\alpha$-synuclein measured under (A) protein matching conditions, 100% $D_2O$, or (B) lipid matching conditions, 18% $D_2O$. Insets: Change in scattering function at $q = 0.0073$ Å$^{-1}$ as a function of temperature. Black: pure DMPS (SANS profiles not shown). (C) and (D) $p(r)$-distribution generated from the data demonstrating the system is undergoing structural changes.

for the pure DMPS model membranes with increasing temperatures (Figure S14 and inset to Figure S13 A), suggesting no structural re-arrangement of the lipids within the length scales corresponding to this range of $q$-values and similar to the observation by SAXS. Under lipid-matching conditions, i.e. 18% $D_2O$, we observed an increase in the intensity at low $q$ of the scattering profile of the mixture $\alpha$-synuclein-DMPS with increasing temperature, pointing towards an increase in the apparent size of the protein upon binding to the DMPS

19

model membranes (Figure S13B). This transition also occurs at a temperature around 23 °C, the same value as that of the structural re-arrangement observed for the lipid model membranes (Figure S13B, inset). We once again interpret the decrease in forward scattering of the lipids upon protein binding as a break-up of the initial DMPS structures into smaller particles. We observe a transition of the slope from $q^{-2}$ to $q^{-1}$, indicating the formation of rods as in the static SANS experiments above. The $p(r)$-distribution profiles also reflect a transition from very large polydisperse objects at low temperatures to rod-like particles at higher temperatures ($> 25°$C).

We again attempted to refine form factors from this data series (Figure S15). Due to the
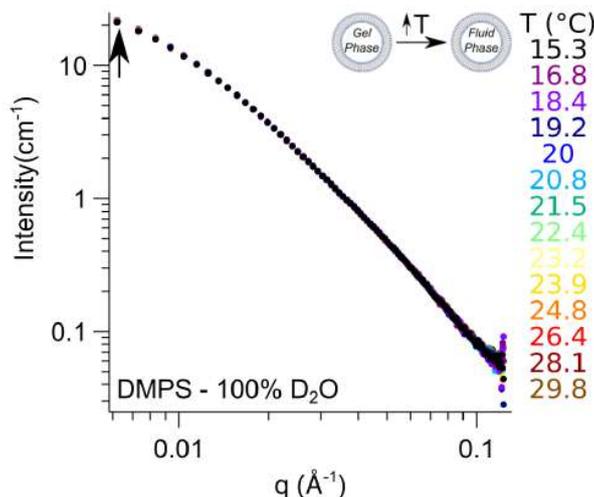


Figure S14: Neutron scattering function of DMPS model membranes (3 mM lipids) measured under protein (100% $D_2O$) contrast matching conditions with increasing temperature.

more simple lipid contrast in SANS, as well as our limited $q$-range, we only considered simple geometrical volumes: either planes or cylinders. The particles here were assigned a "bulk" DMPS scattering length density. We find that the best description for the neutron scattering profile of the DMPS:$\alpha$-synuclein system is as polydisperse planar structures, between 15 and 20 °C, or as rods between 26 and 30 °C with a length of ca. 430 Å and a radius decreasing from 26 to 22 Å (Figure S15 and Tables S6 and S7). We note that we can initially fit the model on absolute scale for the pure DMPS sample. Throughout the rest of the data series,

20

however, the scale is a free fitting parameter. As the temperature is increased the refined scale decreases, reflecting some potential variation in SLD of the particle as d-$\alpha$-synuclein binds to the lipids reducing the overall contrast.

Our analysis suggests that the observed transition of the vesicles to rod-like structures upon $\alpha$-synuclein binding might pass through disc-like particles, specifically at 21 °C where cylinders of radius 130 Å and height 34 Å are refined from the data Figure S15). Around the transition temperature, around 23 °C, the data cannot be described adequately with either model, presumably due to the data containing a multi-modal distribution of structures.
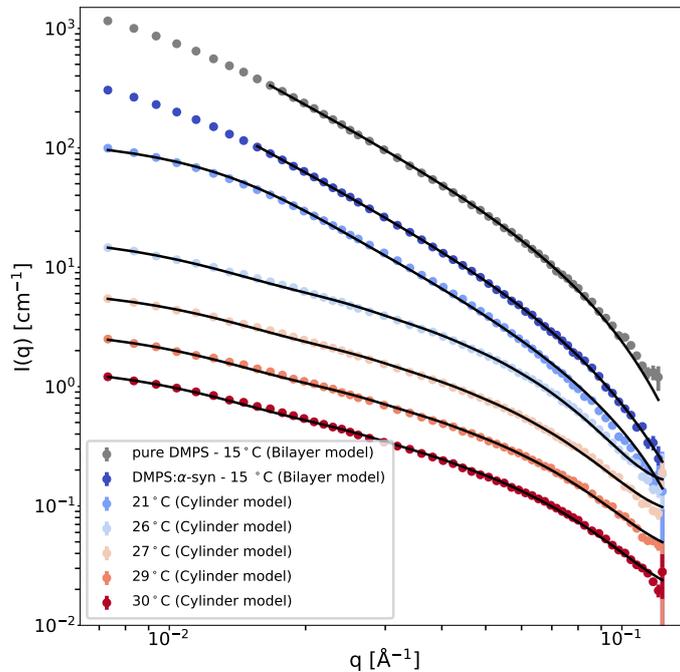


Figure S15: Structural modelling of the SANS data showing the evolution in the phospholipid particles with increasing temperature and hence binding of $\alpha$-synuclein. For ease of viewing, the data are scaled by $2^n$ where $n$ is the profile number, so the bottom scattering profile (red, 30°C) has n = 0 and remains on absolute scale, and the topmost profile (black) has n = 6. The refined structural parameters are listed in Tables S6 and S7.

The results of these temperature ramp experiments agree well with our initial SANS titration experiments (Figure S12 D,E), whereby protein and lipid were mixed at a temperature

Table S6: Structural parameters refined from the SANS data in Figure S15 using a simple planar model. *Parameter fixed during refinement.

| Parameter | DMPS, 15°C | DMPS + $\alpha$-syn, 15°C |
|---|---|---|
| Scale | 1* | 0.58 ± 0.04 |
| $t_{lipid}$ [Å] | 18.6 ± 0.3 | 17.3 ± 0.2 |
| $\sigma_t$ | 0.31 ± 0.07 | 0.30 ± 0.002 |
| $\chi^2$ | 3.81 | 1.77 |

Table S7: Structural parameters refined from the SANS data in Figure S15 using a simple cylinder model model. - means the parameter was not required to during refinement from that data set.

| Parameter | 21°C | 26°C | 27°C | 29 °C | 30 °C |
|---|---|---|---|---|---|
| Scale | 0.59± 0.06 | 0.43 ± 0.01 | 0.42 ± 0.01 | 0.42 ± 0.01 | 0.42 ± 0.01 |
| Radius [Å] | 134 ± 2 | 25.7 ± 0.3 | 22.9 ± 0.3 | 21.6 ± 0.4 | 21.4 ± 0.4 |
| $\sigma_{Radius}$ | 0.33 ± 0.04 | - | - | - | - |
| Length [Å] | 34.4 ± 3.85 | 451 ± 49 | 421 ± 45 | 431 ± 51 | 432± 52 |
| $\chi^2$ | 23.9 | 20.8 | 14.2 | 11.3 | 8.5 |

where the binding immediately induces melting of the lipid. In both scenarios the addition of the protein to the outer layer of the bilayer induces a decrease in protein-matched lipid SANS intensity at low q values and an increase in lipid-matched protein SANS intensity at low q values. Our modeling of the corresponding SAXS data shows that the interaction between $\alpha$-synuclein and DMPS planar bilayer structures leads to a break-up of the lipid structures into significantly smaller particles that are best modelled as discs. The SANS data are in qualitative agreement with this conclusion, and show a decrease of the size of the protein-matched planar bilayer structures and an increase of the apparent size of $\alpha$-synuclein from that of a single protein molecule towards that of a larger lipid-protein particle. However, in the case of the SANS data, the modelling suggests that the particles are best described as short, thin rods. This discrepancy between these models will be discussed below.

## The reversible break up of DMPS and DLPS vesicles into small particles upon $\alpha$-synuclein binding is confirmed by dynamic light scattering measurements

We also performed DLS experiments in microcapillaries and we subjected different mixtures of $\alpha$-synuclein and DLPS/DMPS lipid structures to thermal scanning at two different rates

22

(0.5 and 5°C/min) from 15 to 35°C, and back. Figure S16 shows representative data for 1 mM DLPS in the presence of 95 $\mu$M (excess) and 9.5 $\mu$M (sub-stoichiometric) $\alpha$-synuclein, acquired at scan rates of 0.5 °C/min and 5 °C/min. Similarly, Figure S17 shows representative data for 1 mM DMPS in the presence of 95 $\mu$M (excess) and 9.5 $\mu$M (sub-stoichiometric) $\alpha$-synuclein, acquired at scan rates of 0.5 °C/min and 5 °C/min. The results of these DLS experiments, in which also the absolute light scattering intensity was measured, can be summarised as follows. The addition of $\alpha$-synuclein to DLPS vesicles leads to a decrease in size and scattering intensity at both protein concentrations, with the strongest effect at the highest concentration. Only at the higher concentration, a clear evolution with increasing temperature is observed, whereby the particles appear to become gradually smaller. This process is not reversible upon cooling back down to 15°C, because even at this lower temperature limit attainable in this experiment, the lipid is in the fluid phase if protein is bound. In the case of DMPS, the particles in the protein-lipid mixture at temperatures below the melting temperature are bigger than pure lipids (reflecting the size of the vesicles with a protein corona), but upon heating both the size and scattering intensity decrease significantly and sharply. It is worth noting here that the size range meaured by DLS for the DMPS is very similar to that of DLPS and hence indicative that vesicles form a significant fraction of the structures. However, together with the SAXS and SANS data, a picture of more complex composition, comprising both vesicles and large bilayer structures, emerges for DMPS. The transition into smaller structures observed for DMPS upon heating is in large parts reversible (more so at the lower protein concentration), but displays significant hysteresis. The change in size and scattering intensity induced by protein binding is more pronounced for DLPS compared to DMPS. In the case of DMPS, a clear scan rate dependence can be observed, but only during the cooling step. This is consistent with stopped-flow SAXS data that showed that the initial binding step is approximately diffusion limited. The detachment of protein induced by cooling down, on the other hand, appears to be significantly slower and a clear difference can be seen between 0.5 °C/min and 5 °C/min. The overall conclusions from the

DLS experiments are thus in excellent agreement with, and provides a model-free confirmation of, the results obtained from the modeling of our SAXS and SANS data. The finding that in the case of DMPS a protein concentration (9.5 $\mu$M) three-fold below the nominal saturation concentration leads to an effect on particle size and scattering intensity of very similar magnitude compared to a concentration ca. 3-fold above the saturation concentration (95 $\mu$M) provides additional evidence for the low degree of cooperativity of the system. If the binding were highly cooperative, in the case of 9.5 $\mu$M, only about 1/3 of the vesicles should be disrupted and hence one would at most expect a drop in scattering intensity by 1/3. However, the observed decrease is of the order of 4-5, suggesting that most, if not all, vesicles are significantly disrupted and that is only possible if the protein is distributed approximately equally over all vesicles, in turn suggesting a low degree of cooperativity.
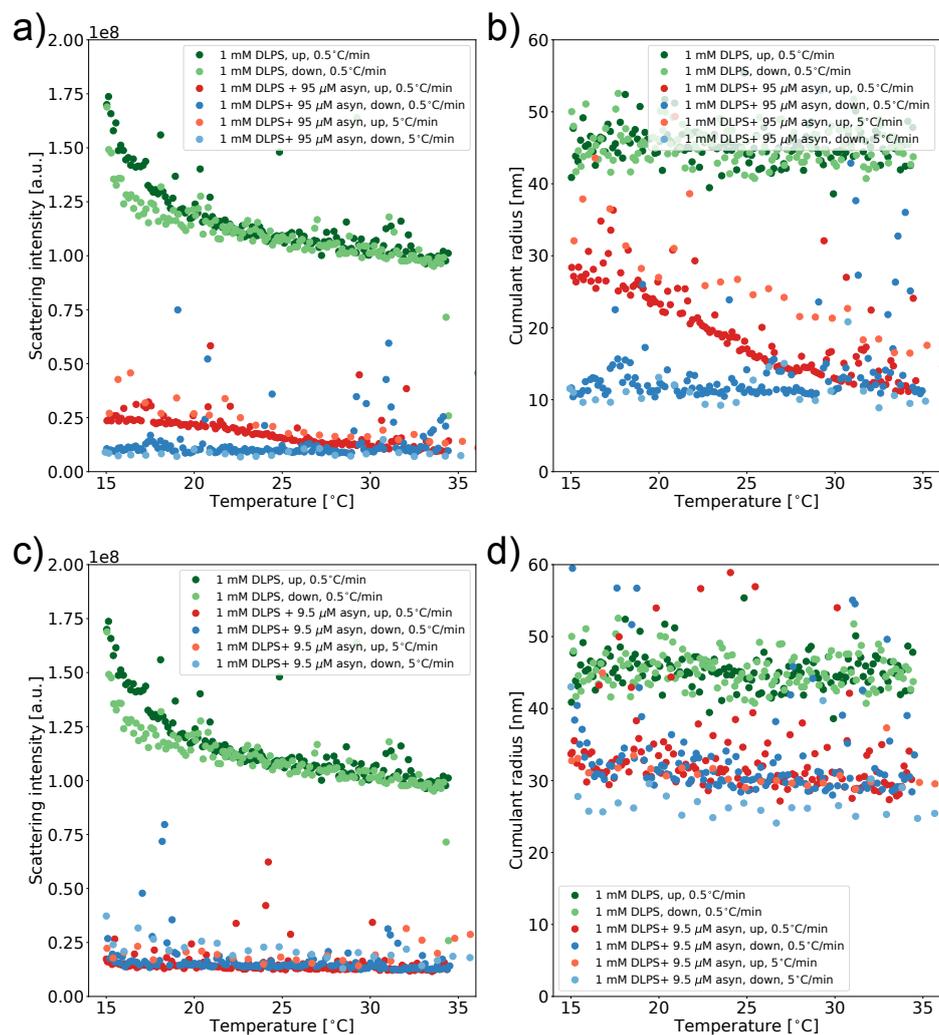
Figure S16: Temperature-dependent DLS experiments of DLPS vesicles in the absence or presence of different concentrations of $\alpha$-synuclein. DLPS (1 mM ) vesicles were incubated in the absence (green) or the presence of 95 $\mu$M (a,b) or 9.5 $\mu$M (c,d) $\alpha$-synuclein at increasing (dark colours) or decreasing (light colours) temperature and scan rates of 0.5°C / min and 5°C / min. The scattering intensity (a,c) and cumulant radius (b,d) are shown.
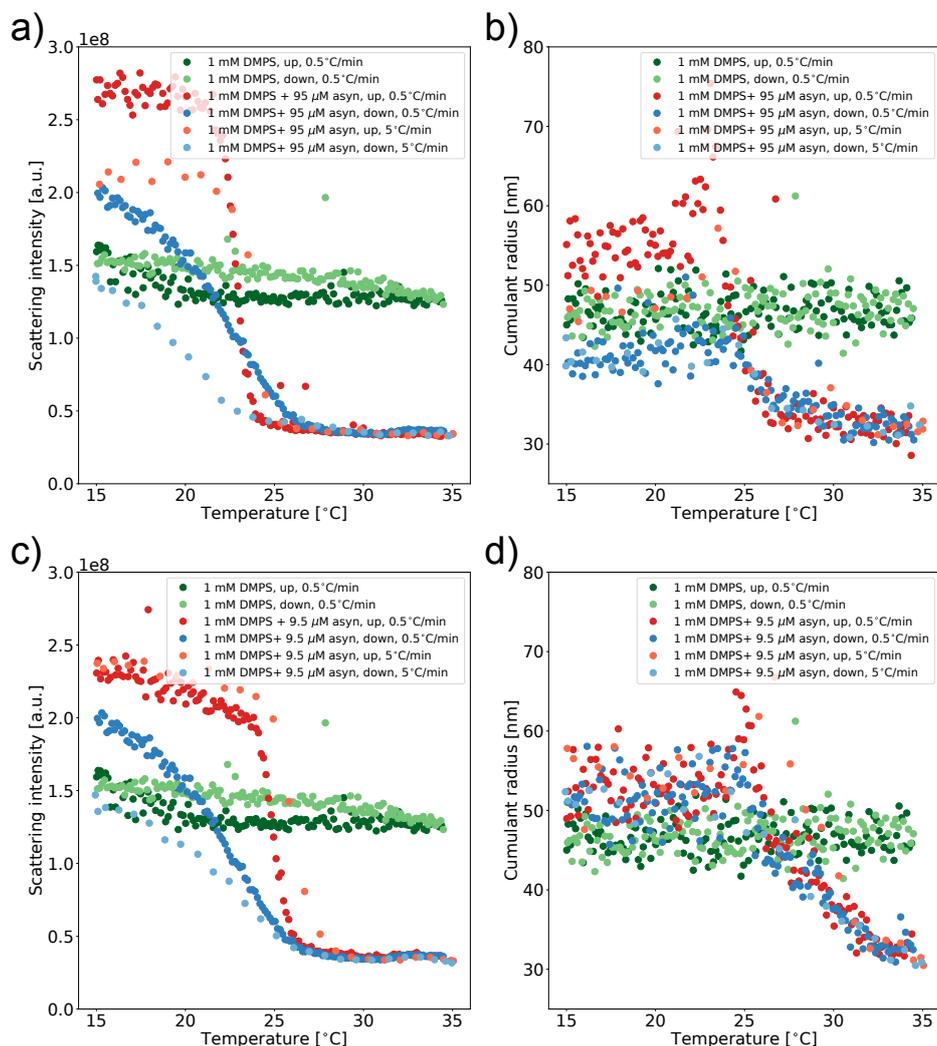
Figure S17: Temperature-dependent DLS experiments of DMPS vesicles in the absence or presence of different concentrations of $\alpha$-synuclein. DMPS (1 mM) vesicles were incubated in the absence (green) or the presence of 95 $\mu$M (a,b) or 9.5 $\mu$M (c,d) $\alpha$-synuclein at increasing (dark colours) or decreasing (light colours) temperature and scan rates of 0.5°C/min and 5°C/min. The scattering intensity (a,c) and cumulant radius (b,d) are shown.

# References

(1) Nallet, F.; Laversanne, R.; Roux, D. Modelling X-ray or neutron scattering spectra of lyotropic lamellar phases: interplay between form and structure factors. *Journal de Physique II* **1993**, *3*, 487–502.

26

(2) Berghausen, J.; Zipfel, J.; Lindner, P.; Richtering, W. Influence of water-soluble polymers on the shear-induced structure formation in lyotropic lamellar phases. *The Journal of Physical Chemistry B* **2001**, *105*, 11081–11088.

(3) Debye, P. Molecular-weight determination by light scattering. *The Journal of Physical Chemistry* **1947**, *51*, 18–32.

(4) Skar-Gislinge, N.; Simonsen, J. B.; Mortensen, K.; Feidenhans'l, R.; Sligar, S. G.; Lindberg Møller, B.; Bjørnholm, T.; Arleth, L. Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: Casting the roles of the lipids and the protein. *Journal of the American Chemical Society* **2010**, *132*, 13713–13722.

(5) Skar-Gislinge, N.; Arleth, L. Small-angle scattering from phospholipid nanodiscs: Derivation and refinement of a molecular constrained analytical model form factor. *Physical Chemistry Chemical Physics* **2011**, *13*, 3161–3170.

(6) Cholak, E.; Bucciarelli, S.; Bugge, K.; Vestergaard, N. T. J. B.; Arleth, L.; Kragelund, B. B.; Langkilde, A. E. Distinct $\alpha$-Synuclein:Lipid Co-Structure Complexes Affect Amyloid Nucleation Through Fibril Mimetic Behavior. *Biochemistry* **2019**, *58*, 5052–5065.

(7) Pedersen, J. S.; Gerstenberg, M. C. Scattering form factor of block copolymer micelles. *Macromolecules* **1996**, *29*, 1363–1365.

(8) Pedersen, J. S. Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores. *Journal of Applied Crystallography* **2000**, *33*, 637–640.

(9) Arleth, L.; Ashok, B.; Onyuksel, H.; Thiyagarajan, P.; Jacob, J.; Hjelm, R. P. Detailed structure of hairy mixed micelles formed by phosphatidylcholine and PEGylated phospholipids in aqueous media. *Langmuir* **2005**, *21*, 3279–3290.

27

(10) Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: Modeling and least-squares fitting. *Adv. Colloid Interface Sci.* **1997**, *70*, 171–210.

(11) Pedersen, M. C.; Arleth, L.; Mortensen, K. WillItFit: a framework for fitting of constrained models to small-angle scattering data. *Journal of Applied Crystallography* **2013**, *46*, 1894–1898.

(12) Als-Nielsen, J.; McMorrow, D. *Elements of Modern X-ray Physics: Second Edition*; John Wiley and Sons, 2011.

(13) Cholak, E.; Bugge, K.; Khondker, A.; Gauger, K.; Pedraz-Cuesta, E.; Pedersen, M. E.; Bucciarelli, S.; Vestergaard, B.; Pedersen, S. F.; Rheinstädter, M. C.; Langkilde, A. E.; Kragelund, B. B. Avidity within the N-terminal anchor drives $\alpha$-synuclein membrane interaction and insertion. *FASEB Journal* **2020**, *34*, 7462–7482.

(14) Fusco, G.; Simone, A. D.; Gopinath, T.; Vostrikov, V.; Vendruscolo, M.; Dobson, C. M.; Veglia, G. Direct observation of the three regions in $\alpha$-synuclein that determine its membrane-bound behaviour. *Nat Commun* **2014**, *5*, 3827.

(15) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R., et al. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences* **2004**, *101*, 12491–12496.

(16) Mylonas, E.; Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of applied crystallography* **2007**, *40*, s245–s249.

(17) Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F. Structure and Fluctuations of Charged Phosphatidylserine Bilayers in the Absence of Salt. *Biophysical Journal* **2004**, *86*, 1574–1586.

(18) Armen, R. S.; Uitto, O. D.; Feller, S. E. Phospholipid component volumes: Determination and application to bilayer structure calculations. *Biophys. J.* **1998**, *75*, 734–744.

(19) Hansen, S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *Journal of Applied Crystallography* **2000**, *33*, 1415–1421.

(20) Savelyev, A.; Brookes, E. GenApp: Extensible tool for rapid generation of web and native GUI applications. *Future Generation Computer Systems* **2019**, *94*, 929–936.

(21) Glatter, O. In *Scatt. Methods their Appl. Colloid Interface Sci.*; Glatter, O., Ed.; Elsevier, 2018; pp 33–74.

(22) Makasewicz, K.; Wennmalm, S.; Stenqvist, B.; Fornasier, M.; Andersson, A.; Jonsson, P.; Linse, S.; Sparr, E. Cooperativity of $\alpha$-Synuclein Binding to Lipid Membranes. *ACS Chemical Neuroscience* **2021**, *12*, 2099–2109.

(23) Galvagnion, C.; Brown, J. W. P.; Ouberai, M. M.; Flagmeier, P.; Vendruscolo, M.; Buell, A. K.; Sparr, E.; Dobson, C. M. Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of $\alpha$-synuclein. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 7065–7070.

# 11

## Φ-VALUE ANALYSIS OF PI3-KINASE SH3 AMYLOID FIBRIL ELONGATION

CONTRIBUTIONS

The project was conceptualised and led by Alexander Buell. Jacob Aunstrup Larsen conducted all of the experiments and wrote most of the manuscript. I performed the MD simulations and FoldX calculations, and wrote these sections of the manuscript. I was supervised by Kresten Lindorff-Larsen.

# Φ-value analysis of PI3-kinase SH3 amyloid fibril elongation

Jacob Aunstrup Larsen[1], Abigail Barclay[2,3], Nicola Vettore[1], Louise K. Klausen[1], Lena Mangels[4], Kresten Lindorff-Larsen[3], and Alexander K. Buell[1]

[1]Section for Protein Chemistry & Enzyme Technology, Technical University of Denmark
[2]Niels Bohr Institute, University of Copenhagen
[3]Structural Biology and NMR Laboratory, University of Copenhagen
[4]Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf

**Abstract**

Amyloid fibrils are thermodynamically highly stable protein assemblies playing an important role in several neurodegenerative and systemic diseases, such as Parkinson's disease and Alzheimer's disease. While the kinetics of amyloid growth are well known and structural information of amyloid is becoming increasingly abundant, the transition state between the monomeric species and the fibrillar state remains elusive. Here we characterize effects on the thermodynamic stability and reaction barrier height of elongation of fifteen point-mutated species of PI3K-SH3 amyloid fibrils. We calculate Φ-values and bias MD-simulations to provide the first view of the TS of amyloid elongation and identify the primary contacts between the monomer and fibril end. The presented framework provides a strong foundation for thorough mechanistic study of the amyloid growth mechanisms.

1

# 1    Introduction

In the last three decades amyloid fibrils and the misfolding reaction have received increasing interest as the amyloid conformation of proteins is related to several neurodegenerative diseases and systemic pathologies[1–3]. Misfolding of proteins into the amyloid conformation has been treated largely as a kinetic problem[4]. However, the aggregation kinetics of protein variants does not necessarily correlate with the onset of disease, as in the case of Parkinson's disease[5]. Early attempts to develop pharmaceutical aggregation inhibitors have been largely unsuccessful, a contributing factor is poor understanding of the interaction between the inhibitor compound and the various conformational states populated during the aggregation reaction[1,6]. It is evident that increased understanding of the microscopic steps of the aggregation reaction would be of significant value in tackling the challenge of diseases associated with amyloid structures[7]. Evolution in structural biology techniques, such as electron cryo-microscopy, have allowed determination of amyloid fibril structures at an atomic level[8]. However, information of the final amyloid structure does not intrinsically identify residues governing the misfolding reaction. Such information may be accessed by a combination of structure-guided mutagenesis experiments and simulation.

The elongation of amyloid is well defined. In general, this involves a fibril end where a soluble monomer can attach by changing its conformation and reaching a more favourable energetic state[2]. Both the change in the energetic state and in conformation resemble closely the classical folding process, with the difference of being a bi-molecular process, which leads to a concentration dependence of the kinetics of the reaction. Several lines of evidence illuminate the rate limiting ensemble as a desolvated state with positive activation entropy[9] and is similarly dependent on ionic strength as the fully misfolded amyloid state[10]. It is clear that the conformations defining the energy barriers of amyloid elongation can best be described as spatially close and structurally similar the fibril end, or product-like. However, whether the amyloid elongation reaction is best described by an unbiased energy-landscape of kinetic traps[11] or by a single energy-barrier with a defined transition state[12] (TS) remains a point of discussion.

Φ-value analysis is a powerful tool for investigation and mapping of folding pathways[13]. The analysis infers structural information of the transition state ensamble (TSE) from changes in energetics upon mutagenesis (change of free energy of activation of folding,

2

$\Delta\Delta G_{\ddagger-D}$, and change of equilibrium free energy of folding, $\Delta\Delta G_{N-D}$)[14–16] and can obtain atomic level information[17]. The approach has been extended from protein folding to protein binding[18,19] and can reasonably be extended to protein misfolding. Indeed, Wang and Fersht have applied $\Phi$-value analysis in an effort to investigate the conformational rearrangement of the p53 monomer along with fibril incorporation[20]. $\Phi$-value analysis seemingly prove suitable for investigation of the amyloid elongation reaction of proteins, where no monomeric unfolding is necessary.

Here we apply $\Phi$-value analysis to protein misfolding, the amyloid elongation reaction, for the first time. We characterize PI3K-SH3 amyloid fibril elongation kinetics and thermodynamics at an unprecedented level of detail. We present a strategy to study the influence of individual residues in the elongation reaction, through both kinetic measurements of change of the activation energy of elongation upon mutagenesis ($\Delta\Delta G_{\ddagger-U}$) and equilibrium measurements, characterizing the free energy of polymerization ($\Delta G_{A-U}$) and the change upon mutagenesis ($\Delta\Delta G_{A-U}$). Guided by structure, we study fifteen mutations covering the entire fibril fold and combine experimental $\Phi$-values with Molecular Dynamics (MD) simulations to provide the first estimation of an amyloid elongation TSE. Our work demonstrates a high-resolution mechanistic study, enabled by recent breakthroughs in amyloid structure determination, and provides an advance towards a complete mechanistic understanding of protein misfolding.
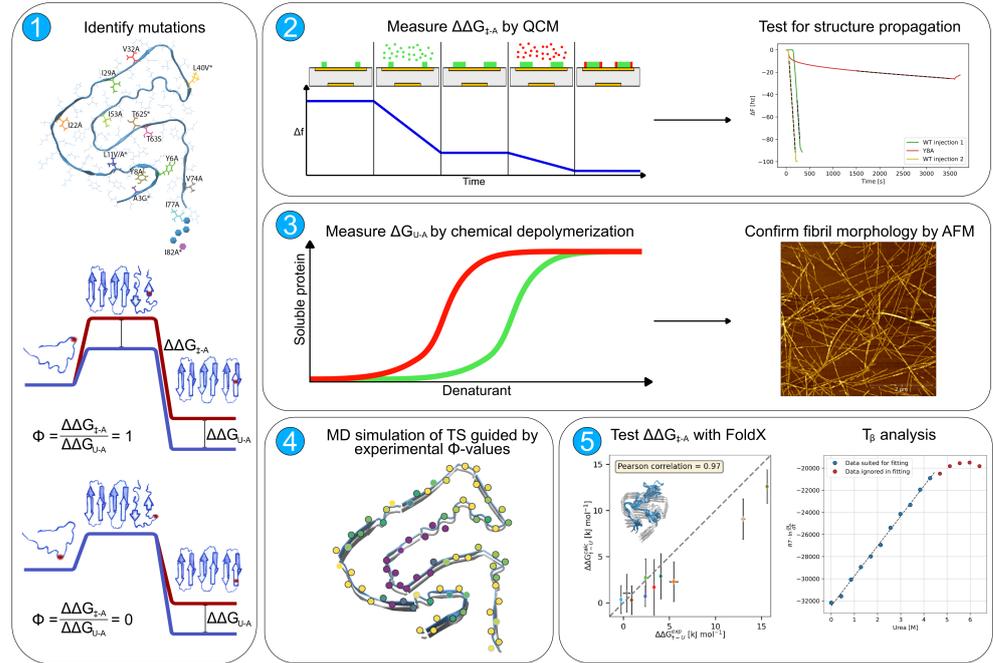
Figure 1: Graphical abstract of Φ-value analysis of amyloid elongation. 1: Identify aliphatic residues viable for hydrophobic deletion, such that the all structural elements of the fibril structure are probed. 2: Relative elongation rates of all mutants are investigated by Quartz Crystal Microbalance, by incubating immobilized WT fibrils to both WT and mutant monomer. WT monomer is demonstrated to maintain elongation rates after mutant growth, demonstrating fibril structure is propagated. 3: A denaturation series is performed on all mutants capable of propagating the WT fibril structure. Mutant protein is seeded with WT fibrils to induce the correct fibril structure upon the mutant protein. The denaturation series are globally fitted with the cooperative polymerization model. Mutant fibrils are imaged by AFM to demonstrate that fibril morphology is maintained. 4: Calculated Φ-values are used to bias MD-simulations of the TS. If certain regions of the fibril are lacking experimental Φ-values, return to step 1 and select new residues for mutations to increase resolution. 5: Use FoldX to calculate the $\Delta\Delta G_{\ddagger-A}$ from the simulated TSE structures to check if the simulation captures the experimental kinetic data. Measure depolymerization kinetics of the WT fibrils from a denaturation series, calculate the m-value and the Tanford value. Compare this to the simulated TSE to further comfirm the validity of the simulated structure.

## 2    Methods

### 2.1    Selection of residues of interest

We identified fourteen residues viable for conservative hydrophobic deletions: residues A3, Y6, Y8, L11, I22, I29, V32, L40, I53, T62, T63, V74, I77, I82, which cover the entire amyloid fold[21]. We performed fifteen hydrophobic deletions: A3G, Y6A, Y8A, L11V, L11A, I22A, I29A, V32A, L40V, I53A, T62S, T63S, V74A, I77A, I82A.

## 2.2   Proteins

We expressed, produced and purified the WT and mutants of the bovine PI3K-SH3 domain according to a previously published protocol[10]. We dissolved lyophilized protein in the required buffer before use.

## 2.3   Fibril preparation

We formed PI3K-SH3 WT fibrils from protein solutions in 10 mM glycine hydrochloride pH 2 at a final concentration of $200\mu$M. The solutions were stirred at $42°$C overnight and the presence of fibrils was verified through atomic force microscopy (AFM). These fibrils were sonicated with a VialTweeter-sonotrode (Hielscher, Teltow, Germany) for 10 s at 100% amplitude in a total volume of $800\mu$l. We used the so obtained seed fibrils, at 5% (equivalent monomer mass), in a fresh solution of PI3K-SH3 WT whose concentration ranges between 200 and $400\mu$M. We left solutions overnight at room temperature, and investigated the day after for fibrils through AFM and supernatant absorbance measurements. We used the fibrils so prepared both for attachment to QCM sensors and for further seeding. We prepared fibrils for equilibrium denaturation experiments starting from solutions of WT and mutants at $200\mu$M. We mixed identical volumes of these solutions with a volume of WT fibrils prepared as above described, to reach a final seeding concentration of 5% in equivalent monomer mass. We verified the fibrils by AFM.

## 2.4   AFM imaging

We diluted the fibril samples in 10 mM glycine-hydrochloride, pH 2 to a concentration of $5\mu$M and pipetted $10\mu$l of the solutions on freshly cleaved mica substrates. After 10 min of incubation, we washed the mica extensively with milliQ water and dried them under a nitrogen gas flush. We imaged WT, I22A, I29A, I53A, I77A, I82A fibrils in tapping mode on a Bruker Multimode 8 (Billerica, Massachusetts, USA) using OMCL-AC160TS cantilevers (Shinjuku, Tokyo, Japan) and imaged WT, A3G, Y6A, Y8A, L11V, L11A, V32A, L40V, T62S, T63S, V74A fibrils in tapping mode on a DriveAFM (Nanosurf, Liestal, Switzerland) using PPP-NCLAuD cantilevers (Nanosensors, Neuchatel, Switzerland).

## 2.5    Depolymerization experiments

We mixed fibril samples with increasing volumes of urea stock solutions (see below) and decreasing volumes of buffer to create denaturation series of linearly increasing urea concentration. We measured fibrils at a final protein concentration of $40\mu$M for I22A, I29A, I53A, I77A, I82A and one WT reference and 30 $\mu$M for all other mutants as well as four WT references. In order to maintain a solution pH of 2 constant throughout the whole denaturation series, we prepared an 8 M urea stock solution by dissolving 12 g of urea (SigmaAldrich) in 16 ml of concentrated buffer with added HCl, to yield a final concentration of 10 mM glycine hydrochloride and 16 mM HCl. The HCl is necessary, as the urea is weakly basic. We confirmed the urea concentration by refractive index measurements with an Ataga R-5000 refractometer. In order to keep the ionic strength constant in all samples, we added NaCl to the 10 mM glycine HCl buffer to a final concentration of 16 mM. We allowed samples to equilibrate for two weeks at 25 °C, which has been demonstrated to be sufficient for these conditions[10]. We distributed mutants into four experimental sets, each containing an internal WT reference as a control of reproducability. The I22A, I29A, I53, I77A and I82A mutants are experimental set 1. We measured the I77A, I82A mutants and the WT reference in a 26 sample series from 0 to 6.4 M urea, while I22A, I29A and I53 mutants are measured in a 24 sample series from 0 to 6 M urea. L11V, V32A and L40V mutants are experimental set 2. The T62S, T63S and V74A mutants are experimental set 3. The A3G, Y6A, Y8A and L11A, mutants are experimental set 4. We measured the mutants and the WT sample of set 2, 3 and 4 in a 16 sample series from 0 to 6.4 M urea.

## 2.6    Fluorescence measurements

We recorded the fluorescence spectra of experimental set 1 on a Tecan M1000pro instrument using a Greiner UV-transparent 96 well plates. The wells were filled with 140 $\mu$l of solution and the temperature was maintained constant at 25 °C. The fluorescence spectra were measured by top reading, exciting at 290 nm (5 nm bandwidth), while the emission was recorded between 300 nm and 420 nm (5 nm bandwidth). A blank spectrum (buffer) resulting from the average of 10 different spectra is subtracted before analysis. In order to determine the relative populations of soluble vs. fibrillar protein, we computed the ratios of the fluorescence intensities at 340 and 310 nm for each spectrum. We measured the flu-

6

orescence of experimental set 2, 3 and 4 in two systems. Firstly, on a FLUOstar Omega instrument using a Hellma Analytics quartz microplate. The wells were filled with 45 $\mu$l of solution and the temperature was maintained constant at 25 °C. The fluorescence intensities were measured by top reading, exciting at 280 nm, while the emission was recorded at 310, 320, 330, 340, 350 and 360 nm (10 nm bandwidth). A blank spectrum (buffer) is subtracted before analysis. In order to determine the relative populations of soluble vs. fibrillar protein, we computed the ratios of the fluorescence intensities at 340 and 310 nm for each spectrum. Secondly, we also measured experimental set 2, 3 and 4 in a Prometheus Panta instrument using Prometheus NT.48 Series nanoDSF Grade High Sensitivity Capillaries. The fluorescence emission was recorded at 330 and 350 nm, exciting at 280 nm. In order to determine the relative populations of soluble vs. fibrillar protein, we computed the ratios of the fluorescence intensities at 350 and 330 nm for each spectrum, assuming an insignificant background spectrum from the buffer.

## 2.7   Analysis of chemical depolymerization

We fitted depolymerization data to the cooperative polymerization model, which is an extension of the isodesmic model[22,23]. The final fitting equation takes the form:

$$x_{tot} = \sigma^{-1} \left( \frac{(\sigma x)^{n+1}(n\sigma x - 1)}{(\sigma x - 1)^2} + \frac{\sigma x}{(\sigma x - 1)^2} \right) - \sigma^{n-1} \left( \frac{x^{n+1}(nx - n - 1)}{(x - 1)^2} \right) \quad (1)$$

Where n is the maximal aggregation number for a nucleus-like species. The fit quality is relatively insensitive to exact choices of n below values of $10^{10}$, so we set n to 3, where we obtained the best results. $\sigma$ is defined by $K_n = \sigma K_e$, where $K_e$ is the equilibrium constant between monomers and species of higher aggregation number than n, and $K_n$ is the equilibrium constant between monomers and species of aggregation number up to n. $x = K_e[M]$ and $x_{\text{tot}}$ are the normalized soluble and total monomer concentration: $x_{\text{tot}} = K_e[M]_{\text{tot}}$. $[M]_{\text{tot}}$ is the total protein concentration, $K_e = K_e([D]) = \exp\left((\Delta G_0^{\text{Buff}} + m[D])/(RT)\right)$. R is the ideal gas constant, T is the absolute temperature, [D] is the independent variable, denaturant concentration, $\Delta G_0^{\text{Buff}}$ is the free energy of polymerization in the absence of denaturant for each variant and $\Delta G([D]) = \Delta G_0^{\text{Buff}} + m[D]$. m and $\sigma$ are shared fitting parameters, while $\Delta G_0^{\text{Buff}}$ is the independent fitting parameter. The fitting equation is solved numerically through the use of the *least_square* module of the *scipy.optimize* python library. The change in free energy of

7

polymerization is calculated as: $\Delta\Delta G_{A-U} = \Delta G_{0_{\text{Mutant}}}^{\text{Buff}} - \Delta G_{0_{\text{WT}}}^{\text{Buff}}$. We fit the data glob-
ally, with WT references sharing $\Delta G_0^{\text{Buff}}$. We assume independent measurements when
propagating errors.

## 2.8    Elongation measurements

We measured the elongation rate of PI3K-SH3 fibrils through immobilization of fibrils on
a Quartz Crystal Microbalance (QCM) sensor and subsequent incubation with monomer
solution[12]. To immobilize the fibrils on the sensor, chemical modification is necessary[24].
We pelleted the fibrils, washed and resuspended them in 10 mM HCl twice, before mix-
ing at a final concentration of 50 $\mu$M in buffer (10 mM HCl) with EDC (1-ethyl-3-
(3-dimethylaminopropyl)carbodiimide hydrochloride) (1M) and cystamine hydrochloride
(0.5mg/ml). After pelleting and washing the chemically modified fibrils, we sonicated
them in an Eppendorf tube in a volume of ca. 500 $\mu$l with a Bandelin Sonoplus using a
MS72 probe (10% amplitude, 15 s, 1 s"on", 2 s"off"). We incubated gold sensors (Biolin
Scientific, Gothenburg, Sweden) with the above-mentioned solution overnight in a 100%
humidity environment. The measurements were performed with a QSense Pro (Biolin
Scientific, Gothenburg,Sweden) by measuring the elongation rate as change in resonant
frequency over time. With the temperature set at 25 °C, we injected at least 3 cell vol-
umes (60 $\mu$l) of the monomer solutions and the measurement lasted until a stable slope
was reached. To obtain the relative rates, we injected the monomeric mutant solutions
in different sensor chambers after a WT injection, the latter being used as normalization
reference. We measured the rate as the slope of the third overtone frequency and av-
eraged among the multiple injections. We measured the elongation rate at two protein
concentrations, 5 $\mu$M and 20 $\mu$M.

## 2.9    Calculation of change in energy of activation of elongation

The elongation rate kinetics is described by the master equation[25]:

$$\frac{\partial \text{M}([\text{D}])}{\partial t} = 2P(x_{tot} \cdot k_{\text{on}} - k_{\text{off}}) \tag{2}$$

Where $P$ is the number of fibril ends, $x_{tot}$ is the monomer concentration, $k_{\text{on}}$ is the on
rate and $k_{\text{off}}$ is the off-rate. When normalized by the number of fibril ends and at high
monomer concentrations, where $x_{tot} \cdot k_{\text{on}} \gg k_{\text{off}}$ the elongation kinetics reduces to:

$$\frac{\partial M([D])}{\partial t} = x_{tot} \cdot k \tag{3}$$

Where $k$ is the rate of elongation.

We calculated the change of energy of activation of elongation upon mutation from the relative elongation rate according to:

$$k = \tau \exp\left(-\frac{\Delta G_{\ddagger-U}}{RT}\right) \rightarrow \Delta G_{\ddagger-U} = -RT \ln\left(\frac{k}{\tau}\right) \tag{4}$$

Where $\Delta G_{\ddagger-U}$ is the energy of activation of elongation. It is reasonable to assume similar kinetic prefactors ($\tau$) for elongation of WT protein and mutants, when fibril structure is maintained, as all mutations investigated here are conservative hydrophobic deletions.

$$\Delta\Delta G_{\ddagger-U} = \Delta G_{\ddagger-U}^{mut} - \Delta G_{\ddagger-U}^{WT} = -RT \ln\left(\frac{k_{mut} \cdot \tau}{k_{WT} \cdot \tau}\right) = -RT \ln\left(\frac{k_{mut}}{k_{WT}}\right) \tag{5}$$

Standard errors are calculated assuming each measurement is independent.

## 2.10   Φ-value calculation

We calculate Φ-values as:

$$\Phi = \frac{\Delta\Delta G_{\ddagger-U}}{\Delta\Delta G_{A-U}} \tag{6}$$

Errors are propagated assuming all measurements are independent.

## 2.11   Depolymerization kinetics

We measured the unfolding rate of SH3 WT fibrils in a Prometheus Panta instrument using Prometheus NT.48 Series nanoDSF Grade High Sensitivity Capillaries. We record the fluorescence emission at 330 and 350 nm, exciting at 280 nm. We perform an Isothermal unfolding for WT fibils and WT monomer at 16 linearly spaced urea concentrations, from 0 to 6.4 M urea, for 10 hours at 30 $\mu$M protein. Depolymerization buffers were prepared as described above. We ensured consistent fibril concentrations and homogeneous fibril lengths by extending the sonication step to 30 seconds.

We calculate the fraction of soluble monomer from the fluorescence ratio, assuming fully fibrillar samples at $t = 0$ with 0 M urea and no fibrils present for monomer samples at

respective urea concentrations.

We calculated monomer release is by linear regression of the first 15 to 400 minutes, depending on urea concentration.

## 2.12   Tandford $\beta$ calculation

The m-value to the depolymerization barrier can be accessed from the rate of monomer release with the following assumptions:

Fibril elongation, nucleation, fragmentation and secondary nucleation are insignificant compared to disassociation on the fitted timescale. The number of fibril ends are similar in all samples and constant over the fitted timescale. The kinetic master equation[25] then reduces to:

$$-\frac{\partial \mathrm{M}([\mathrm{D}])}{\partial t} = 2P{\cdot}k_{\mathrm{off}}([\mathrm{D}]) \rightarrow \frac{\partial \mathrm{c}([\mathrm{D}])}{\partial t} = 2P{\cdot}k_{\mathrm{off}}([\mathrm{D}]) = 2P{\cdot}\tau \exp\left(\frac{\Delta G^{\ddagger\mathrm{Buff}} + \mathrm{m}[\mathrm{D}]}{\mathrm{RT}}\right) \quad (7)$$

Where M is the fibril concentration, c is monomer concentration, D is denaturant concentration, P is number of fibril ends, $k_{\mathrm{off}}$ is the off-rate, $\tau$ is the kinetic pre-factor and $\Delta G^{\ddagger\mathrm{Buff}}$ is the kinetic barrier at 0 M denaturant

It follows that:

$$\mathrm{RT} \cdot \ln\frac{\partial \mathrm{c}([D])}{\partial t} = m[\mathrm{D}] + c, \quad c = \Delta G^{\ddagger\mathrm{Buff}} + \mathrm{RT} \cdot \ln\left(2P \cdot \tau\right) \quad (8)$$

We calculated the m-value of depolymerization from linear regression of the fitted unfolding rates according to equation 8. We did not include samples where more than 25% monomer was released prior to the measurement, as the fundamental assumption of constant number of fibrils may no longer hold true.

We calculated the Tanford$\beta$-value as[17]:

$$T_\beta = \frac{\mathrm{m}_{A-\ddagger}}{\mathrm{m}_{A-U}} \quad (9)$$

Where $\mathrm{m}_{A-\ddagger}$ is the m-value for depolymerization kinetics and $\mathrm{m}_{A-U}$ is the m-value of chemical stability, obtained from the cooperative polymerization model fitted to the equilibrium measurements.

## 2.13 Molecular Dynamics Simulation

We genereted the structural ensembles of PI3K-SH3 fibrils and the corresponding TS by MD simulations with GROMACS 2021[26,27] and PLUMED2 software[28]. We made use of the amber99sb-disp forcefield[29] and a TIP4P-D water model[30]. For the starting conformation we extracted four stacked subunits (chains A, C, E and G) from the published atomic structure PDB: 6r4r[21], (Figure S3), thereby choosing to simulate a single protofilament instead of the double protofilament. This allows us to drastically increase conformational sampling. Given the evidence that the single protofilament can exist on its own[21], as well as only having one data point at the filament-interface (V74A) which has a low $\Phi$-value and therefore not predicted to make any contacts in the transition state, this is a sufficient model for investigating the association of a monomer to the existing fibril end.

he structure was protonated in line with the acidic solutions used in the experiments. The structure was solvated, neutralized by chloride ions and energy minimized using steepest descent. Initially, the temperature was incrementally increased from 0 K to 300 K using the Berendsen thermostat[31]. The fibril structure is not expected to be stable with only four subunits. To mimic the increased stability and structural order of long fibrils, we restrained all backbone atom positions with a potential of 100 kJ mol$^{-1}$ nm$^{-2}$ during equilibration. In the production simulations, bonds involving hydrogens were constrained with the LINCS algorithm[32], electrostatics were treated with the particle mesh Ewald scheme[33] with a short-range cutoff of 0.9nm, and the van der Waals interaction was set to 0.9nm. We ran production simulations at 300 K, maintained with the velocity-rescaling thermostat[34] and pressure of 1 atm, maintained with the Parrinello-Rahman barrostat[35].

## 2.14 Molecular Dynamics $\Phi$-value calculation

We followed a procedure where $\Phi$-values are modeled as the fraction of native contacts in the TS for that residue[36], so that for each conformation in the simulation the $\Phi$-value for residue $i$ is calculated with:

$$\Phi_i^{\mathrm{sim}}(t) = \frac{N_i(t)}{N_i^{\mathrm{Native}}} \qquad (10)$$

where $N_i(t)$ is the number of contacts residue $i$ makes at step $t$ and $N_i^{\text{Native}}$ is the number of native contacts for residue $i$. We introduced an energy perturbation to the force field to minimise deviations between $\Phi^{\text{exp}}$ and $\Phi^{\text{sim}}$[37,38]. The force constant in our Φ-value restraint was set to 100 kJ mol$^{-1}\mu$M$^{-1}$. We avoided overfitting to our Φ-values by scanning force constants from 1 to 1000 kJ mol $^{-1}\mu$M$^{-1}$ and selecting the one that gave a $\chi^2_{\text{red}}$ for $\Phi^{\text{calc}}$ closest to 1 (Figure S4). (During this procedure we additionally included a 'computational' error of 0.1 $\mu$M for each Φ-value in line with observations from[39]. This computational error accounts for inaccuracies in the forward model as well as in the native contact approximation (Equation 10). The experimental error for V74A specifically seemed improbably low and therefore the computational error is required so that the $\chi^2_{\text{red}}$ calculation is not dominated by one data point.)

To determine the native contacts of the top subunit, we performed a 160 ns reference state simulation with position restraints on all backbone atoms. We determined native contacts for each residue as heavy side-chain atom pairs within 0.6 nm in at least 75% of the frames, excluding atoms from directly neighboring residues. The set includes atom pairs within the top subunit as well as contacts to the unit directly below.

We included 10 Φ-values from our experimental data set at 20 $\mu$M: Y6A, Y8A, L11A, I22A, I29A, V32A, I53A, T63S, V74A, I77A. I77A has Φ-value = -0.12 which we approximated to be equal to zero. We omitted A3G, L11V, L40V and T62S due to their large uncertainties, and I82A since it does not form part of the rigid $\beta$-sheet core of the amyloid fibril. We applied the Φ-value restraints only to the top unit in the structure (Figure S3) in order to model the association of the monomer to a pre-existing fibril end. For each step of the simulation, native contacts were back-calculated using a rational switching function:

$$s(r) = \frac{1 - (\frac{r-d_0}{r_0})^n}{1 - (\frac{r-d_0}{r_0})^m} \tag{11}$$

with r$_0$=0.85, n=15 and m=30. In order to maintain the rigid fibril structure in the three lower subunits, we applied position restraints to all backbone atoms apart from those in the top subunit. We ran the simulation for a total of 4 $\mu$s and the TS was determined using only the conformations sampled in the last 2 $\mu$s where the RMSD and total fraction of native contacts for the top subunit plateaued (Figure S4).

A 400 ns control simulation was also performed without the Φ-values restraints, with position restraints applied to the three lower subunits. We observed that our set-up is

stable since after discarding the first 200 ns, the RMSD to the native structure is steady $\sim 0.28$ nm. Furthermore, 63 out of 77 residues in the top subunit maintain over 80% of their native contacts on average across the control simulation (Figure S4). There is some dissociation of the top subunit in residues 1, 16, 17 and the disordered loop region between $\beta$-sheets 4 and 5, but the overall structure seems unaffected.

## 3    Results

### 3.1    Mutant fibril preparation

We designed fifteen different mutants of PI3K-SH3 to probe the degree of structure in the TSE. These mutations span all of the seven $\beta$-strands in the fibril structure, the non-core region, as well as the hydrophobic cluster of the interfilament interface. We have demonstrated that the mutations does not alter fibril morphology or interfere with WT elongation propensity of fibrils, thereby confirming that each mutant elongates WT seeds and propagate the seed structure (Supporting Results).

### 3.2    Elongation kinetics

We proped the elongation rates of the mutants by QCM, a surface-based biosensing technique[40]. Seeds of PI3K-SH3 WT fibrils were immobilized on all the sensors and solutions of the WT and the different mutants were brought into contact with the different sensors to measure the elongation rates. We measured the rates of elongation for two different concentrations of monomeric protein, 5 and 20 $\mu$M (Table 1). Since destabilizing mutations can be highly destabilizing in some cases, this approach is essential as the lower monomeric concentration may lead to a saturation concentration ($c_{sat}$) of the fibrils, that is significantly higher than for the WT and might approach 5 $\mu$M. As the measurement relies on relative elongation rates, a similar degree of absolute super-saturation is necessary to produce comparable data. As protein concentration approaches significantly higher concentrations than $c_{sat}$ for both WT and the mutant, an increasingly similar degree of absolute super-saturation is reached. Hence the kinetic measurements at higher protein concentration should be considered the more robust parameter.

This seemingly highlights a distinctive feature of $\Phi$-value analysis of fibril growth as opposed to protein folding, namely its concentration dependence. In protein folding $\Phi$-

13

values are normally concentration independent, whereas for a bi-molecular reaction, such as fibril growth, they will depend on concentration. This is because the forward reaction, i.e. fibril growth, depends on monomer concentration, whereas the back-reaction, i.e. fibril dissociation, does not. Indeed, mutations such as I53A changes the solubility of the monomer with more than four orders of magnitude.

| | $\Delta\Delta G_{\ddagger-U}$ kJ/mol 5 $\mu$M | $\Delta\Delta G_{\ddagger-U}$ kJ/mol 20 $\mu$M | $\Delta\Delta G_{A-U}$ kJ/mol | Φ-value 5 $\mu$M | Φ-value 20 $\mu$M |
|------|------|------|------|------|------|
| A3G | 2.5 ± 0.3 | 2.7 ± 0.2 | 2.41 ± 3.08 | 1.06 ± 1.35 | 1.13 ± 1.45 |
| Y6A | 1.0 ± 0.4 | 2.4 ± 0.1 | 6.13 ± 2.86 | 0.16 ± 0.10 | 0.39 ± 0.19 |
| Y8A | 13.0 ± 0.9 | 13.0 ± 0.2 | 15.40 ± 2.32 | 0.85 ± 0.14 | 0.84 ± 0.13 |
| L11V | -1.9 ± 0.5 | -0.5 | 1.32 ± 3.16 | -1.41 ± 3.40 | -0.35 ± 0.84 |
| L11A | 1.3 ± 0.1 | 2.4 ± 0.1 | 7.28 ± 2.80 | 0.18 ± 0.07 | 0.33 ± 0.13 |
| I22A | 9.9 ± 0.3 | 5.5 ± 0.5 | 13.20 ± 2.47 | 0.75 ± 0.14 | 0.41 ± 0.08 |
| I29A | 9.7 ± 0.2 | 4.1 ± 0.0 | 5.51 ± 2.94 | 1.76 ± 0.94 | 0.74 ± 0.39 |
| V32A | 3.3 ± 0.1 | 3.3 | 7.64 ± 2.77 | 0.43 ± 0.16 | 0.44 ± 0.16 |
| L40V | -0.6 ± 0.1 | -0.9 | -1.27 ± 3.31 | 0.46 ± 1.19 | 0.74 ± 1.92 |
| I53A | 14.7 ± 0.7 | 15.6 | 15.89 ± 2.31 | 0.92 ± 0.14 | 0.98 ± 0.14 |
| T62S | 0.2 ± 0.3 | 1.6 ± 0.2 | 1.96 ± 3.11 | 0.09 ± 0.21 | 0.80 ± 1.28 |
| T63S | -0.2 ± 0.3 | 0.9 ± 0.1 | 3.30 ± 3.03 | -0.09 ± 0.11 | 0.28 ± 0.25 |
| V74A | 0.0 ± 0.2 | 0.4 ± 0.2 | 9.96 ± 2.64 | 0.00 ± 0.02 | 0.04 ± 0.02 |
| I77A | -0.2 ± 0.3 | -0.2 ± 0.1 | 2.02 ± 3.14 | -0.12 ± 0.23 | -0.12 ± 0.20 |
| I82A | -1.1 ± 0.3 | -0.9 ± 0.2 | -2.85 ± 3.44 | 0.40 ± 0.49 | 0.32 ± 0.39 |

Table 1: All relative elongation rates and derived $\Delta\Delta G_{\ddagger-U}$-values. Elongation rates are measured by Quarts Crystal Microbalance, as the slope of the third overtone frequency during incubation of respective monomer concentrations over a sensor of immobilized WT fibrils. Each elongation rate is normalized to a WT injection prior to mutant injections.

Four mutants are particularly kinetically different from the WT; Y8A, I22A, I29A and I53A, which elongate at least an order of magnitude slower than WT. A3G, Y6A, L11A and V32A elongates several times slower than the WT, while elongation of L11V, L40V, I77A and I82A are slightly accelerated. T62S and T63S elongate similarly to the WT under 5 $\mu$M conditions, but are slower at 20 $\mu$M, while V74A elongates as fast as the WT in both measurements (Table S1).
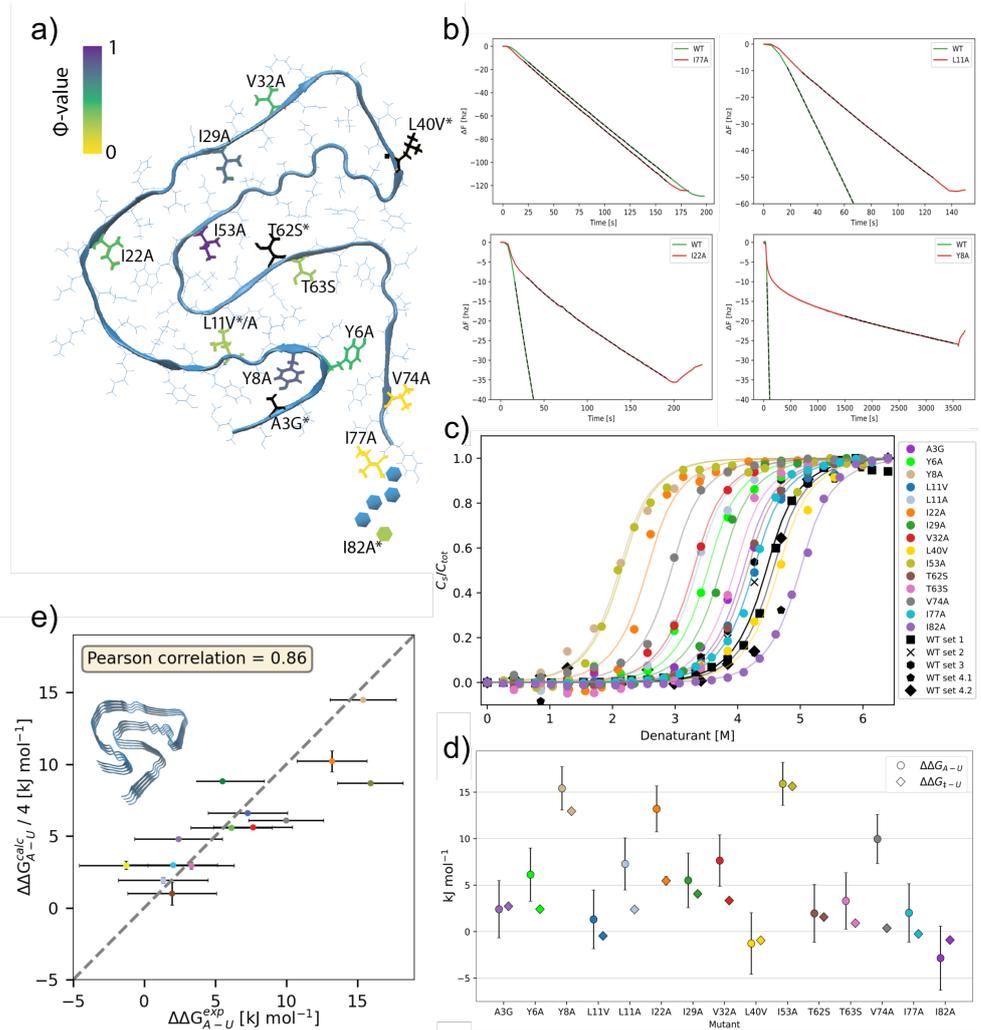
Figure 2: Fibril structure of the PI3K-SH3 fibril, mutations investigated, full thermo-dynamic analysis of mutation and sample data. (a) Top-view of the fibril with the point-mutations highlighted. Residues are color-coded according to their experimentally determined $\Phi$-values. * annotations mark residues, which $\Phi$-value was not included as restraints in the TSE simulations. (b) Sample kinetic data, measured at 20 $\mu$M, WT data in green and mutant data in red. Dashed lines mark data fit for linear regression and rate comparison. (c) Full denaturation series of all mutants fibrils and WT controls. Markers denote measurement points and the cooperative polymerization fit is in solid. (d) $\Delta\Delta G_{A-U}$ values and $\Delta\Delta G_{\ddagger-U}$ values, measured at 20 $\mu$M protein concentration, for all mutants. (c) Calculated $\Phi$-values for all mutants. The values are calculated from the 20 $\mu$M kinetic experiment. The full values can be found in table 1. Mutants, the $\Phi$-values of which have not been used as restraints in the simulation have been marked with a red slash. (e) Comparison of experimental and FoldX-calculated $\Delta\Delta$Gs for the fibril structure, PDB 6r4r. The mutations are modeled in every unit of the fibril. The error bars in the x-axis signify the experimental uncertainty and the error bars in the y-axis signify the standard deviation between repeat calculations.

15

### 3.3    Equilibrium denaturation

We formed WT and mutant fibrils from WT fibril seeds, to ensure a structural coherence between all samples and exposed them to increasing concentrations of urea (Section 2.5). Once the samples were fully equilibrated, we measured the soluble protein fraction from intrinsic fluorescence emission ratios of 340/310 nm or 350/330nm. We have demonstrated that different measurement approaches used here are comparable (Supporting Results) and globally fit the full data-set (Figure 2) using the cooperative polymerization model (Methods) and calculated $\Delta\Delta G_{A-U}$ (Table 1).

We find that eight mutations, Y6A, Y8A, L11A, I22A, I29A, V32A, I53A and V74A are highly destabilizing ($[\Delta\Delta G_{A-U}] > 5$ kJ mol$^{-1}$), and span the full fibril structure (Figure 2). Highly destabilizing mutations, are commonly associated with truncation of large residues in the hydrophobic cores[41], which holds true for the Y6A, Y8, L11, I22, I29A and I53 residues, which are all facing into the hydrophobic core of the amyloid fibril, while V74A is buried in the filament interface. The highly destabilizing effect of V32A is unexpected, as the residue faces into the solvent. We identify destabilizing mutants, which have $[\Delta\Delta G_{A-U}]$ similar to our mean SD value (2.9 kJ mol$^{-1}$), A3G, L11V, T62S, T63S and I77A. These mutations are generally more conservative compared to the more destabilizing mutations, with the exception of I77A. However, since the I77 residue is at the edge of the core structure and facing into solution, a small $[\Delta\Delta G_{A-U}]$ is expected. We find that two mutations have a slightly stabilizing effect on the fibril structure, L40V and I82A. It is striking to compare the V74A and L40V mutation, due to their structural proximity at the hydrophobic filament interface core and similar chemistry, but vastly different effects on the fibril stability. The stabilizing effect of the I82A mutation is notable as I82 is not located within the fibril core sequence (Figure 2) and hence an effect on the stability of the fibril upon aliphatic mutation is unexpected. The mutated residue would be expected to affect the disordered monomeric state as well as the disordered non-core region similarly.

### 3.4    Φ-value calculation

By comparing the effects of the different mutants on the energetic barrier of elongation and the stability of the amyloid fibril, information on the role of the mutated residues on the TS can be obtained. We calculated the Φ-values for both the concentrations used

16

for the kinetic experiments (Table 1). All mutations with $\Delta\Delta_{A-U}$ significantly above the SD produce classically meaningful $\Phi$-values, demonstrating the validity of the $\Phi$-value analysis strategy we have deployed. It is important to note that while I82A may produce a seemingly meaningful $\Phi$-value, since the residue is not within the structured core region, strict interpretation of this $\Phi$-value is not possible.

Mutations within the same $\beta$-strand demonstrate distinctly different $\Phi$-values. It follows, that the protocol is capable of distinguishing effects of mutations at single residue level resolution.

Within our data-set we produce only two extreme $\Phi = 1$ and $\Phi = 0$, the I53A and I74A mutations respectively. The $\Phi$-value of the I74A mutation is in agreement with earlier findings[21].

## 3.5   Molecular Dynamics simulation of the TS

Experimental $\Phi$-values and molecular simulations form a powerful combination for investigating the structure of transition state ensembles (TSE) at atomic-resolution[38], giving insights beyond what experiments can directly. This has been successfully applied to cases of protein folding many times, where specific structural features of the folding nucleus could be determined, including secondary structural elements[38,42,43], key native contacts[36,37,44], as well as non-native interactions which stabilise the TS[45]. More recently $\Phi$-values integrated with simulations have also been used to capture the TS of binding reactions for intrinsically disordered proteins[18,19].

Here we extend the methodology to simulate the TSE of PI3K-SH3 amyloid fibrils, specifically investigating the critical interactions of a free monomer with the pre-existing fibril end upon binding. As described in Section 2.13, ten of our experimental $\Phi$-values are incorporated into the simulation. The experimental $\Phi$-values are approximated as the fraction of native contacts, where each residue's selection of native contacts is derived from the amyloid state. The simulation is then biased to generate an ensemble of conformations that closely match the desired fraction of native contacts in the TS.

The main results from our simulation of the transition state are reported in Figure 3. The $\chi^2$ between the experimental and the calculated $\Phi$-values is close to one, suggesting that the simulated conformations are representative of the TS of amyloid elongation. The calculated $\Phi$-values are mapped onto the fibril structure (Figure 3(e)).
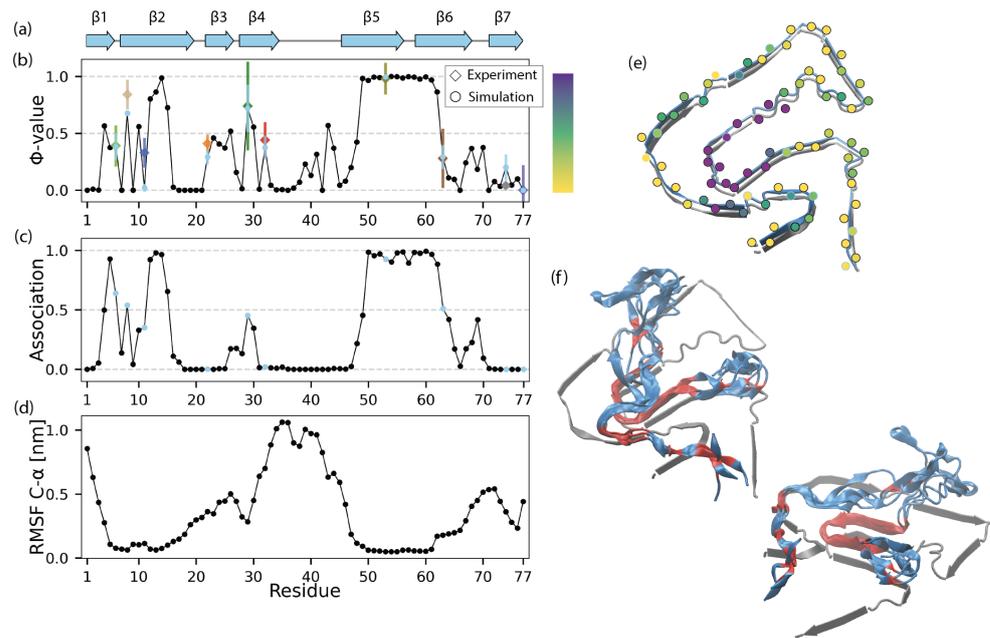
17

Figure 3: Characterisation of the TS. (a) Positions of the seven $\beta$-sheets in the fibril structure (b) The average $\Phi$-value calculated from the simulation for each residue (black circles). Blue circles indicate residues which were restrained by their experimental counterpart, along with their standard deviation across the simulation. Colourful diamonds indicate experimental $\Phi$-values and their associated error. Therefore the $\Phi$-value restraints are met within error in all cases but two (Y8A and L11A). (c) Average fraction of native contacts maintained to the second unit for each residue, so 1 means all inter-chain contacts are maintained and 0 means complete dissociation of the top unit. (d) $C_\alpha$ root-mean-square fluctuation per residue. (e) Mapping of the $\Phi$-values obtained from the simulation on the fibril structure. (f) Two views of a small ensemble of structures in the TS. Red indicates regions that maintain, on average, at least 40% of their native contacts to the chain below. Only the top two units are shown for clarity.

Both the N-terminus and the residues close to the C-terminus have $\Phi$-values close to zero, indicating they do not form many of their native contacts the TS.

Between positions 4 and 12 the $\Phi$-values jump discretely between 0 and 0.5, where tyrosine interactions appear critical since Y6, Y8, Y12 and Y14 maintain contact with the corresponding tyrosines in the unit below, while the alternate residues point away from the fibril. Along the long edge of the fibril, positions 16 to 48, the $\Phi$-values shift smoothly between 0 and 0.5, indicating patches of varying degrees of fibril structure. The calculated $\Phi$-values tend to equal zero in the loops between the $\beta$-sheets (Figure 3(a)).

The most prominent feature of the TS is the 13 residue-stretch between positions 49 and 62 which have $\Phi$-values consistently at 1. Situated at the centre of the fibrillar core, this region sits in the deepest part of the interface due to the staggered arrangement of fibril units. Consequently, it is a region which establishes numerous contacts in the final fibril state. These include both intra-chain (within the top monomer) bonds across $\beta$-sheets 5 and 6, as well as inter-chain interactions which span from the top monomer to the lower three subunits of the fibril and help stabilise the overall inter-twisted pattern (e.g. E52 to L26 in chain 2 and chain 3, [2,3] and I29[2,3]; W55 to E19[2,3], I22[2,3], L24[2,3]; E61 to R9[3,4]). These 13 residues appear to serve as the main anchor to stabilise the incoming monomer and facilitate amyloid elongation.

We investigated which regions of the monomer are attached to the fibril end in the TS (3(c)). Approximately 30% of the monomer is permanently attached, while 50% is completely dissociated, and the remaining 20% make fractional interactions. Positions 50 to 61 constitute the largest contact region to the interface and is present in every conformation of the TSE. In addition residues E4, Y12, D13 and Y14 serve as stable touch-points. The segment running from position 16 to 48 is detached from the fibril end, except for residues I29 and L30, which on average maintain almost 50% of their native contacts to the lower unit. The retention of these contacts keep this segment of residues somewhat in proximity to the fibril end. Furthermore, residue F69 maintains over 40% of its inter-chain native contacts, preventing the C-terminal-end from fully dissociating. These fractional contacts could play a crucial in guiding the monomer to find its proper position on the fibril template. Figure 3(f) shows a small ensemble of structures in the TS. The lower unit is shown in gray and the incoming monomer is coloured. Regions which are in contact with the lower unit are represented in red while regions that are detached are represented in blue. The detached regions are mainly flexible but have some rigid,

19

non-native elements that will be discussed in Section 3.6

By comparing to Figure 3(b) and (c) it is evident that residues in positions 22, 23, 24, 32, and 38 to 44 are dominated by native intra-chain contacts in the TS, since they are detached from the fibril end but have calculated Φ-values greater than zero. Furthermore there is some minor intra-chain interactions occurring among residues 71 to 77.

In Figure 3(d) the $C_\alpha$ root-mean-square fluctuation (RMSF) plotted. Unsurprisingly, the RMSF closely mirrors Figure 3(c) since regions that are detached from the lower unit exhibit higher flexibility and variability compared to the bound regions.

In short, in the TS, residues at positions 49 to 62 are fully structured and form the largest interaction site of the incoming monomer with the fibril-end. Y14 also forms all of its native contacts in the TS. Presumably it is critical for these contacts to form as a nucleation site for the rest of the monomer to sequentially fall into place.

## 3.6   Investigating contacts in the TS

To investigate the critical contacts between the incoming monomer and fibril end, which are formed in the TS, we compared the contact maps of the incoming monomer in our native fibril simulation and our TS simulation (Figure S3(c)). As expected the native state simulation shows a stable fibril structure with little variation across the simulation. The off-diagonal points indicate residues in contact across the turns of the fibril interface. The TS contact map depicts a much more dynamic system, since the majority of the incoming monomer is unbound.

In the native state ensemble, A48 is positioned to make contact with T31. In the TS, A48 interacts strongly with I29 and stabilises the neck of the loop of dissociated residues in positions 30 to 47 (Figure S3). The contact map shows variability in this region, displaying many fractional non-native contacts. The detached regions show a few notable non-native structural features (Figure S3). Firstly, $\beta$-sheet 2 is held on top of $\beta$-sheet 5 due to L26 interactions with the lower unit. Secondly, a small non-native kink appears to be maintained through interactions between Y8 and E19. Finally, the last ten residues turn back on themselves, mainly driven by interactions of Y76 with E65 and R66. These residues also make non-native contacts to residues in the lower chain around E65[2], F42[2] and D44[2]. Non-native TS contacts are not observed in the experimental Φ-values, however, the non-native contacts observed in the simulation could potentially be important

for overcoming the kinetic barriers by stabilising a network of interactions that play a positive role in elongation. Non-native interactions have been shown to accelerate protein folding in various cases[46–48]. The non-native contacts are responsible for maintaining regions not directly connected to the fibril end in close vicinity to it, but the effect of these contacts should be investigated further.

## 3.7   Computational calculations of free energy changes

We investigated how our experimental $\Delta\Delta G$s compare directly with $\Delta\Delta G$s calculated from all-atom structures of the PI3K-SH3 fibril. We used FoldX[49], an energy function which quantifies the important interactions in protein stability, to calculate free energy changes. FoldX has been shown to accurately predict the effect of point mutations on the stability of folded proteins[50,51] as well as transition state structures[39], and was recently used to study fibril structural stability but without comparison to thermodynamic data from experiments[52].

Firstly, as an independent test, we used the cryo-EM structure (PDB 6r4r) and calculated the energy difference between the input structure and the mutant structure for each mutation in Table 1 (excluding I82A which is not present in the structure) to find $\Delta\Delta G^{\mathrm{calc}}_{A-U}$ . The mutation is modeled in all four monomer units of the fibril structure simultaneously. $\Delta\Delta G^{\mathrm{calc}}_{A-U}$ is then divided by the number of units in order to be directly comparable in magnitude with the experimental values. The comparison (Figure 2) shows a good agreement between $\Delta\Delta G^{\mathrm{exp}}_{A-U}$ and $\Delta\Delta G^{\mathrm{calc}}_{A-U}$, with a pearson linear correlation coeffiecient of 0.86. This is inline with the performance of FoldX for thermodynamic calculations of folded proteins[53], demonstrating this could be a very valuable tool for investigating the effect of deletion mutations on amyloid fibril stability in future studies, although it should be tested on a larger data set. This high agreement also acts as a validation of our depolymerization kinetic data.

Next, we followed the procedure outlined in[39] to further analyse and validate the TS structures from our $\Phi$-value-restrained simulation. We used an ensemble of 150 TS structures to calculate $\Delta\Delta G^{\mathrm{calc}}_{\dagger-U}$ using FoldX. The agreement between $\Delta\Delta G^{\mathrm{exp}}_{\dagger-U}$ and $\Delta\Delta G^{\mathrm{calc}}_{\dagger-U}$ is strikingly good with a linear correlation of 0.97 (Figure S5) and shows that modeling $\Delta\Delta G^{\mathrm{calc}}_{\dagger-U}$ as the removal of native state contacts is a reliable approximation for fibril elongation. The TSE structures are consistent with the kinetic data and we could use the

method to examine TS free energies of residues that we do not have experimental data for. Further analysis is shown in the SI.

## 4    Discussion

Here we have introduced the methodology necessary to approach the study of the TS of the amyloid elongation reaction. By investigating fifteen mutations spanning the PI3K-SH3 sequence, we have produced the first glimpse of the spatial role of the residues in the formation of the TS. The approach we present here is focused at the description of the TS of the elongation reaction only, starting from an unstructured monomeric starting state[54–56]. This is opposed to the pioneering work by Wang and Ferhst, who characterizes different rate constants simultaneously, in an effort to investigate the conformational rearrangement of the p53 monomer along with fibril incorporation[20]. To our knowledge, we also present the largest experimental study to date of amyloid thermodynamic stability by chemical denaturation of a single protein and its mutation variants.

In our methodological framework, the kinetic analysis proved to be a robust method to probe the elongation barrier, able to determine $\Delta\Delta G_{\ddagger-U}$ with errors of 0.7 kJ mol$^{-1}$ or less. Within our analysis, we have extrapolated the $\Delta\Delta G_{A-U}$ from a global fit of denaturation series using the cooperative polymerization model. We have maximally constrained the fitting protocol by sharing m and $\sigma$-values across all species, sharing $\Delta\Delta G_{A-U}$-values between all WT samples, allowing $\Delta\Delta G_{A-U}$ to be the only variable parameter between species. It feeds into an interesting discussion, within protein folding, regarding the appropriate way to analyse a data-set such as this, which we expand our thoughts upon in the Supplementary Information. The average standard deviation obtained by this fitting protocol is 3 kJ mol$^{-1}$, which would define the absolute lower limit of $\Delta\Delta G_{A-U}$ values suitable for analysis.

We investigated the expected $\Delta\Delta G_{A-U}$ values upon mutagenesis by FoldX, which we demonstrate are predicted well in Figure 2. Considering allowed Alanine mutations within traditional Φ-value analysis (Val, Leu, Ile, Phe and Tyr) we obtain a mean $\Delta\Delta G_{A-U}$ of $7.99 \pm 3.09$ kJ mol$^{-1}$, identifying 21 of 23 residues viable for alanine-mutation and further analysis. It is apparent that many residues are viable for this type of analysis and a general applicability of the approach is expected for structurally well-defined amyloid systems.

It is intriguing that we find no hydrophobic truncation that stabilizes the amyloid structure. Given that no evolutionary pressure has optimized the sequence for this fold, it could be assumed that some mutations would remove strain from the amyloid fold and stabilizes the structure. This may elude to optimal hydrophobic packing of the PI3K-SH3 amyloid fibrils, despite no evolutionary pressure to guide the sequence to obtain such properties.

We should ensure that amyloid polymorphism is addressed, an interesting property of the amyloid fibrils, which, seemingly due to the lack of evolutionary pressure, can obtain numerous different amyloid folds[57–59]. The presence of different amyloid states deriving from polymorphism under a given set of solution conditions, could render accurate $\Phi$-value analysis of the elongation reaction very difficult. In order to avoid such complexity, both experimental methods and theoretical framework must be adequately adjusted. Here we formed a highly pure preparation of fibrils which has previously been characterized as an isolated single fibril morphology[21]. We only used this sample for further seeding of experimental samples. Combined with the utilization of the templating effect to induce the correct morphology, mentioned above, sources of variation have been minimized.

We find a concentration-dependence of the amyloid elongation $\Phi$-values. The bi-molecular nature of the elongation reaction is consistent with this observation. However, as we normalize one species during the kinetic measurements, the fibrilar end, this effect should not be significant. Since the folding mechanism is not concentration dependent[60], $\Phi$-values as folding probes should per extension also be concentration independent. The phenomenon can be explained by changes in the critical saturation concentration, between the WT and the mutant. It follows that this phenomenon would be prominent for mutations with high $\Delta\Delta G_{A-U}$ values and intermediate to low $\Phi$-values.

Kinetic measurements should be performed at high excess protein concentrations similar degrees of absolute super-saturation is achieved for both mutant and WT species. However, one should be careful to not saturate the fibril ends as the kinetic barrier of reorganization is dominating the kinetics under such conditions, rather than the kinetic barrier for attachment of soluble monomer[12].

We analyzed the elongation kinetics of PI3K-SH3 WT and mutant at 5 and 20 $\mu$M, well within the linear regime, which continues above 100 $\mu$M for the WT[61], ensuring the fibril end is not saturated.

It should be noted that the concept of what is a disruptive mutation in the amyloid elon-

gation reaction may have to be redefined. In traditional folding Φ-value analysis, a protein carrying a non-disruptive mutation is expected to acquire the WT fold, maintain a similar TSE (avoiding Hammond and anti-Hammond effects) and form no new interactions, as the mutation is a minimal deletion and should not interfere with the overall pattern of interactions. In the amyloid elongation reaction, the templating effect of the existing fibril induces the same amyloid fold on the mutated monomer[7,62,63] and more radical mutations may still undergo structural transition, catalyzed by the fibril end. Maintaining the same TSE is less trivial, as radical mutations may alter the misfolding pathway, especially in an energetic landscape which have not been under pressure for evolutionary optimization. It is interesting to further investigate whether more radical mutations, such as truncation of charged residues, still fulfill the necessary criteria of Φ-value analysis such that the role of charged residues in the misfolding reaction could be experimentally accessed. The protocol we are presenting intrinsically demonstrates the mutated proteins capability of adapting the native fold by elongation-specific kinetic measurements.

A concern to the protocol may be mutations capable of adapting to the native fold, but increase the reaction barrier of elongation significantly as to limit the seeding capacity of WT fibrils to an extend where de novo nucleation rates of the mutant fibril becomes significant within the recruitment time of monomer to the fibril seeds. Conservation of the native fibrillar morphology should be demonstrated with rigor for mutations with extreme increases in the reaction barrier. Global fitting of the denaturation series proves useful as a mutant which has adapted a non-native fold is likely have a different m-value compared to the native species and hence the fit of that particular mutant should prove poor. Additionally, one should consider the produced Φ-value of such mutants. Mutants adapting a non-native fold are likely structurally compensating for the change in sequence and should adapt a fold more stable than the seeded conformation of the mutant. Hence, one should expect $\Delta\Delta G_{A-U}$ values below the $\Delta\Delta G_{\ddagger-U}$ value and obtain a Φ-value above 1. In cases of extreme concern, it may be useful to investigate the thermodynamic stability of de novo fibrils of the mutant as well as the seeded mutant fibrils and demonstrate non-compatibility between the two. In our ensemble, we find no mutant demonstrating all of the above mentioned characteristics and detailed analysis of de novo fibril formation is not necessary.

In this work we investigate a single mutation not within the structured amyloid core, I82A. This mutation affects both the kinetics and thermodynamic stability of the fibrils,

though to a small extent. It is challenging to determine how this mutation is affecting the final structure or TS, as the residue is not located within an ordered segment. These results eludes to non-core interactions playing a quantifiable but unknown role in amyloid thermodynamics.

We have used guided MD-simulations to capture an appropriate estimation of the amyloid elongation TSE. It is essential to evaluate whether the identified TSE predicts behaviour of the amyloid, which have not been used for the guided simulation. To this effort we have applied Tanford analysis, which estimates the relative degree of residue exposure in the TS. Estimating the $\beta_{T_{sim}}$ is not trivial, as some non-native contacts are formed in the simulated TSE (Figure S3), whereas Tanford analysis assumes that the contacts investigated in the folded state and the TSE are similar. To simplify our analysis, we have constrained our $\beta_{T_{sim}}$ calculations to native contacts only. Under these simplifications the $\beta_{T_{sim}}$ is highly compatible with the experimentally identified Tanford value $\beta_{T_{exp}}$, supporting the viability of the MD-simulations. Additionally, we have used FoldX to calculate the expected $\Delta\Delta G_{\ddagger-A}$ from the simulated TSE, where we find good agreement between the experimental and predicted data (Figure S5).

# 5   Conclusion/Cover letter text

In summary, we have developed a protocol, which utilizes independent kinetics and stability measurements to apply the powerful tool of $\Phi$-value analysis to probe the TS of the elongation reaction of amyloid fibrils. The protocol introduces intrinsic control of acceptable mutations through the kinetic analysis and morphological constraints. We have applied this protocol to fifteen, non-disruptive hydrophobic deletions in the PI3K-SH3 domain, which has allowed us to perform the first appropriate MD simulations of the TS. We have experimentally validated the predictions of the simulations and can identify key monomer recognition sites within the fibril end, which could be key targets for inhibition of amyloid elongation.

# References

1. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **86,** 27–68 (2017).

2. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890 (2003).

3. Dobson, C. M. The Amyloid Phenomenon and Its Links with Human Disease. *Cold Spring Harb Perspect Biol* **9** (2017).

4. Buell, A. K. Stability matters, too – the thermodynamics of amyloid fibril formation. *Chem. Sci.* (2022).

5. Li, J., Uversky, V. N. & Fink, A. L. Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human alpha-synuclein. *Biochemistry* (2001).

6. Habchi, J. *et al.* Systematic development of small molecules to inhibit specific microscopic steps of A&#x3b2;42 aggregation in Alzheimer&#x2019;s disease. *Proceedings of the National Academy of Sciences* **114,** 200–208 (2017).

7. Eichner, T., Kalverda, A., Thompson, G., Homans, S. & Radford, S. Conformational conversion during amyloid formation at atomic resolution. *Mol Cell* **41,** 161–172 (2011).

8. Lövestam, S. & Scheres, S. H. W. High-throughput cryo-EM structure determination of amyloids. *The Royal Society of Chemistry* **240,** 243–260 (2022).

9. Buell, A. K. *et al.* Detailed Analysis of the Energy Barriers for Amyloid Fibril Growth. *Angevandte Chemie* **51,** 5247–5251 (2012).

10. Vettore, N. & Buell, A. K. Thermodynamics of amyloid fibril formation from chemical depolymerization. *Phys. Chem. Chem. Phys.* **21,** 26184–26194 (2019).

11. Jia, Z., Schmit, J. D. & Chen, J. Amyloid assembly is dominated by misregistered kinetic traps on an unbiased energy landscape. *Proceedings of the National Academy of Sciences* **117,** 10322–10328 (2020).

12. Buell, A. K. *et al.* Frequency Factors in a Landscape Model of Filamentous Protein Aggregation. *Phys. Rev. Lett.* **104,** 228101 (2010).

13. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annu Rev Biophys* **37,** 289–316 (2008).

14. Matouschek, A., Kellis, J. J., Serrano, L., Bycroft, M. & Fersht, A. Transient folding intermediates characterized by protein engineering. *Nature* **346,** 440–445 (1990).

15. Matouschek, A., Kellis, J., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340,** 122–126 (1989).

16. Matouschek, A., Serrano, L. & Fersht, A. R. The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* **224,** 819–835 (1992).

17. Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysisand Protein Folding* (Freeman, New York, 1999).

18. Karlsson, E. *et al.* Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins. *Journal of Biological Chemistry* **295,** 17698–17712 (2020).

19. Yang, J., Gao, M., Xiong, J., Su, Z. & Huang, Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Science* **28,** 1952–1965 (2019).

20. Wang, G. Z. & Fersht, A. R. Mechanism of initiation of aggregation of p53 revealed by Φ-value analysis. *PNAS* **112,** 2437–2442 (2015).

21. Röder, C. *et al.* Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy. *Nat Commun* **10,** 3754 (2019).

22. Zhao, D. & Moore, J. S. Nucleation–elongation: a mechanism for cooperative supramolecular polymerization. *Org. Biomol. Chem.* **1,** 3471–3491 (2003).

23. Korevaar, P. A., Schaefer, C., de Greef, T. F. A. & Meijer, E. W. Controlling Chemical Self-Assembly by Solvent-Dependent Dynamics. *J. Am. Chem. Soc.* **134,** 13482–13491 (2012).

24. Buell, A. K., White, D. A., Welland, M. E., Knowles, T. P. & Dobson, C. M. Surface Attachment of Protein Fibrils via Covalent Modification Strategies. *J. Phys. Chem.* **134,** 10925–10938 (2010).

25. Cohen, S. I. *et al.* Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. *J Chem Phys* **135** (2011).

27

26.   Lindahl, Abraham, Hess & van der Spoel. *GROMACS 2021 Manual* version 2021. Jan. 2021. `https://doi.org/10.5281/zenodo.4457591`.

27.   Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1,** 19–25 (2015).

28.   Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer physics communications* **185,** 604–613 (2014).

29.   Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115,** E4758–E4766 (2018).

30.   Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The journal of physical chemistry B* **119,** 5113–5123 (2015).

31.   Berendsen, H. J., Postma, J. v., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81,** 3684–3690 (1984).

32.   Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry* **18,** 1463–1472 (1997).

33.   Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103,** 8577–8593 (1995).

34.   Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126,** 014101 (2007).

35.   Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52,** 7182–7190 (1981).

36.   Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409,** 641–645 (2001).

37.   Paci, E., Vendruscolo, M., Dobson, C. M. & Karplus, M. Determination of a transition state at atomic resolution from protein engineering data. *Journal of molecular biology* **324,** 151–163 (2002).

38. Gsponer, J. & Caflisch, A. Molecular dynamics simulations of protein folding from the transition state. *Proceedings of the National Academy of Sciences* **99,** 6719–6724 (2002).

39. Lindorff-Larsen, K., Paci, E., Serrano, L., Dobson, C. M. & Vendruscolo, M. Calculation of mutational free energy changes in transition states for protein folding. *Biophysical journal* **85,** 1207–1214 (2003).

40. Buell, A. K., Dobson, C. M. & Welland, M. E. in *Amyloid Proteins* (eds Einar M. Sigurdsson, M. C. & Gasset, M.) 101–119 (Humana Press, New York, 2012).

41. Serrano, L., Kellis, J. T. J., Cann, P., Matouschek, A. & Fersht, A. R. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* **224,** 783–804 (1992).

42. Gsponer, J. *et al.* Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proceedings of the National Academy of Sciences* **103,** 99–104 (2006).

43. Sato, S., Religa, T. L., Daggett, V. & Fersht, A. R. Testing protein-folding simulations by experiment: B domain of protein A. *Proceedings of the National Academy of Sciences* **101,** 6952–6956 (2004).

44. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophysical journal* **83,** 3525–3532 (2002).

45. Paissoni, C. *et al.* Converging experimental and computational views of the knotting mechanism of a small knotted protein. *Biophysical Journal* **120,** 2276–2286 (2021).

46. Faisca, P. F., Nunes, A., Travasso, R. D. & Shakhnovich, E. I. Non-native interactions play an effective role in protein folding dynamics. *Protein Science* **19,** 2196–2209 (2010).

47. Shao, Q. & Zhu, W. Nonnative contact effects in protein folding. *Physical Chemistry Chemical Physics* **21,** 11924–11936 (2019).

48. Mouro, P. R., de Godoi Contessoto, V., Chahine, J., de Oliveira, R. J. & Leite, V. B. P. Quantifying nonnative interactions in the protein-folding free-energy landscape. *Biophysical Journal* **111,** 287–293 (2016).

49. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of pro-
    teins and protein complexes: a study of more than 1000 mutations. *Journal of molec-
    ular biology* **320,** 369–387 (2002).

50. Potapov, V., Cohen, M. & Schreiber, G. Assessing computational methods for pre-
    dicting protein stability upon mutation: good on average but not in the details.
    *Protein engineering, design & selection* **22,** 553–560 (2009).

51. Buß, O., Rudat, J. & Ochsenreither, K. FoldX as protein engineering tool: better
    than random based approaches? *Computational and structural biotechnology journal*
    **16,** 25–33 (2018).

52. Van der Kant, R., Louros, N., Schymkowitz, J. & Rousseau, F. Thermodynamic
    analysis of amyloid fibril structures reveals a common framework for stability in
    amyloid polymorphs. *Structure* **30,** 1178–1189 (2022).

53. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic acids
    research* **33,** W382–W388 (2005).

54. Zurdo, J. *et al.* Dependence on Solution Conditions of Aggregation and Amyloid
    Formation by an SH3 Domain. *Journal of Molecular Biology* **311,** 325–340 (2001).

55. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid
    fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95,** 4224–8 (1998).

56. Polverino de Laureto, P. *et al.* Protein aggregation and amyloid fibril formation by
    an SH3 domain probed by limited proteolysis. *J Mol Biol* **334,** 129–141 (2003).

57. Fändrich, M. *et al.* Amyloid fibril polymorphism: a challenge for molecular imaging
    and therapy. *J Intern Med* **283,** 218–237 (2018).

58. Tycko, R. Amyloid polymorphism: structural basis and neurobiological relevance.
    *Neuron* **86,** 632–645 (2015).

59. Close, W. *et al.* Physical basis of amyloid fibril polymorphism. *Nat Commun* **9,** 699
    (2018).

60. Samiotakis, A, Wittung-Stafshede, P & Cheung, M. Folding, stability and shape of
    proteins in crowded environments: experimental and computational approaches. *Int
    J Mol Sci* **2,** 572–588 (2009).

61. Coden, A. *Biophysical characterization of amyloid fibril elongation: a surface-based
    biosensing approach* MA thesis (University of Padua, 2018).

62.  Riek, R. & Eisenberg, D. The activities of amyloids from a structural perspective. *Nature* **539,** 227–235 (2016).

63.  Del Mar, C, Greenbaum, E., Mayne, L., Englander, S. & Woods, V. J. Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. *Proc Natl Acad Sci* **102,** 15477–15782 (2005).

# 6    Supplementary Material

| | Rel. elongation $k_{mut}/k_{WT}$ 5 $\mu$M | Rel. elongation $k_{mut}/k_{WT}$ 20 $\mu$M | $\Delta G_{A-U}$ kJ/mol |
|---|---|---|---|
| A3G | $0.359 \pm 0.037$ | $0.333 \pm 0.024$ | $-53.60 \pm 1.47$ |
| Y6A | $0.672 \pm 0.095$ | $0.378 \pm 0.021$ | $-49.88 \pm 1.25$ |
| Y8A | $0.005 \pm 0.002$ | $0.005 \pm 0.000$ | $-40.61 \pm 0.71$ |
| L11V | $2.152 \pm 0.418$ | $1.205$ | $-54.69 \pm 1.54$ |
| L11A | $0.598 \pm 0.016$ | $0.384 \pm 0.021$ | $-48.74 \pm 1.18$ |
| I22A | $0.019 \pm 0.002$ | $0.111 \pm 0.021$ | $-42.81 \pm 0.86$ |
| I29A | $0.020 \pm 0.002$ | $0.194 \pm 0.004$ | $-50.50 \pm 1.33$ |
| V32A | $0.268 \pm 0.009$ | $0.260$ | $-48.37 \pm 1.16$ |
| L40V | $1.265 \pm 0.058$ | $1.460$ | $-57.28 \pm 1.70$ |
| I53A | $0.003 \pm 0.001$ | $0.002$ | $-40.13 \pm 0.70$ |
| T62S | $0.937 \pm 0.115$ | $0.532 \pm 0.049$ | $-54.05 \pm 1.50$ |
| T63S | $1.102 \pm 0.123$ | $0.693 \pm 0.027$ | $-52.71 \pm 1.42$ |
| V74A | $0.998 \pm 0.097$ | $0.862 \pm 0.061$ | $-46.05 \pm 1.02$ |
| I77A | $1.109 \pm 0.127$ | $1.106 \pm 0.053$ | $-53.99 \pm 1.53$ |
| I82A | $1.589 \pm 0.194$ | $1.442 \pm 0.102$ | $-58.87 \pm 1.83$ |
| WT | —— | —— | $-56.01 \pm 1.61$ |

Table S1: All relative elongation rates and $\Delta G_{A-U}$-values. Elongation rates are measured by Quarts Crystal Microbalance, as the slope of the 3rd overtone frequency during incubation of respective monomer concentrations over a sensor of immobilized WT fibrils. Each elongation rate is normalized to a WT injection prior to mutant injections. $\Delta G_{A-U}$-values are derived from the cooperative elongation model fitting (Figure 2) of denaturation series by chemical denaturation.

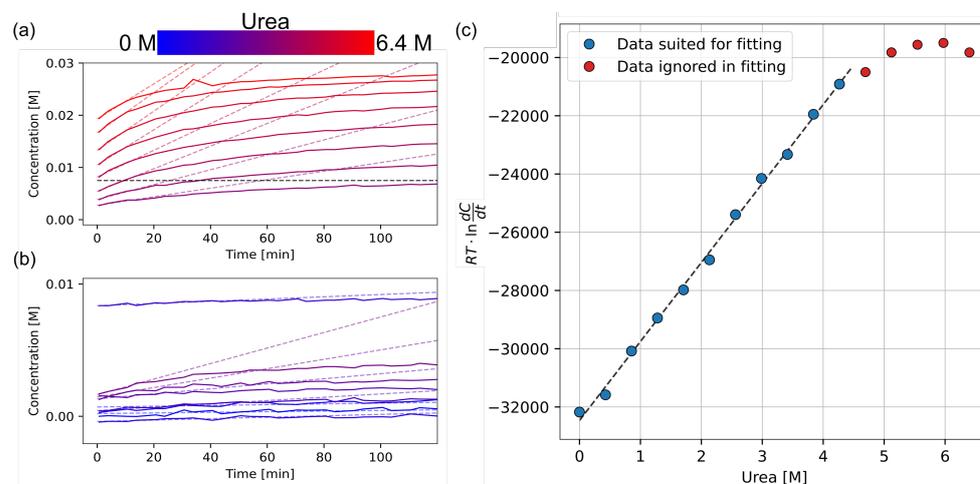| Equillibrium denaturation | |
|---|---|
| $m_{A-U}$ | $6567.62 \pm 401.53 \ J \cdot mol^{-1}[D]^{-1}$ |
| $\sigma$ | $0.214 \pm 0.046$ |
| Unfolding Kinetics | |
| $m_d$ | $2711.76 \pm 43.82 \ J \cdot mol^{-1}[D]^{-1}$ |
| c | $-32462.69 \pm 110.61 \ J \cdot mol^{-1}$ |
| Tanford $\beta$ | |
| Experimental | $0.41 \pm 0.03$ |
| Simulation | $0.45$ |

Table S2

31

Figure S1: Unfolding kinetics of PI3K-SH3 WT fibrils with increasing concentrations of Urea. (a) and (b) release of monomer upon incubation with increasing concentrations of Urea. Starting fibril concentrations are 30 $\mu$M equivalent monomer mass and 10 mM glycine hydrochloride at pH 2. The buffer capacity of urea is compensated with HCl and ionic strength is kept constant with NaCl. Soluble protein is calculated from the 350/330 nm ratio of intrinsic flourescence exited at 280 nm, compared to 30 $\mu$M monomer samples under the same conditions. Similarly colored dashed lines mark initial slope fitted by linear regression, the slope equals $\frac{\partial c([D])}{\partial t}$. The black dashed line marks the upper limit for starting concentrations, where initial assumptions described in the method section are assumed to hold true (25% monomer release). The different starting concentrations of monomer is caused by denaturation suring sample handling prior to measurement. The single outlier starting concentration in (b) is caused by of-center alignment of the measurement capillary causing background noise. The calculated monomer release should still hold true, and fits within the expected rates, shown in (c) as measurement point 4 at 1.28M Urea. (c) $RT \cdot \frac{\partial c([D])}{\partial t}$ vs. [D] plot. Circles denote calculated values from initial slope fits in (a) and (b). Red data points are ignored in the fitting as the free monomer concentration upon first measurement is too high for the assumptions of similar number of fibril ends to hold true. Dashed line is linear fit to the blue data-points, slope of the fit equals the $m_d$-value.
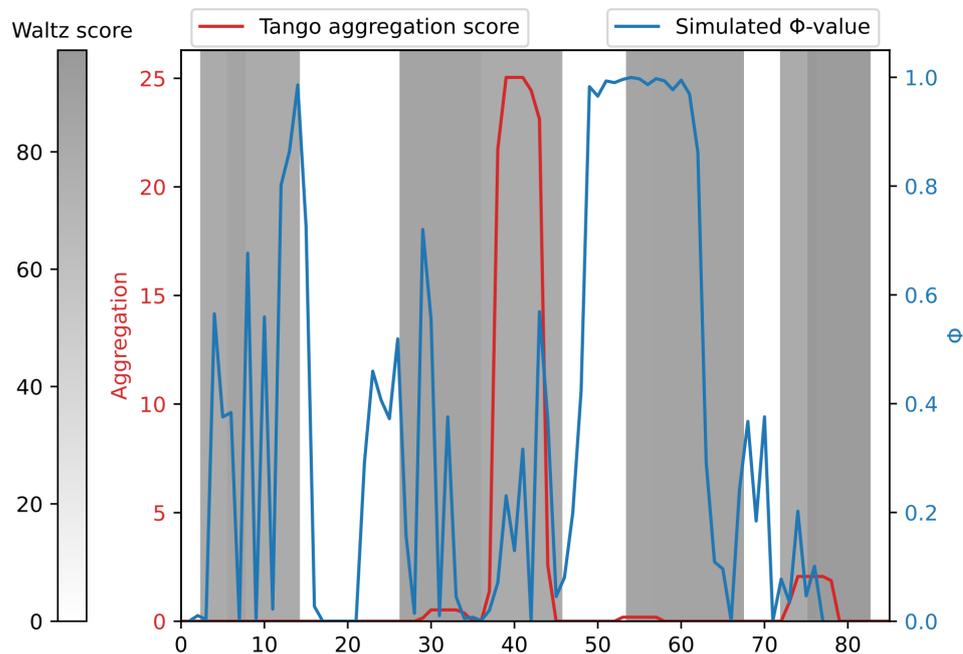
Figure S2: Tango and Waltz predictions. Tango aggregation score in red, Waltz score is shaded in grey and Φ-values are in blue. Tango aggregation score is calculated with no terminus modifications, for 298 K and with an ionic strength of 0.02 M. The Waltz score is calculated using the high sensitivity threshold and pH 2.6.

## 6.1   Computational calculations of free energy changes

We used ensembles of structures from our fibril state simulation and TS simulation as input in order to calculate $\Delta\Delta G^{\mathrm{calc}}_{A-U}$ and $\Delta\Delta G^{\mathrm{calc}}_{\dagger-U}$ respectively. By taking the ratio we obtain $\Phi^{\mathrm{calc}}$ based on the free energies of the structures rather than the native-contact interpretation utilised in the simulations. Each mutation is modeled individually for each conformation and then averaged over the respective ensembles. Mutations A3G, L11V and T62S which do not give reliable $\Phi^{\mathrm{exp}}$ were excluded from the analysis. In order to be able to discriminate between free energies of fibril and TS structures, which share the same structure in three lower chains, the mutation was only modeled in the top chain. The comparison between $\Delta\Delta G^{\mathrm{exp}}_{A-U}$ and $\Delta\Delta G^{\mathrm{calc}}_{A-U}$, generated from an ensemble of 25 amyloid structures, is shown in Figure S5(a). Overall the agreement is good, although $\Delta\Delta G^{\mathrm{calc}}_{A-U}$ fall systematically below $\Delta\Delta G^{\mathrm{exp}}_{A-U}$ and there is a major outlier, Y8A, which has $\Delta\Delta G^{\mathrm{exp}}_{A-U} = 15\,\mathrm{kJ\,mol^{-1}}$ and $\Delta\Delta G^{\mathrm{calc}}_{A-U} = 3\,\mathrm{kJ\,mol^{-1}}$ . This is probably the result of only modeling mutations in the top chain which is less destabilising to the fibril structure than mutating chains in the middle.
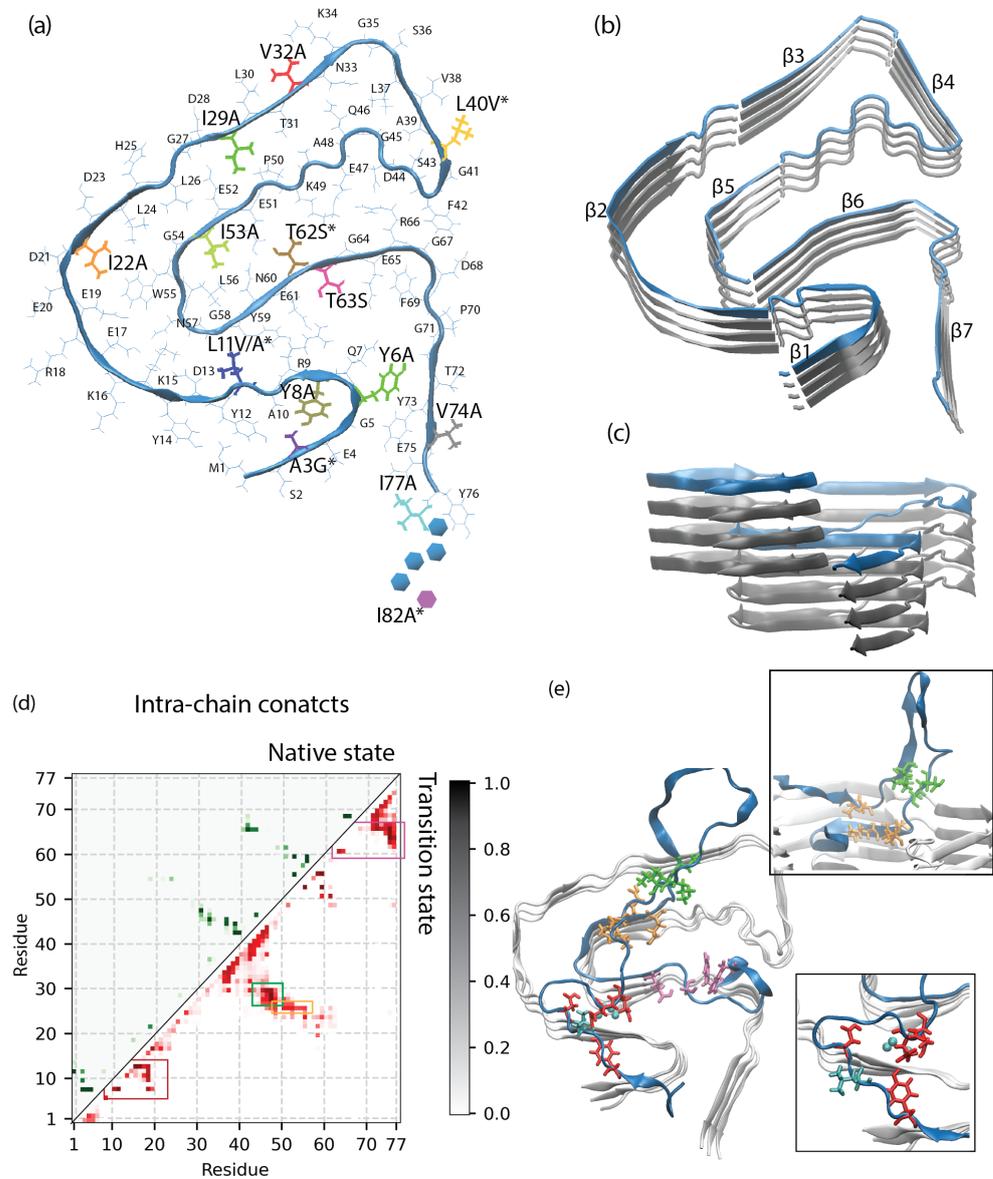
Figure S3: Structures of the P13K-SH3 fibril. (a) Top-view of the fibril with the point-mutations highlighted. * means that residue's Φ-value was not included as a restraints in the simulation. (b) (c) Two views of the starting structure for our simulations. The structure is made of four staggered subunits. The Φ-value restraints are implemented to simulate the transition state of the top unit (blue) binding to the pre-exising fibril (grey). (d) Comparison of contact maps of the top unit in our fibril state simulation (green) and TS simulation (red). The TS displays more variation and some non-native self-interactions within the top unit. (e) One snapshot of the TS. The coloured side-chains correspond with the boxed regions in (d), highlighting non-native contacts which appear to determine intra-chain structure in the TS. Within the red cluster, L11 is shown in cyan and L11's native atom contacts, which are in the lower chain, are shown as cyan spheres. L11 is therefore blocked from its native contacts in the TS.
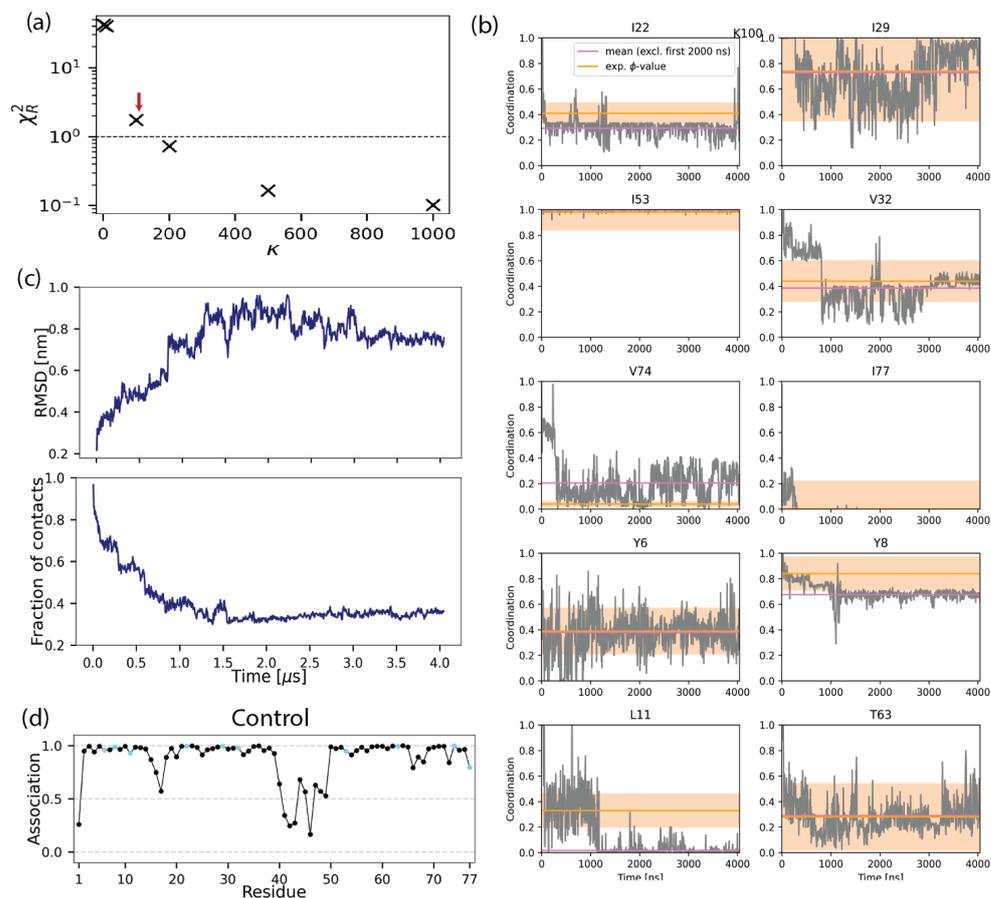
Figure S4: (a) Scan of force constants, $\kappa$, used for the $\Phi$-value restraints. A short simulation was performed with each $\kappa$ and the agreement between $\Phi^{\mathrm{exp}}$ and $\Phi^{\mathrm{sim}}$ was assessed. We chose to continue with $\kappa = 100$ since it gives $\chi^2_{\mathrm{red}}$ close to 1. (b) For each $\Phi$-value restraint, the target, $\Phi^{\mathrm{exp}}$, is shown in orange and the fraction of native contacts per simulation frame is plotted in gray. The average fraction of native contacts, discarding the first half of the simulation, is shown in pink, $\Phi^{\mathrm{sim}}$. (c) Top: the root mean square deviation from the fibril structure. Bottom: The fraction of total contacts between the top and second chains in the fibril. Both appear to have converged after $2\mu s$. (d) The average fraction of native contacts maintained to the second unit for each residue in the control simulation (simulation without $\Phi$-value restraints)
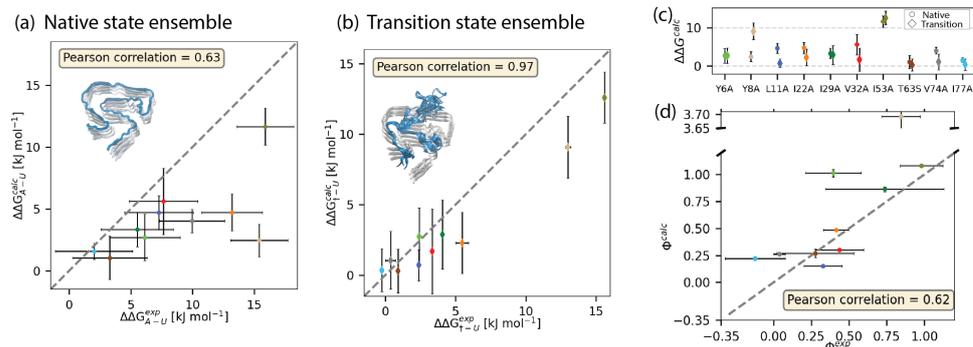
35

Figure S5: Comparison of experimental and FoldX-calculated ΔΔGs of MD simulated structures. (a) Comparison between $\Delta\Delta G_{A-U}^{\mathrm{exp}}$ and $\Delta\Delta G_{A-U}^{\mathrm{calc}}$ for structures from our fibril state ensemble. The mutations are modeled in the top unit only. The error bars in the x-axis signify the experimental uncertainty and the error bars in the y-axis signify the standard deviation of the ensemble. The inset shows a small representation of the ensemble. (b) Comparison between $\Delta\Delta G_{\dagger-U}^{\mathrm{exp}}$ and $\Delta\Delta G_{\dagger-U}^{\mathrm{calc}}$ for structures from our transition state ensemble. The mutations are modeled in the top unit only. The error bars in the x-axis signify the experimental uncertainty and the error bars in the y-axis signify the standard deviation of the ensemble. The inset shows a small representation the ensemble. (c) Plot of $\Delta\Delta G_{A-U}^{\mathrm{calc}}$ and $\Delta\Delta G_{\dagger-U}^{\mathrm{calc}}$ per mutation. (d) Comparison between $\Phi^{\mathrm{exp}}$ and $\Phi^{\mathrm{calc}}$.

The good agreement between $\Delta\Delta G_{\dagger-U}^{\mathrm{exp}}$ and $\Delta\Delta G_{\dagger-U}^{\mathrm{calc}}$ , generated from an ensemble of 150 TS structures, is shown in Figure S5(b). Finally, $\Phi^{\mathrm{calc}}$ versus $\Phi^{\mathrm{exp}}$ is plotted in Figure 2(e), where the uncertainty for $\Phi^{\mathrm{calc}}$ was propagated using the standard errors of the mean of the two ensemble-derived $\Delta\Delta G^{\mathrm{calc}}$'s. The agreement is good overall, thereby validating our fibril simulation results with a second method. Most of the calculated values fall between 0 and 1. FoldX predicts the mutation Y8A to be more destabilising in the TS than the amyloid state because Y8 has an increased number of (non-native) contacts in the TS structures. This is not observed in the experimental data. This leads to $\Phi^{\mathrm{calc}} = 3.7$, much larger than 1.

## 6.2   Discussion of fitting strategy for independent measurements

We have applied the cooperative elongation model to globally fit 20 denaturation curves of five WT fibril samples and fifteen mutant fibril samples. To constrain our fitting protocol we have imposed the highest degree of constraints on the fitting protocol by allowing only the $\Delta G$ parameter to vary between WT and mutant samples to minimize the risk of over-fitting. In the following section we will discuss how these constraints were chosen and what considerations should be made with global fittings of these type of data-sets.

It is important to note that the denaturation series investigated here are performed over a period of approximately four years, between two different labs among several researchers. The comparability allowing global fitting arises from diligent adherence to the experimental protocols, all samples being seeded with the same well-defined fibril stock and the use of several WT references demonstrating replicability across time, labs and researchers.

Firstly we must consider two of the fitting parameters; m and $\sigma$. The m-value is the sensitivity to denaturant; $m = \frac{\mathrm{d}\Delta G}{\mathrm{d}[D]}$. According to the transfer model, the m-value is associated with difference in exposed surface area between the folded and unfolded state[1]. Hence m-values are correlated with the unfolded structural ensemble, or here the monomeric structural ensemble. It is known that mutations in folded proteins can alter both the stability of the native fold and the m-value. The effect of mutations upon m-values be quite significant, altering the m-value by up to 50 %[2]. It is also important to note, that the seeding protocol applied in this work ensures that the examined fibril structure is the same across all samples. It would then follow that when the same structure is investigated, and any change in m-value would originate from the monomeric structural ensemble. PI3K-SH3, in the conditions investigated here, contains no secondary structure in its monomeric form, but is more compact compared to its fully denatured state[3–5]. M utations may alter the m-value, if the radius of gyration ($R_g$) of the monomeric state is significantly altered. With no significant remaining structure, it is a fair assumption that conservative mutations, as we investigate here, will not alter the monomeric structural ensemble significantly. With these sound assumptions, we have chosen to minimize the risk of overfitting and kept the m-parameter global.

The $\sigma$-value of the cooperative elongation model effectively describes the energy penalty of the amyloid nucleus compared to the mature fibril. The energy-penalty arises from the increased surface to volume ratio of the nucleated or oligomeric species, since the surface-exposed monomers are incapable of forming all contacts, resulting in decreased enthalpy-gain but with similar entropy-penalties upon growth. When considering two amyloid species with the same mature structures and similar monomeric strucural ensembles, it seems reasonable that nucleation pathways would also be similar. If that is the case, the energy-penalty of the oligomeric state should manifest as a similar $\sigma$-value, meaning that fitting $\sigma$ as a globally shared parameter is reasonable.

## 6.3 Supporting results

### 6.3.1 Mutant fibril preparation

We investigated the resulting fibrils by AFM. The overall fibrillar morphology of the WT seeds is maintained for all mutants. Intrinsically for our protocol, the ability of the mutant to elongate the WT fibrils is established through QCM measurements.

It is generally accepted that fibrils robustly propagate their structure during elongation[6–9] of the same protein through a dock-lock mechanism[10], especially at high seed to monomer ratio[11]. We tested whether or not the elongation with mutant monomer would interfere with the propagation of the WT seed structure, by incubating the seed fibrils on the QCM again with WT protein after the injection of mutant protein. The growth rate upon WT monomer injection was comparable to the initial WT-induced growth rate, suggesting that the fibril structure was maintained during the entire experiment.
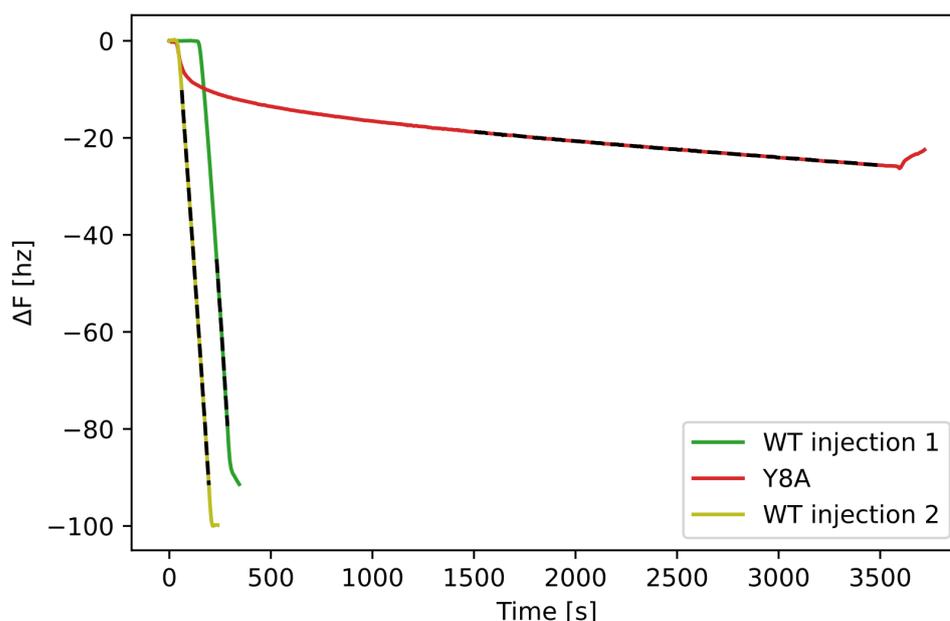


Figure S6: Elongation kinetics measured with QCM. Immobilized WT seed fibrils are exposed to WT injection 1, then Y8A and then WT injection 2. Dashed lines denote data fit for linear regression and rate comparison. WT elongation rate is maintained within 5 % of initial rate.

### 6.3.2 Equilibrium denaturation

We measured the soluble protein fraction from intrinsic fluorescence emission ratios of 340/310 nm or 350/330nm. This methodology was possible due to the presence of tryp-

tophan in the protein sequence[12], which has also been demonstrated for this particular system, where the tryptophan is fully quenched in the fibrillar state and fluorescent in the monomeric state[13]. In order to utilize the accurate, and low-volume method, of capillary based nanoDSF, compatibility between the different fluorescence ratios must be demonstrated. We demonstrated compatibility between the 340/310 and the 350/330 ratio by comparing the calculated free energies of polymerisation measured with both methods. Soluble WT protein was calculated from the normalised fluorescence ratio of 340 nm / 310 nm and 350 nm / 330 nm using a FLUOstar Omega instrument as well as the 350 nm / 330 nm ratio from nanoDSF using a Prometheus Panta. A global fit, using the cooperative polymerization model with shared m and $\sigma$-values, is performed and plotted (S7). The difference between free energies of polymerisation is below 0.6 $kJmol^{-1}$ and well within the fitting error. Hence the estimated energies derived from each measurement protocol can be considered effectively identical.
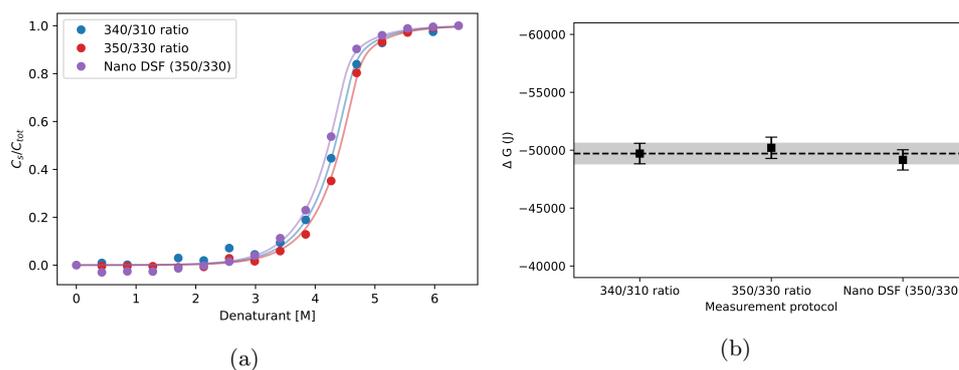


Figure S7: Cooperative elongation model fits to PI3K-SH3 denaturation series with shared m and $\sigma$-values. **(a)** fraction of soluble protein plotted against urea concentration. Circles are calculated soluble protein from intrinsic fluorescence ratio. Solid lines are the reported fits. **(b)** error-bar plot of calculated $\Delta G_{A_U}^{Buff}$ energies. The dashed line denotes the energy calculated from the 340 nm / 310 nm ratio while the shading denotes the SD limits.

# References

1. O'Brien, E. P., Brooks, B. R. & Thirumalai, D. Molecular origin of constant m-values, denatured state collapse, and residue-dependent transition midpoints in globular proteins. *Biochemistry* **48,** 3743–3754 (2020).

2. Shortle, D. in *Protein Stability* (eds Anfinsen, C., Richards, F. M., Edsall, J. T. & Eisenberg, D. S.) 217–247 (Academic Press, 1995). `https://www.sciencedirect.com/science/article/pii/S0065323308603368`.

3. Zurdo, J. *et al.* Dependence on Solution Conditions of Aggregation and Amyloid Formation by an SH3 Domain. *Journal of Molecular Biology* **311,** 325–340 (2001).

4. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95,** 4224–8 (1998).

5. Polverino de Laureto, P. *et al.* Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J Mol Biol* **334,** 129–141 (2003).

6. Riek, R. & Eisenberg, D. The activities of amyloids from a structural perspective. *Nature* **539,** 227–235 (2016).

7. Del Mar, C, Greenbaum, E., Mayne, L., Englander, S. & Woods, V. J. Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. *Proc Natl Acad Sci* **102,** 15477–15782 (2005).

8. Eichner, T., Kalverda, A., Thompson, G., Homans, S. & Radford, S. Conformational conversion during amyloid formation at atomic resolution. *Mol Cell* **41,** 161–172 (2011).

9. Petkova, A. T., Leapman, R. D., Guo, Z., Yao, W. & Tycko, R. Self-Propagating, Molecular-Level Polymorphism in Alzheimer's &#xdf;-Amyloid Fibrils. *Science* **307,** 262–265 (2005).

10. Esler, W. P. *et al.* Alzheimer's disease amyloid propagation by a template-dependent dock- lock mechanism. *Biochemistry* **39,** 6288–6295 (2000).

11. Peduzzo, A., Linse, S. & Buell, A. K. The Properties of α-Synuclein Secondary Nuclei Are Dominated by the Solution Conditions Rather than the Seed Fibril Strain. *ACS Chem. Neurosci.* **11,** 909–918 (2020).

12. Pedersen, J. S., Dikov, D., Hjuler, H. A, Christiansen, G. & Otzen, D. E. The changing face of glucagon fibrillation: structural polymorphism and conformational imprinting. *J. Mol. Biol.* **355,** 501–523 (2005).

13. Vettore, N. & Buell, A. K. Thermodynamics of amyloid fibril formation from chemical depolymerization. *Phys. Chem. Chem. Phys.* **21,** 26184–26194 (2019).