GÖTEBORGS UNIVERSITET

Experimental Protein Dynamics and its Role in Predicting Protein Function

Maja Jensen

Department of Chemistry and Molecular Biology
Gothenburg 2022

Thesis for degree of Doctor of Philosophy in Natural Science, Biophysics

Maja Jensen

Experimental Protein Dynamics and its Role in Predicting Protein Function

ISBN: **978-91-8009-841-0**

Department of Chemistry and Molecular Biology

Division of Biochemistry and Structural Biology

University of GothenburgSE-405 30, Göteborg, Sweden

Printed by Stema Specialtryck AB, Borås, Sweden, 2022

“Sometimes it is the people no one can imagine anything of who do the things no one can imagine.”
- Alan Turing

Abstract

By studying proteins we learn about the processes that control life, such as important procedures in the body, diseases and eventually find more targeted cures for many diseases.

Proteins are constantly being built up and decomposed in living organisms. Many proteins move and interact with other proteins or small molecules in the cell. They can for example have enzymatic functions where they catalyze a biochemical reaction, or they can be situated in the cell membrane controlling the flow of smaller molecules. Some proteins are involved in several different processes depending on with which proteins they interact. By investigating the dynamics and interactions of proteins we can learn more about their function.

In this work I have been involved in several project with focus on method development. X-ray crystallography diffraction experiments were performed at the short pulse facility beamline FemtoMAX at MAXIV, with and without effect of terahertz radiation. We found and investigated an interesting connection between survivin and PRC2, which are both involved in several diseases. Bayesian machine learning methods were implemented in the analysis of MST data.

Sammanfattning på svenska

Genom att studera proteiner lär vi oss mer om processerna som kontrollerar liv, så som viktiga funktioner i kroppen, sjukdomar och vi kan hitta riktade sätt att bota sjukdomar.

Proteiner byggs upp och bryts ner i en ständig process i levande organismer. Många proteiner flyttar sig och interagerar med proteiner eller andra typer av molekyler i cellerna. De kan tex fungera som enzymer vilka katalyserar biokemiska reaktioner, eller så kan de fungera som ”portar” i cellmembranet där de reglerar vilken typ av molekyler som kan tränga igenom. Vissa proteiner ingår i flera olika funktioner beroende på vilka andra proteiner de interagerar med. Genom att undersöka dynamik och interaktion hos proteiner kan vi lära oss mer om hur de fungerar.

I det här arbetet har jag varit inblandad i flera olika projekt med fokus på metodutveckling. Röntgenkristallografi utfördes i MAXIV, Lund, med och utan terahertzstrålning. Vi hittade ett intressant samband mellan survivin och PRC2, vilka båda är involverade i många sjukdomar. Bayesiansk machine learning implementerades i analys av MST data.

List of publications

Paper I

Maja Jensen, María-José García Bonéte Atsarina Larasati Anindya, Karin Andersson, Malin C. Erlandsson, Nina Oparina, Ulrika Brath, Venkataragavan Chandrasekaran, Maria Bokarewa and Gergely Katona **“Survivin prevents the Polycomb Repressor Complex 2 from methylating Histone 3 lysine 27”** *Manuscript*

Contribution: Designed peptide microarray, designed longer peptides, produced survivin, produced isotope labeled survivin. Performed and analyzed NMR spectroscopy titration experiments. Performed and analyzed SAXS experiment. Took part in planning of BLI experiments. Prepared some of the figures.

Paper II

María-José García Bonéte, **Maja Jensen**, Christian V Recktenwald, Sandra Rocha, Volker Stadler, Maria Bokarewa, and Gergely Katona. **"Bayesian Analysis of MicroScale**

Thermophoresis Data to Quantify Affinity of Protein:Protein Interactions with Human Survivin." Scientific Reports, 2017, Vol. 7, Iss. 1 7.1 (2017): Scientific Reports, 2017, Vol. 7, Iss. 1. Print.

Contribution: Took part in production and labeling of survivin and performed MST experiments.

Paper III

Atsarina Larasati Anindyaa, María-José García Bonéte, **Maja Jensen**, Christian V. Recktenwald, Maria Bokarewa and Gergely Katona **"Bayesian Progress Curve Analysis of MicroScale Thermophoresis Data"** Digital Discovery, 2022

Contribution: Took part in production and labeling of survivin and performed MST experiments.

Paper IV

Jensen, Maja, Viktor Ahlberg Gagnér, Juan Cabello Sánchez, Åsa U J Bengtsson, J Carl Ekström, Tinna Björg Úlfarsdóttir, Maria-Jose Garcia-Bonete, Andrius Jurgilaitis, David Kroon, Van Thai Pham, Stefano Checcia, Hélène Coudert-Alteirac, Siawosch Schewa, Manfred Rössle, Helena Rodilla, Jan Stake, Vitali Zhaunerchyk, Jörgen Larsson, and Gergely Katona. "**High-resolution Macromolecular Crystallography at the FemtoMAX Beamline with Time-over-threshold Photon Detection.**" Journal Of Synchrotron Radiation, 2021, Vol. 28, Iss. Pt 1, Pp. 64-.70 28.Pt 1 (2021): 64-70. Print.

Contribution: Prepared crystals, took part in planning and conducting the experiment at the synchrotron. Analyzed the data and prepared figures.

Paper V

Viktor Ahlberg Gagnér, **Maja Jensen**, Juan Cabello Sánchez, Åsa U. J. Bengtsson, J. Carl Ekström, Tinna Björg Úlfarsdóttir, Zhedong Zhang, Andrius Jurgilaitis, David Kroon, Van-Thai Pham, Stefano Checcia, Hélène Coudert-Alteirac, Siawosch Schewa, Manfred Rössle, Helena Rodilla, Jan Stake, Vitali Zhaunerchyk, Stefano A. Mezzasalma, Ran Friedman,

Jörgen Larsson and Gergely Katona **“Ultrafast structural response of a protein crystal to a strong pulsed THz field”** *Manuscript*

Contribution: Prepared crystals, took part in planning and conducting the experiment at the synchrotron.

Related Publications

García-Bonete, María-José, **Maja Jensen**, and Gergely Katona. "**A Practical Guide to Developing Virtual and Augmented Reality Exercises for Teaching Structural Biology."** Biochemistry and Molecular Biology Education 47.1 (2019): 16-24. Web.

Ahlberg Gagnér, Viktor, **Maja Jensen**, and Gergely Katona. **"Estimating the Probability of Coincidental Similarity between Atomic Displacement Parameters with Machine Learning."** Machine Learning-Science And Technology, 2021, Vol. 2, Iss. 3 2.3 (2021): Machine Learning-Science And Technology, 2021, Vol. 2, Iss. 3. Print.

Chandrasekaran, Venkatagaran, N. Oparina, Maria-Jose Garcia-Bonete, C. Wasen, Malin Erlandsson, Eric Malmhäll-Bah, Karin Andersson, **Maja Jensen**, Sofia Töyrä Silfverswärd, Gergely Katona, and Maria Bokarewa. "**Cohesin-Mediated Chromatin Interactions and Autoimmunity**." Frontiers In Immunology, 2022, Vol. 13 13 (2022): Frontiers In Immunology, 2022, Vol. 13. Print.

Contents

Introduction

Atoms can be considered as small building blocks of all life in universe, they each have its own unique property. When two or several atoms are bound they form molecules, which also have their own unique properties. Molecules could be small, such as water molecules, or it could be thousands or millions of atoms creating huge biomolecules, like DNA or proteins. Complex and extensive combinations of small and large molecules are required to create living organisms. Proteins are found everywhere in living organisms, where they serve as enzymes, perform cell membrane transport, participate in signaling and much more. Malfunction, abundance or deficiency of specific proteins may lead to an imbalance causing diseases in the human body. For us to learn more about these biological processes it is of great value to be able to examine the dynamics, structure and function of proteins.

A variety of different methods have been developed with the purpose of performing structural and functional studies of proteins. One of the oldest and still most commonly used methods for structure determination of proteins is X-ray crystallography. By measuring x-ray diffraction of a protein crystal a detailed high-resolution model of a protein structure can be obtained. The limitation is that this method gives relatively little information about how the protein behaves in its natural environment. Other X-ray methods, e.g. solution scattering, offer more information about proteins in solution, yet at much lower resolution.

Time-resolved studies at X-ray free electron lasers are an emerging area, which presents high resolution and information about structural changes and protein dynamics. The limiting step with this method is often sample preparation since large quantities of microcrystals of good quality with high internal order are needed.

For studies of smaller proteins in solution Nuclear Magnetic Resistance (NMR) spectroscopy is an excellent method as it provides a more accurate model of a protein in its natural environment. Cryogenic Electron Microscopy (cryo-EM) has lately been a rapidly emerging method for protein structure determination with the advantage of the possibility to capture different states of the protein at a rather good resolution. Often a combination of several methods is used to determine the structure and investigate the dynamics and function of a protein.

In my research group we have been working with several of these methods, in both conventionally and in less conventional manners. We have introduced a terahertz pump to X-ray crystallography leading the way to perform time-resolved crystallography experiments at the short pulse facility FemtoMAX in MAXIV (Lund, Sweden). We are also presenting alternative approaches of analyzing the results, including machine learning and Bayesian statistical analysis.

Electromagnetic Energy

Electromagnetism is a physical phenomenon that can be described interchangeably either as light, electromagnetic waves or as radiation. Electromagnetic waves have properties similar to mechanical waves, with the difference that they do not need a medium to propagate through. Instead, they can travel through a medium or through vacuum. The phenomena of electromagnetic waves can also be described as photons, or packages of light, which behave as particles.[1, 2] The fact that light behaves both as particles and waves and these properties can be used depending on what one wants to detect. The properties elastic scattering for example is used in X-ray experiments, such as Small Angle X-ray Scattering (SAXS) or X-ray diffraction where the wavelength and the energy of the scattered wave doesn't change, only the direction is altered.[3]

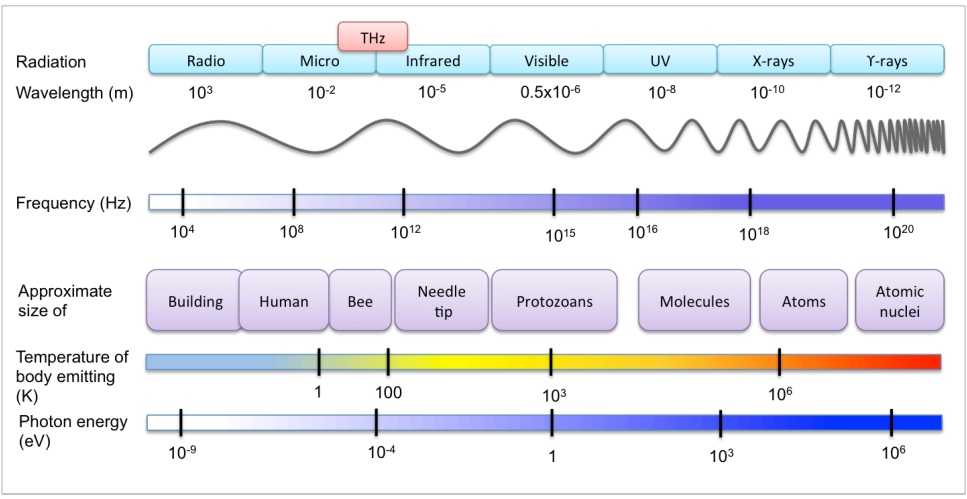


Figure 1, The electromagnetic spectrum.

Frequency, energy and wavelength are different ways to describe electromagnetic energy. Commonly larger waves, such as radio- and microwaves, are measured in frequency, ν (Hertz; Hz), visible light is measured in wavelengths, λ (meters, m) and x-rays in energy, E (electron volts, eV). The different ways of measuring electromagnetic energy are related by the following equations;

$$E = h\nu$$

Equation 1

Where h is the Planck's constant ($6{,}62607015 \times 10^{-34}$ Js)

Planck's constant relates the energy in one photon of electromagnetic radiation to the frequency of that photon.

The frequency, ν is related to the wavelength by;

$$\nu = \frac{c}{\lambda}$$

Equation 2

Where the speed of light in vacuum, c = 299 792 458 m/s

Already the 19th century Maxwell explained electromagnetic waves and how they propagate at constant speed. By a number of equations he explains that an electromagnetic field consist of an electric and a magnetic field, which are perpendicular in relation to each other and at the same time constantly oscillating (Fig 2). [4]

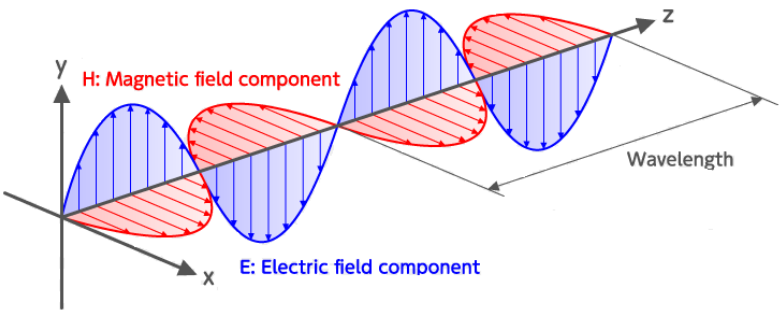


Figure 2, The electromagnetic wave displaying the magnetic and the electric components.

When a photon hits a matter there are some different possible outcomes, which depend on the energy of the photon and the nature of the matter. Elastic scattering means that there is no energy is absorbed and hence there is no change in energy, only in the direction of the scattered waves. Examples of elastic scattering are diffraction or reflection.[5] Inelastic scattering occurs when the matter that interacts with the light absorbs some of its energy. An example of this is Raman or Compton scattering where the light is scattered with lower energy than the incoming energy. [5, 6]

Protein Dynamics

Proteins are large biological molecules consisting of hundreds of covalently bound amino acids, which possess different internal properties. Consequently many internal and external forces are affecting these large biomolecules. Motions in proteins are occurring on a wide range of time and length scale and the dynamic of proteins is a highly complex phenomena. Covalent bonds are constantly vibrating with very high frequencies, from femtoseconds to picoseconds. Side chain rotations with in the proteins are occurring with a frequency of picoseconds to nanoseconds. Protein backbone fluctuations take place within nanoseconds. Dissociation and association of protein ligands occurs in nanoseconds to microseconds. Catalytic events may take place in a matter of microseconds to milliseconds. Protein folding and allosteric changes occur within seconds or even hours. [7-9]

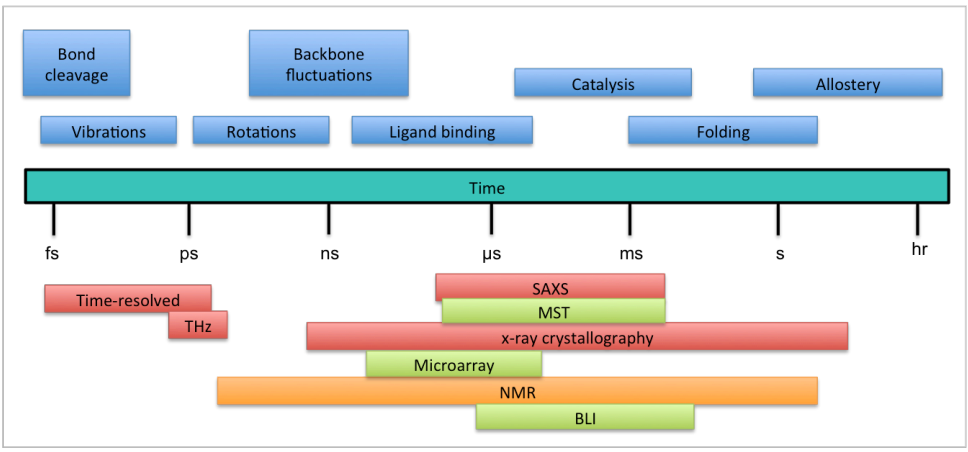


Figure 3, Protein dynamics (blue) portrayed in relation to time scale, including an estimation of the range different biophysical techniques (red) can capture. The methods I have used in interaction studies are in green and orange have been used for both.

By learning more about these motions, where they come from and how they proceed we can learn more about the proteins and their function.

Chemical Reactions in Proteins

Any chemical reactions generally involve some kind of energy exchange. Within living organisms processes that release energy are favorable and biochemical reactions follow the laws of thermodynamics.[10] Gibbs free energy, G, is a thermodynamic potential which can be used to determine if a reaction will occur spontaneously or not. The change in Gibbs free energy (J), ΔG, is calculated by;

$$\Delta G = \Delta H - T\Delta S$$

Equation 3

Where H is enthalpy (J), or internal energy of a system, T is temperature (K) and S is entropy (JK^{-1}), or level of disorder.

If $\Delta G<0$ the reaction is exergonic and spontaneous - energy is released
If $\Delta G=0$ equilibrium - no energy exchange
If $\Delta G>0$ the reaction is endergonic and not spontaneous - energy need to be added

By maintaining a steady state entropy is produced at a minimal rate in living systems[11]. This agrees with the second law of thermodynamics, which in essence say that the entropy in the universe is steadily increasing. However, in order to maintain the stability in cells energy need to be added to the system, even for spontaneous reactions (Figure 4).

The law of mass action explains the relation between the concentration of reactants and the velocity of a reaction in biochemical reactions[12]. The reaction rate is directly proportional to the activity. This assumption is recurring in the calculations of enzymatic activity and protein-ligand binding calculations.

Protein Catalysis - Enzymes

Enzymes are a type of proteins that can catalyze chemical reactions in living organisms. Most reactions would be too slow without catalysis hence this is a very important function of proteins. Catalysts increase a chemical reaction without being consumed. Enzymes are known to be very efficient catalysts, the can increase a reaction by a factor of 10^{20}. The exact mechanism is not known, however according to the transition state theory of enzyme catalysis the enzyme is catalyzing a reaction by forming a more favorable transition state, which decreases the activation energy needed for the reaction to occur.[13] A substrate or substrates bind to the biding site of the enzyme where favorably specific interactions occur. The reaction is promoted in the catalytic active site where the reaction takes place and the product or products are formed. [14, 15]

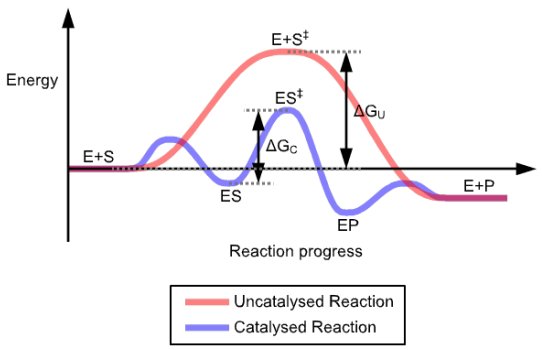


Figure 4. Representation of a biochemical reaction, with and without enzyme. E is the enzyme, S is the substrate and P is the product. ES‡ is the transition state. Less input of energy is needed with the enzyme, however net ΔG will be negative in both cases.

The kinetics of a catalytic process can be described by;

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} ES \xrightarrow{k_{cat}} E + P$$

Equation 4

The Michaelis-Menten equation can be used to calculate the initial reaction rate;

$$V_0 = V_{max} \frac{[S]}{[S] + K_M}$$

Equation 5, The Michaelis-Menten equation

Where;

V_0 = initial velocity (moles/time)

[S] = substrate concentration

V_{max} = maximum velocity
K_M = Michaelis-Menten constant or substrate concentration at $V_{max}/2$

The Michaelis-Menten equation is useful for investigations of the relation between reaction rates and substrate concentration of enzymatic reactions.[16]

Protein - Ligand Binding

As part of their function proteins constantly interact with other proteins and substrates. By understanding these interactions we can reveal more about the function of proteins. Protein kinetics is a useful tool in the investigations of protein folding mechanism, protein-protein interaction studies, when examining enzymatic reactions and when investigating how proteins interact with other substrates, such as drugs or metals. All these interactions can be described with protein kinetics.[14]

Protein – ligand binding can be explained with similar equations as enzyme reactions. The dissociation constant, K_d, is often used to describe the strength of an interaction between a protein and a ligand (which can be e. g. another protein, a drug or a peptide). K_d defines the concentration of ligand at which half of the protein is bound (Equation 3).

$$K_d = \frac{[L][P]}{[LP]}$$

Equation 6

[P] and [L] are the concentrations of protein and ligand respectively and [LP] is the concentration of protein-ligand complex. Binding strengths varies a lot however a strong binding can be considered when K_d<10 nM and a weak binding when K_d<10 μM.

The Hill equation describes saturation of protein-ligand binding;

$$\theta = \frac{[L]^n}{K_d + [L]^n}$$

Equation 7, The Hill equation

Where;
θ = fraction of protein bound to ligand (%)

n = Hill coefficient

The Hill equation is very useful for calculating the fraction of protein bound to ligand, and estimate when saturation is reached.[17]

Protein Folding

When a protein is being constructed in the cell there are a number of reactions and activities taking place. The primary structure of the protein is defined by the order of amino acids, which are covalently bound to each other in the ribosome. The primary structure of a protein contains all information necessary for the higher orders of structure. The secondary structure determines the hydrogen bonds of adjacent amino acids, forming structures as α-helices and β-sheets. With additional hydrogen bonds, disulphide bridges, ionic bonds as well as hydrophilic and hydrophobic interactions being induced within the protein its tertiary structure, or final three-dimensional form, is created. Some proteins also have a quaternary structure, which means that they consist of two or several subunits bound by similar forces. All these processes depend on ambient conditions such as pH, temperature and concentration of substrates. Some of these folding events are happening spontaneously directly after the N-terminal of the protein exits the ribosome while some need help from chaperones and other proteins (cofactors) to speed up these processes.[15, 18]

Protein Motions in the Terahertz Range

It is know that many low frequency motions and vibrations in proteins are occurring within the terahertz range.[19, 20] Low-frequency vibrations of proteins have been vastly investigated together with spectroscopic methods and molecular dynamics (MD) simulations.[21, 22]

In our group we have been working on methods to introduce terahertz radiation to X-ray methods. The goal has been to be able to follow the protein motions induced by the terahertz radiation.

In paper V a pump/probe experiment with these questions in mind is described. By pumping a protein crystal with terahertz radiation while probing it with X-rays, the difference with and without THz-pumping was investigated.

Target Proteins

In this world there are zillions of proteins, although we cannot study them all. In this thesis I have been focusing on two different of proteins, and their dynamics and function.

In paper I, **II** and **III** I have been studying human survivin and its interaction with synthesized peptides from partner proteins. Survivin has been well studied for its involvement in diseases and its involvement with many cellular processes including cell division and apoptosis.

In **paper IV** and **paper V**, pancreas bovine trypsin was crystallized and used for X-ray diffraction experiments, with and without THz-pumping. Trypsin was chosen for our X-ray crystallography experiments, as it is a well-characterized enzyme with a well-known mechanism of action. It is also easy to crystallize and known for diffracting at high resolutions.

Survivin

Survivin is the smallest member of the Inhibitor of Apoptosis Protein (IAP) family, with 142 amino acids and a mass of 16.5 kDa.[23] It belongs to the inhibitor of apoptosis protein (IAP) family. A common feature of members of the IAP family is that they possess a zinc-binding baculoviral inhibitory (BIR) domain. The BIR domain of survivin is connected to a long α-helical C-terminal via a short loop. This region is essential for the homodimer formation of survivin.[24, 25]

Survivin has been thoroughly studied for its involvement in human cancers, inflammatory diseases and in autoimmune diseases in humans.[26] It is indirectly involved in apoptosis[27] and it has an important role in cell division where a monomer of survivin takes part of the chromosomal passenger complex (CPC) together with Aurora B kinase, INCENP and borealin.[28]

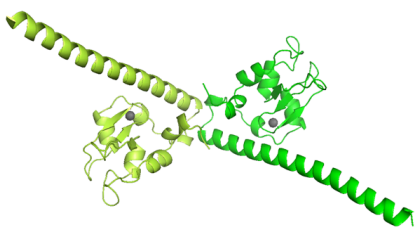


Figure 5, Crystallography model of survivin dimer, PBD: 3UEI

Survivin Interaction Partners

Survivin and its function have been extensively studied. The three dimensional structure of human survivin was solved with X-ray diffraction and published by two research teams already in the year 2000.[23, 29] Survivin was successfully co-crystallized with a part of borealin and INCENP in 2007 and interactions between the proteins were demonstrated.[30] In 2011, Jeyaprakash with colleagues solved the crystal structure of survivin bound to a short peptide of the N-terminal sequence of Sgol1; Ala1-Lys2-Glu3-Arg4. In their study they suggested that the sequence of Sgol1, which is common in mitotic proteins, is recognized and bound by survivin.[31]

Direct interaction has also been reported between survivin and smac/Diablo, which is also involved in apoptosis.[32, 33] Another direct interaction has been found between histone H3 and the monomer of survivin in the chromosomal passenger complex.[34]

We used this knowledge in the search for additional interaction partners of survivin. A peptide microarray assay including peptides from proteins already known to interact with survivin as well as peptides from proteins that were thought to be interacting or which are included in the same cellular processes as survivin.

The Chromosomal Passenger Complex

The chromosomal passenger complex (CPC) is composed by the four proteins Aurora B, INCENP, Borealin and survivin and this complex play a crucial role in cell division. Aurora B kinase operates as the enzymatically active part of the CPC whereas the functions of the other three proteins are regulatory and targeting. The localization bundle of the CPC is formed by the N-terminus of INCENP linked to the baculoviral inhibitory repeat (BIR) domain of survivin and the carboxyl terminus of borealin.[30]

Shugoshin Proteins

Shugoshin proteins, or protecting proteins, fulfill an important function in cell division by protecting centromeric cohesion from premature cleavage during meiosis and mitosis. Shugoshin 1 (Sgol1) is operating during mitosis in vertebrates. Sgol1 protects cohesion from being prematurely removed from chromosome during mitosis.[35] Shugoshin 2 (Sgol2) is crucial during meiosis. When cohesin along the chromosome arms is cleaved, Sgol2 protects centromeric cohesin from being prematurely cleaved from the centromeres.[36] Furthermore,

shugoshin proteins are believed to be involved in several processes during cell division, including recruitment of the CPC.[37]

The Polycomb Repressive Complex 2

Polycomb Repressive Complex 2 (PRC2) is a histone methyltransferase, which is responsible for methylation of lysine 27 on histone 3 (H3K27) and thus the regulation of gene expression by epigenetic modification. When H3K27 is trimethylated (H3K27me3) the target genes are maintained in a repressed state, preserving the identity of the cell.[38]

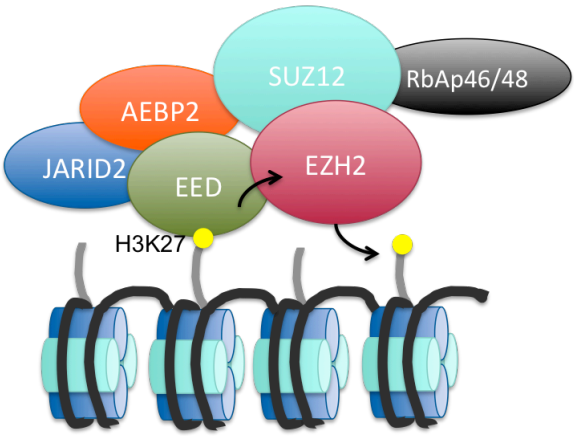


Figure 6, PRC2 and its mechanism with histone 3

PRC2 have been extensively studied for its presence in various cancers.[38, 39] The core of PRC2 involves four subunits; enhancer of zeste homologue 2 (Ezh2), embryonic ectoderm development (EED), suppressor of zeste 12 (Suz12) and retinoblastoma binding protein 46/48 (RBAP46/RBAP48). The structures of the subunits and how they interact have been well studied.[40, 41] Ezh2 contains the enzymatic component of the PRC2, this is where the trimethylation of H3K27 is taking place. The seven WD40-repeat containing protein EED is the smallest component of the complex. EED has an aromatic cage, which consists of four aromatic side chains, Phe-97, Tyr-148, Trp-364 and Tyr-365. This motif binds the already triple-methylated H3K27, stimulating activity of the complex and autocatalysing the methylation of other histone subunits by Ezh2.[42] Suz12 enfolds the other subunits contributing to the stability of the complex and keeping them together.[40]

Jumonji and AT-rich interaction containing domain 2 (Jarid2) and adipocyte enhancer-binding protein 2 (Aebp2) have been reported as important cofactors of PRC2. Jarid2 is mimicking a methylated H3 tail, which triggers PRC2 activity. Aebp2 on the other hand is mimicking an unmodified H3 tail.[41]

Trypsin

Trypsin is a small enzyme, with a size of ~23 kDa, located in the intestines where it has an essential role in the degradation of proteins. It belongs to the group of serine proteases, which are enzymes that cleave peptide bonds in proteins. They all have a serine in their active site, although the site attracts and cleave different residues. Hence, these proteins are useful in digestion of proteins.

The catalytic site of trypsin consists of the "catalytic triad", which involves histidine-57, aspartate-102 and serine-195.[43] The catalytic triad is present in the catalytic site of many enzymes[44], however in trypsin it hydrolyzes and cleaves peptide bonds on the c-terminal side of lysine and arginine residues. [44-46]

We have been using cationic trypsin for our experiments and crystallized it together with benzamidine, which is an inhibitor of trypsin. It is expressed in pancreas and can be purified with high yields. It is a well-studied protein that is easily crystalized. Crystals of trypsin are rectangular and well ordered and known to diffract at a really high resolution with a reproducible result, up to 0.75 Å.[47]

Scope of the Thesis

In my research a combination of conventional biophysical characterization methods have been used in combination with less established methods. Understanding the dynamics of a protein is the key to comprehending the function of it. The scope of this thesis has been rather broad but with the overall goal to improve the understanding of protein dynamics. In this search we have used recombinant survivin for interaction studies and pancreas bovine trypsin for dynamical crystallography studies.

The results in **paper IV** and **paper V** are from the same experiment in FemtoMAX, an ultrafast beamline at the MAX IV short-pulse facility (SPF) (Lund, Sweden). Diffraction was collected from trypsin crystals in room temperature, with and without simultaneous laser radiation with THz frequency. The goal with paper IV was to describe how an X-ray experiment can be performed at FemtoMAX whereas in paper V the effect of the THz pumping is described.

Different approaches were used to examine the interaction survivin and various interaction partners in **paper I**, **paper II** and **paper III**. In paper I an interesting connection between survivin and PRC2 is described and investigated with several different interaction methods. We have used machine learning as part of analysis of the results. In paper II and III we use interaction between survivin and a peptides of sgol1 and sgol2 to describe a Bayesian statistical approach for analysis of MST-data.

Methodology

Protein Production

A large amount of protein is often needed for biophysical studies. Some proteins can be easily extracted from natural sources, such as trypsin or lysozyme. For most proteins this would be a very resource consuming way to produce protein and larger amounts of recombinant protein is commonly produced by overexpression and production in various types of cells. Every protein is different and there are many things that need to be considered when developing a production protocol. Which expression host to use for expression, bacteria, yeast, mammalian cells or insect cells? Expression in bacteria is the least time consuming and is often enough for small soluble proteins. Usually a bacterial plasmid is used as vector for insertion of the gene into the cells. The choice of plasmid depends on which properties, e.g. specific antibiotics resistance, fusion tag or cleavage site, will be included into the construct.[48]

The gene of survivin was inserted to the pHIS8 expression vector used by Verdecia et al [29], which generates survivin with an N-terminal 8xHis-tag and a thrombin cleavage site. The construct was transformed into *E.coli* BL21(DE3) star (Merck) cells for expression.
Expression of wild type survivin was performed by growing the cells in LB, 37°C, until an OD_{600} between approximately 0.6 and 1 was reached. The cells were induced by adding 500 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and protein was overexpressed during 4 hours in 30°C. With this procedure a final yield of around 10 mg pure protein per liter of cell culture have been obtained.

Expression of Isotope Labeled Survivin

In order to be able to observe protein with 2D NMR it is necessary to exchange some of the atoms to NMR active isotopes. Commonly ^{15}N and/or ^{13}C are used instead of the prevalent ^{14}N and ^{12}C. ^{1}H can also be exchanged with ^{2}H to hide the hydrogen in the NMR spectrum and obtain a clearer result of the protein spectrum. When producing isotope-labeled protein it is important to exclude all naturally occurring sources of the atoms.

Instead of LB medium, M9 minimal medium is typically used as growth medium for isotope-labeled protein. With a minimal medium it is easier to control the conditions and exchange

the components depending on how which labeling is sought. M9 medium is based on a phosphate buffer (M9 salt) to which minerals and vitamins are added.[49] A minimal medium contains just enough components for the bacteria to grow and expressing in M9 medium normally reduces the growth rate and protein yield in contrast to expressing in LB medium.

For isotopic labeling ammonium is exchanged for $^{15}NH_4Cl$ to obtain nitrogen-labeled protein and for carbon-labeling glucose is replaced by ^{13}C-glucose. Water can be exchanged to D_2O if the ambition is to replace ^{1}H to ^{2}H.

I used the original construct of survivin and developed a new protocol for expression and production of ^{15}N-labeled and $^{13}C^{15}N$-labeled survivin in M9 medium, described in paper I.

Protein Purification

It is important to start with a pure sample in order to achieve a satisfying and reliable result in functional and structural studies of proteins. This is achieved by separating the target protein from any other proteins and other impurities that may be present in the sample. Commonly, several purification steps are needed to obtain a pure, monodisperse protein. Our purification protocol of recombinant survivin includes a first purification step with affinity chromatography followed by dialysis and a last purification step with size exclusion chromatography (SEC).

Cell Lysis

In order to access the protein the cells need to be disrupted. This can be done mechanically, e.g. with high pressure or bead mill, or it can be done non-mechanically with enzymes, osmotic shock or with pH change.[50] I used a combination of lysis buffer and sonication to disrupt the cells.

Immobilized Metal Affinity Chromatography (IMAC)

IMAC is an affinity chromatographic method and a commonly used technique for the first purification step of tagged proteins. IMAC columns are preloaded with different types of metal ions for which proteins possess affinity. Most common is to use columns loaded with Ni^{2+} together with His-tagged proteins but sometimes a better result can be achieved with other metal ions, such as Zn^{2+} and Co^{2+}.[51]

A 5 ml HisTrap™ FF column from Cytvia was used as a first step of the purification. The column is prepacked with charged Ni Sepharose® 6 Fast Flow, which is specially designed for purification of His-tagged proteins, which possess high affinity for Ni^{2+}. Tagged proteins are retained in the column whereas untagged proteins will flow through the column or be washed out with low concentrations of imidazole. The His-tagged protein can then be eluted in a controlled manner with a buffer by introducing a gradient or a stepwise increasing concentration of imidazole. Imidazole is used because of its ability to compete with the His-tag for the binding to the metal ions in the resin. Wild type survivin was eluted at around 20% imidazole whereas isotope labeled survivin was eluted at approximately 40% imidazole (of 500 mM imidazole stock)

Dialysis

After IMAC it is important to remove all the imidazole from the sample. This was done by insert the sample to a dialysis bag and placing it inside a buffer without imidazole. The dialyzed protein was the centrifuged in order to separate the protein in solution from precipitated proteins.

Size Exclusion Chromatography (SEC)

As a last purification step gel filtration, or SEC, was used to remove remaining impurities from the sample. The dialyzed protein was the centrifuged in order to separate the protein in solution from precipitated proteins.

With SEC macromolecules are separated by size and molecular weight. A gel filtration column is packed with porous beads in which smaller molecules will be retained while larger molecules bypass the porous beads and travel faster through the column. This means that larger molecules will be eluted prior to smaller molecules and the size of the eluted molecule can be estimated.

A Superdex™75 10/300 GL column from Cytvia was used for gel filtration of the proteins. It is a prepacked high performance glass column, which provides a good separation of proteins between masses of 3-70 kDa, it covers the common oligomeric states of survivin as the monomer has a molecular weight of around 16 kDa and the dimer is around 32 kDa.

Biophysical Methods for Characterization of Proteins

In this work I have been using different methods for characterization of the proteins. I have performed X-ray crystallography diffraction experiments with trypsin at the traditional crystallography beamline BioMAX and at the more unconventional short pulse facility, the FemtoMAX beamline, both situated at the MaxIV synchrotron in Lund, Sweden. With survivin I have conducted SAXS experiments at the beamline BM29 of the ESRF synchrotron (Grenoble, France) and at P12 beamline of PETRA III synchrotron (Hamburg, Germany). I have also preformed Nuclear Magnetic Resistance (NMR) spectroscopy measurements at the Swedish NMR Centre (Gothenburg, Sweden). I'll give a brief introduction of these methods as follows.

X-ray Methods for Structure Determination

X-rays are of very short wavelengths, 0.01-10nm, which is suitable for studying objects at atomic resolution (~1 Å or 0.1 nm).

X-ray crystallography diffraction is still the most common method for biological structure determination with more than 160 000 structures in the protein data bank (PDB)[52] at the time of writing (more than 188 000 in total). With X-ray diffraction a protein structure can be determined with near-atomic resolution.

Small angle X-ray scattering (SAXS) is another X-ray method for protein characterization. SAXS measurements do not reveal atomic structural details, although the method generates information about size, shape and oligomerization state of a protein. This method is often used as a complement to other structure determination methods, such as cryo-EM, NMR and X-ray crystallography.

Synchrotron Light Source

As we need X-rays with a high photon flux, or brilliance, to get good resolution of protein models commonly synchrotron light sources are used for generating X-rays. Electrons are accelerated by an electric field in a linear accelerator (LINAC) and then injected into a storage ring where they are kept with constant speed for long time. In the storage ring there are different type of magnets that "bend" or accelerate the electrons and photons are emitted from the electrons.[53] The photons, or synchrotron radiation, are then directed into beamlines with different purposes, such as X-ray scattering or crystallography diffraction

experiments. In MaxIV (Figure 1) there is an option to lead the electron to one of the two storage rings or directly to the short pulse facility (SPF) and FemtoMAX.

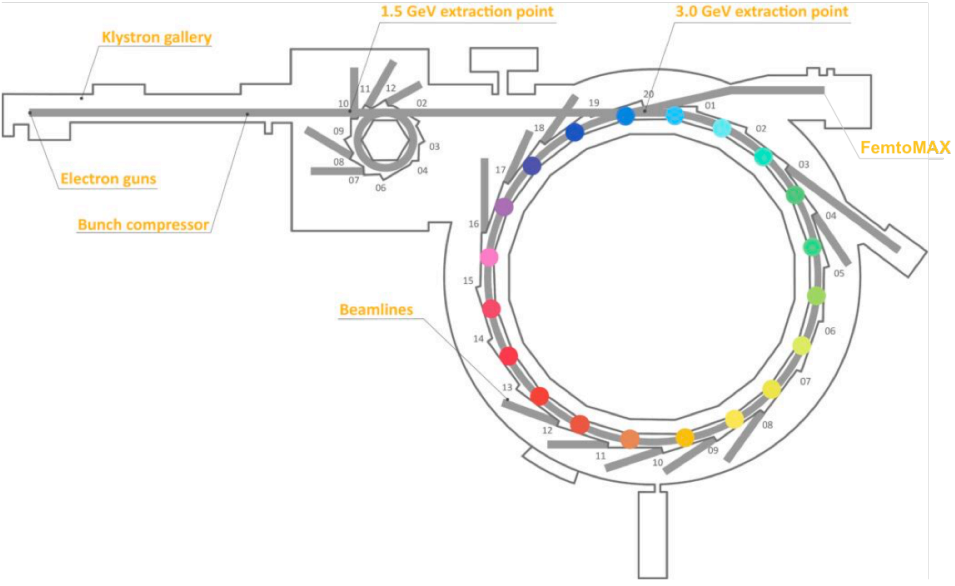


Figure 7, Schematic representation of the synchrotron facility of MAXIV

X-ray Crystallography Diffraction Experiments

The most popular method for determination of protein structures is still X-ray crystallography diffraction, with which a high-resolution model of a protein structure can be obtained.

One molecule itself cannot be studied by X-ray crystallography as it does not possess translational symmetry. Within a crystal many protein molecules are arranged in a specific symmetrical array and when the X-rays penetrate the crystal and hit several molecules the diffraction will be amplified at certain angles. The diffraction pattern is recorded with a detector and from the diffraction pattern, by using Fourier transform, the amplitude of each reflection is obtained. By providing information of the phases (e.g. molecular replacement) the electron density is calculated and a model of the protein can be created.[54, 55]

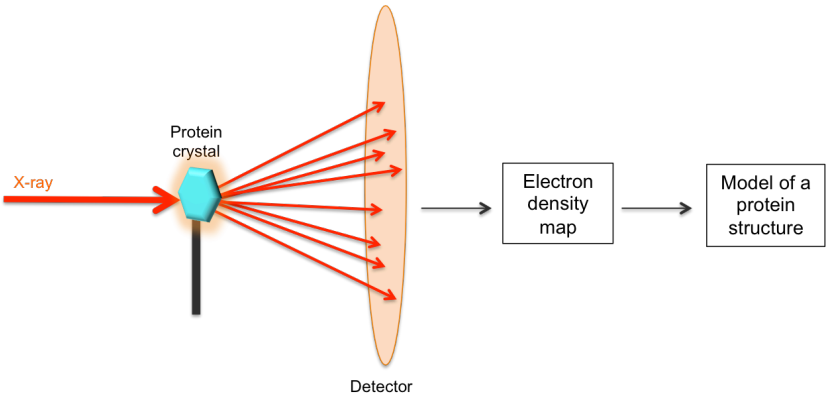


Figure 8, Schematic drawing of an X-ray crystallography diffraction experiment

Sample Preparation - Protein Crystallization

Protein crystallization can be a challenging task and not all proteins have the ability to crystallize satisfactory. It can be the limiting step and it was one of the reasons that we chose to work with trypsin, which in known to crystallize with in well-ordered arrays.[56]

Vapor diffusion is the most commonly used process that is applied in different crystallization techniques. The principle is that a liquid drop is prepared with an aqueous solution containing protein and precipitant, placed in a closed container and allowed to equilibrate with a larger volume aqueous solution via vapor diffusion. I used the hanging drop method [54, 57] for crystallization of trypsin that is described in paper IV and V, where the drop is placed hanging on the inside of a cover slip. The drop should not be in direct contact with the reservoir solution and that the container has to be properly sealed. Water will slowly evaporate from the drop to the reservoir solution until the precipitant concentration is the same in the drop as in the reservoir and the crystal is formed.

The growth conditions typically need to be optimized for the target protein. Various conditions, such as pH, temperature and the constituent concentrations of salt, precipitant and protein, commonly need to be tested. For this reason, crystals are commonly grown in a multiple well plate where different conditions can be set up simultaneously.

It is not fully understood how protein crystal form, however one theory is that it is necessary to reach an energy barrier similar to that of a chemical reaction.[54] A crystallization phase diagram (see Figure 7) is often used to schematically illustrate the crystallization process in two dimensions. At first, the concentration of both protein and precipitant are low and the process is in the undersaturated region. As water gradually diffuses from the drop the protein and the precipitant concentrations increases It is important that the supersaturated region is reached for the nucleation process to start. If the precipitation zone is reached first the protein will precipitate or aggregate before the nucleation process begins. Ideally the labile zone, where the supersaturation is large enough for the nucleation process to begin is reached. Once the nuclei formation starts the protein concentration in the solution decreases and the metastable zone is reached. In the metastable zone no more nuclei are formed, the conditions are ideal for crystals to continue growing.[54, 58]

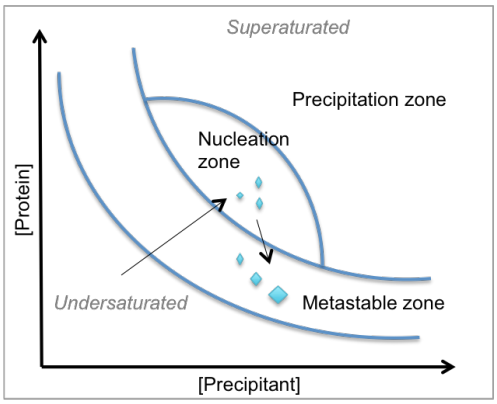


Figure 9, Crystallization phase diagram

Geometry in X-ray Diffraction

In a protein crystal the molecules are arranged in an ordered array in a lattice. The asymmetric unit is the smallest unique part of the crystal that cell by symmetry operations can generate the unit cell. A unit cell is the smallest unit of the lattice, which possesses the full symmetry of the crystal. A unit cell is defined as a geometric figure with the axes a, b, c and with the unique angles α, β, γ. 13 different lattice types are described, however I will not go more into detail. Three integer indices *hkl* identify the planes within a unit cell.[54]

When X-rays hit atoms, the electrons surrounding the atom will start to oscillate and reemit the X-ray (or photon) with the same energy, this is an example of elastic scattering. If several similar atoms in an ordered array are scattered at specific angles the amplify each other, this phenomena is called constructive interference or diffraction. The scattered waves must be in alignment for constructive interference to occur. In order for constructive interference to occur the scattered waves must be in alignment – the second wave must travel a whole

number of wavelength, λ/2 on the incident side and λ/2 on the scattered side. When we have constructive interference, Bragg's law is applicable:

$$n\lambda = 2dsin\theta$$

Equation 8

Where;

d = distance between planes of lattice

θ = the angle of the incident and the scattered X-ray beam

n = integer (whole number of wavelength)

λ = wavelength

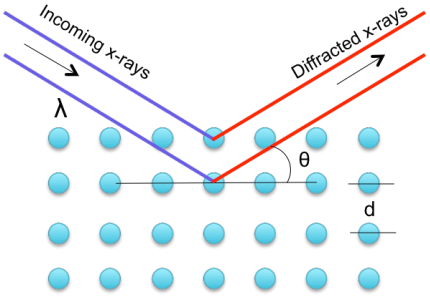


Figure 10, Bragg's law

The intensity of each reflection *hkl* depends on electron density distribution in the crystal. Since the diffraction pattern is the reciprocal lattice of the crystal lattice an inverse Fourier transform is performed to define the real space diffraction before further analysis of the data.

X-ray Data Collection

In conventional X-ray crystallography data is collected from the crystal by rotating it at the same time as is irradiated with a monochromatic X-ray beam while a detector records the diffracted photons from the crystal.

In Laue diffraction, polychromatic X-rays are used. This is commonly used in time-resolved crystallography when the purpose is to observe the dynamics the protein in the crystals. It is useful in observation of the fast rate dynamics in proteins. With the femtosecond pulse of X-ray free electron laser (XFEL) sources the crystal is often completely destroyed and hence a new crystal is needed for every shot and another approach is needed for analysis.[59]

Once the data is collected the analysis and the structure determination is commonly performed with any of the free available software suites, such as CCP4[60] or PHENIX[61].

X-ray Crystallography at FemtoMAX

The FemtoMAX beamline at the MAX IV short-pulse facility (SPF) (Figure x) generates very short and intense X-ray pulses. The pulse lengths are on the time scale of molecular vibrations (100 fs) at wavelengths matching interatomic distances.[62]

FemtoMAX is still under development and we were the first users to perform an X-ray crystallography experiment. So far they have a repetition rate at 2Hz, going for 10 Hz with the goal to reach 100 Hz (compare too 100 MHz in the storage ring). X-ray energy can be varied between 3-12 keV (λ=1-4 Å), with the goal of acquiring an even wider energy range.

This beamline has been designed for performing pump-probe experiments, which makes is useful in time-resolved experiments for investigations of ultrafast dynamics in proteins. With the in-house laser the sample can be pumped with different time delays before it is being probed with X-rays.

One of the advantages in using FemtoMAX for time-resolved experiments is that there is basically no radiation damage over time. In principle the same protein crystal can be used for hours and pumped over and over again with the end station laser while diffraction is detected and recorded. This could be compared to XFEL where one crystal is destroyed per X-ray shot, and thousands of well-diffracting crystals are needed for one experiment.[63]

Together with the beamline team we developed and set up a method for collecting crystallography X-ray diffraction data, explained in paper IV. We also managed to perform initial time-resolved experiments where the crystals were pumped with an end station laser with radiation of THz frequency, paper V.

The multilayer monochromator was set to 11.15 keV. All the measurements were performed at room temperature, keeping the crystals in a plastic capillary. 100 images were collected for every 0.1° of rotation and the images were summed up for further analysis. In the end there were around 1000 summed images for each crystal and X-ray Detector Software (XDS) was used to process (indexing, refinement, integration, scaling and merging) the images. The phases were obtained by molecular replacement with PDB 4I8H. *Phaser[64]* from the PHENIX software suite[61] was used to solve the structure and the software *phenix.refine* was used for refinement.

Small Angle X-ray Scattering (SAXS)

SAXS is another common X-ray method for investigating the size and overall shape of a protein, by measuring the scattering. As the name suggests the measurements are conducted in solution, which is beneficial for investigations of proteins as close as possible for their

natural state. The drawback in comparison to e.g. NMR and X-ray crystallography is that the obtained resolution of the protein is very low. This method is most suitable for investigations of shape and molecular weight of biomolecules, and for determining oligomerization and quarternary structure of a protein.[65] SAXS is an excellent complement to other biophysical characterization methods and is often used in combination with NMR, X-ray crystalloraphy and cryo electron microscopy in the exploring of structure and function of proteins.[66, 67]

With the ability to study shape and size of a particle, SAXS is also a useful technique for studying interactions and forming of complexes. [68] This makes it also a useful technique for performing time-resolved experiments, where some properties or movements of the protein are monitored.[69]

SAXS in Protein Structure Biology

The protein in buffer solution is irradiated with a monochromatic X-ray beam (commonly <0.2 nm). Since the protein molecules are free in solution the scattering of the molecules will come from all directions of the molecule, the averaged scattering from all molecules in the sample is detected. This is why we can only get low-resolution data with SAXS. The scattering is elastic, which mean that the incoming energy will be scattered with the wavelength conserved. Intensity changes at small angles; 0.1-10 degrees are recorded with a detector. The two-dimensional pattern of the scattering is related to the overall size and the shape of the protein molecule.[67]

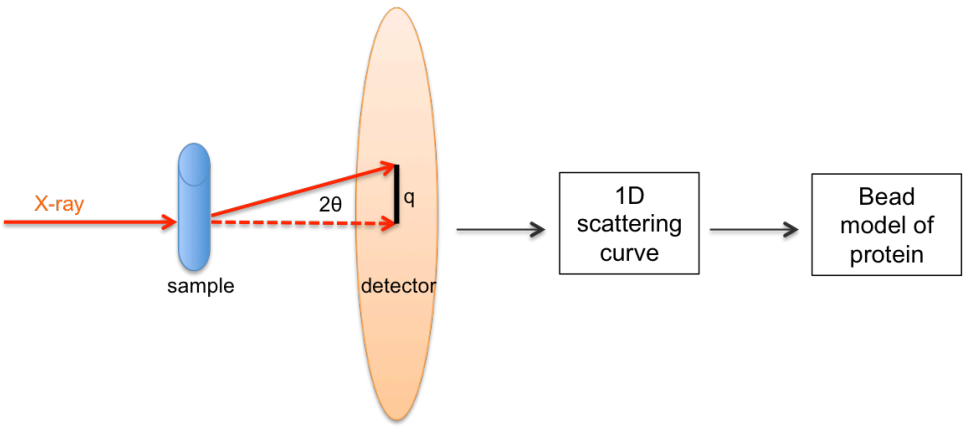


Figure 11, SAXS setup. The X-ray beam is monochromatic and the scattering is measured at small angles, 2θ, generating a two-dimensional pattern which is recorded by the detector. An one-dimensional scattering curve, I(q), is

generated from the one-dimensional scattering pattern. From the one-dimensional curve a bead model of the protein can be calculated.

The scattering pattern is isotropic, which means that the scattered intensity is not depending on the direction of the molecule. Because of this the recorded two-dimensional pattern from the detector can be averaged into an one-dimensional scattering intensity curve, I(q). q is the momentum transfer (Equation 2), where 2θ is the angle between scattered and incident radiation and λ is the wavelength of the radiation.

$$q = \frac{4\pi sin\theta}{\lambda}$$

Equation 9

Sample Preparation and Measurement

Sample preparation for SAXS is by far simpler and more straightforward than for e.g. X-ray crystallography or NMR measurements. However, it is very important to have a monodisperse and evenly distributed sample in solution and this can sometimes be difficult to achieve. Of course it is of great importance that the measured buffer match the buffer of the sample and ideally a dialysis is performed before measurements to ensure this.[70]

A combination of SAXS and size exclusion chromatography (SEC-SAXS) is one way to ensure buffer match, it means that the sample is run through a gel filtration column in line with the SAXS measurement. The eluted sample is directly led to the capillary where hundreds or thousands of images, or frames, are collected. This way one can ensure a more monodisperse sample and choose frames from the fraction of interest for analysis of the protein. The buffer subtraction can be performed with frames from the buffer before the sample peak.[70]

The fastest and most common is to perform batch measurements with a dilution series of the protein. Which concentrations and how many dilutions depend on the protein but at least three different concentrations (approximately 1-10 µl) should be measured to ensure that there is no aggregation or repulsion in the sample. With this method only small amounts of protein is needed, as it is enough with 30 µl per measurement. The sample is flown through a

small capillary while around 10 frames are collected. Buffer is flown through the capillary and frames are collected before and after each sample.

Time-resolved experiments can also be performed with SAXS. In this type of experiments the sample it is common with a continuous flow setup, which mean that the sample is flowing in a capillary while data is collected. A trigger, such as light or temperature is introduced, and the sample can be continuously triggered or a pump-probe setup can be used.[69]

Evaluation of SAXS data

Usually the software at respective synchrotron provides processed data from SAXS measurements. For further analysis there are several commercially available software, In my analysis I have been using is the ATSAS software suite, which have been developed by D. Svergun and his team. ATSAS include a number of programs for processing, visualization, analysis and modeling of SAXS-data.[71, 72]

A number of calculations and plots are provided by the software PRIMUS, which is included in the ATSAS software suite) and used for evaluation and analysis of the protein shape and size from obtained SAXS-data described further in the thesis.

Only a very small part of the incident photons are scattered by the protein molecules, and there scattering from the buffer and capillary are also recorded. An one-dimensional scattering curve, log I(q), is extracted from the two-dimensional scattering curve.

The scattering is always measured for the sample and the buffer separately and the difference curve is calculated and the scattering of the protein is obtained (Equation x).

$$I_{sample}(q) - I_{buffer}(q) = I_{protein}(q)$$

Equation 10

This way eventual scattering from the buffer, capillary and other possible background disturbances are eliminated.[70]

The contrast curve is used to calculate the size and the shape of the molecule. From the scattering curve a bead model can be calculated, with an estimation of the size of the protein.

The scattering intensity at low q contains information about the size and shape of the molecule while in the higher q there is more information about the finer structural details.

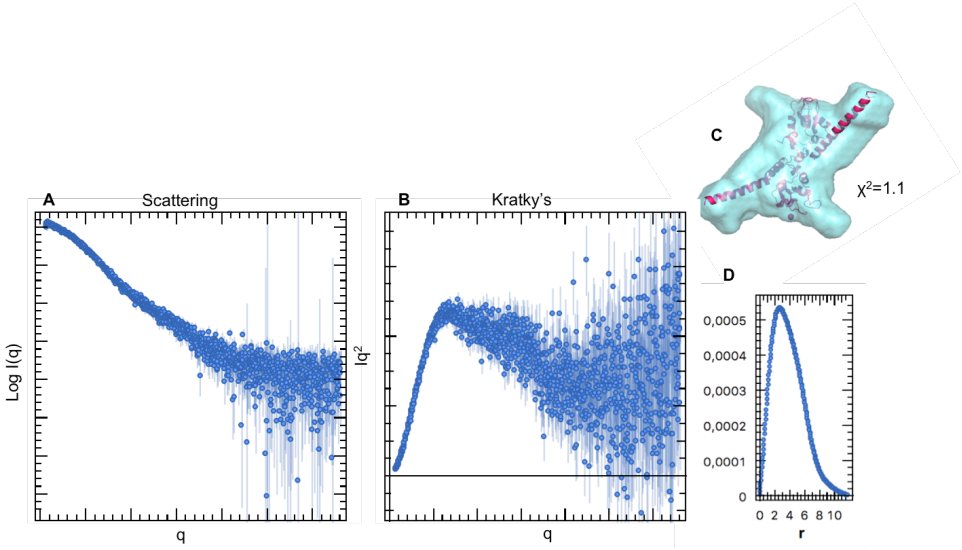
The Kratky plot, $Iq^2(q)$, provides information of how well folded a protein is. A perfectly global protein would follow Porod's law and render a Gaussian shape of the curve, whilst for an unfolded or elongated protein the curve would have a plateau or increase with higher q. This evaluation comes from the assumption that the scattering intensities of macromolecules will behave as Gaussian-like coils and will plateau in a $Iq^2(q)$ plot, originating from Debye's scattering formulation of a Gaussian coil.[73]

The Guinier analysis is performed at very small angles of the scattering curve. Guinier plot, log $I(s^2)$ and radius of gyration, $R_{g,}$ are directly related to the mass and the shape of the molecule. The Guinier plot also reveals deviations from ideal solution, such as aggregations, in the sample.

The distance distribution function, P(r), shows the distribution between all pairs of point in the molecule. From this D_{max} is obtained, which is the maximum distance (longest chord) within the molecule.[72]

Characterization of Survivin with SAXS

I have performed several SAXS measurements with survivin, with and without ligand. Both in batch mode and with size exclusion HPLC, at BM29 in ESRF, Grenoble – France, and in P12, PETRA III, Hamburg - Germany. Here I will present the result from a measurement that was used as a complement to the result in paper I.



Conc (mg/ml)	Rg (nm)	I(0)	Guinier Rg	P(r) Rg	sRgmax	Dmax	Vp	MW (s&s)	MW (bayesian)
1.1	2.98	0.034	3.09	3.09	1.3	11.88	59.32	37933	34575

Figure 12, Result from analysis with the software suit ATSAS. A is displaying the scattering of the protein (buffer subtracted), B the Kratky's plot and C is a model of the protein overlaid with a crystallography model, 3UEI. D is the distance distribution function, P(r). The table displays some statistics from the analysis.

The measurement was performed with three concentrations of survivin, 1.1 mg/ml, 2.4 mg/ml and 4.8 mg/ml. The scattering curve together with the estimated size calculations from the three concentrations indicated that there might be some concentration dependence. There might be a mix of dimer and some larger formations with higher concentrations. Therefore this analysis was done only with the measurement from the sample of 1.1 mg/ml. The bead model was created with Damaver from the ATSAS suit. CRYSOL software has been used to fit the model to PDB crystallography model 3UEI, the overlay was performed with SUPCOMB, both from the ATSAS suit[72]. It should be added that the SAXS model comes from survivin with an N-terminal 8xHis-tag whereas this is not included in the crystallography model.

Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR spectroscopy is since long widely used for smaller molecules and its popularity in biological structure determination is increasing. During the years, with development of two-dimensional and three-dimensional NMR together with stronger magnets and increased computer power NMR has increased substantially in the field of biological structure

determination.[74] At the moment there are more than 13.000 structures solved with NMR deposited in the PDB.[75]

NMR Basics

The measurements are carried out in a strong magnetic field where the resonance transitions between magnetic energy levels. Electro magnetic radiation of a specific frequency is applied to the sample; the signal is absorbed by the nucleus, which resonates with different frequencies depending on the type of atom, the neighboring atoms and the relation in between.

Atom nuclei have a specific spin quantum number that depends on the number of protons and neutrons they contain, and if the net is odd or even. In short, all nuclei with odd mass number have a non-integer spin quantum number that mean that they possess a magnetic moment, which means that they will be affected by an external magnetic field, B_0, thus they are detectable with NMR. While nuclei with even mass number have a spin of zero or an integer and do not possess a magnetic moment, hence they will not be detectable with NMR.

Every isotope has its specific magnetogyric ratio, γ, which is directly proportional to the resonant frequency and is used in the calculation of Larmor precession frequency, ω.

$$\omega = \frac{\Upsilon}{2\pi} B_0$$

Equation 11

Where;
ϒ = magnetogyric ratio (MHz/T)
ω = frequency(MHz)
B = Magnetic field strength (Tesla)

When the sample is placed in a strong magnetic field the nuclei, which possess a spin, will line up in the field and start to precess (spin) in the z plane. A short radiofrequency pulse (Rf) is applied right to the B_0 and the nuclei will flip down to the xy plane and continue to precess unanimously. Rotating magnetization vector induces a signal, free induction decay (FID), which depends on the nuclei but also on its surrounding environment (electron density distribution), hence each nuclei will give rice to an individual signal. The FID is a

representation of the wave in the time domain and a Fourier transform is performed to convert the signals to the frequency domain and visualization in an NMR spectrum.

Precession or Larmor frequency equation dependent on the strength of the magnetic field and the gyromagnetic ratio which is specific for a given isotope. T1, spin-lattice, is the time for one nucleus to return to thermal equilibrium state of the spins. T2 is when the precessing nuclei fall out of alignment with each other and stop producing FID, observed signal. T1 and T2 depend on the rate of molecular motion, gyromagnetic ratios of neighbors in resonance.

Protein NMR

For smaller molecules it is usually enough to identify the composition with a one-dimensional spectrum of where every hydrogen will give a unique signal. Since protein are large molecules with many nuclei another one or two dimensions are commonly added and a two- or three-dimensional spectrum is recorded. In this case it is important to label proteins with detectable isotopes, such as ^{15}N and ^{13}C. In two-dimensional NMR the connections between hydrogens and nitrogens are used to interpret the spectrum. Carbon can be labeled to add an additional dimension to the spectrum and to facilitate the assignment of the peaks.[74]

$^{1}H^{15}N$ HSQC (heteronuclear single quantum coherence spectroscopy) [76] is a commonly used two-dimensional NMR spectroscopy that was performed on survivin with and without ligand in paper I (Figure 13 and 16).

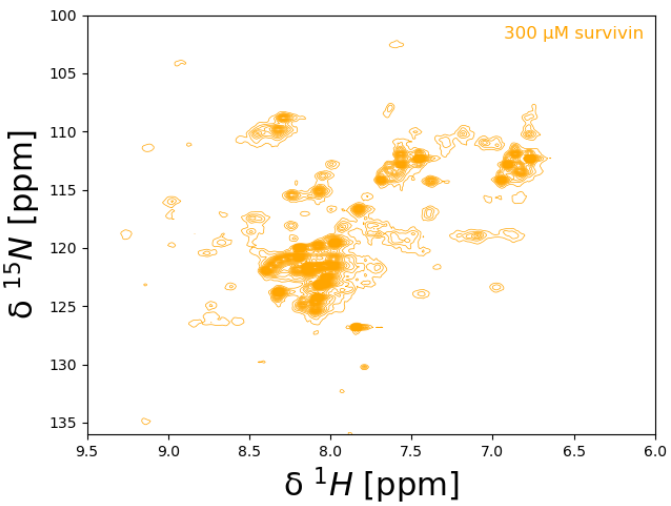


Figure 13. 2D [^{15}N,^{1}H] NMR spectra of 300 μM survivin

Survivin Interaction Studies

One part of my thesis has been to investigate and uncover information about survivin and its function. By studying the interaction between a protein and other proteins or with DNA one can reveal much about its function. In the search for interaction partner proteins several interaction methods were used, by my research group and together with collaborators.

Our collaborators performed a chromatin immunoprecipitation sequencing (CHiP-seq) with antibodies against survivin to investigate which genes it binds to. This method can provide a good indication that the protein would also interact with the protein corresponding to the interacting gene.[77] In addition to that we preformed a protein peptide microarray binding assay with the purpose of screening for interactions between survivin and several potential partner proteins simultaneously.

The result of the CHiP-seq and the peptide microarray together provided identification of some interesting interactions. For further investigation of these interactions longer peptides of the interacting partners were designed and used in further interaction studies with biolayer interferometry (BLI) and NMR titrations specified in paper I. The peptides in paper III were designed from the result of a previously performed peptide microarray and their interaction with survivin were studied with microscale thermophoresis (MST).

There are advantages and disadvantages of all protein interaction methods and in this thesis I am only presenting a selection of available methods. A peptide microarray is a good choice for screening of many binding partners, however it will only indicate if there is a binding or not and reveals nothing bout the binding affinity. The intensity of the fluorescence presents an indication of a strong or week binding between the peptide and the protein and it provides a good basis for further interaction studies.

The dissociation constant, K_D, which provides information about the binding affinity, can be calculated from the results of BLI and MST. In addition to that BLI provides reliable information about kinetic rates, k_{on} and k_{off}.[68] BLI is a label-free bio-sensing method, which is widely used for real-time studies of kinetics and interactions of biomolecules. The procedure is relatively fast and not much sample preparation is needed. One of the binding partners is bound to an optical biosensor (ligand) and the other binding partner is in solution

(analyte). The interference pattern is measured through the biosensor with and without analyte and the result is used to determine whether there is an interaction or not between the two biomolecules.[78]

MST has the advantage that very small amounts of the components are needed, even though the result may be less reliable than with BLI. The analyte (ligand) is titrated into the target (usually protein) and the thermophoresis is measured. The target protein often need to be fluorescently labeled in ordered to be recorded with MST, if it doesn't contain enough aromatics to perform label-free MST.[79, 80]

NMR spectroscopy is another method that can be very powerful in the investigations of protein-ligand interactions. By titrating the ligand into isotope-labeled protein while measuring with HSQC-NMR more details about the interaction can be revealed. If the peaks of the spectrum are assigned to the amino acids of the protein this method can reveal which part of the protein that interact with the peptides. [76, 81] For this type of study large amounts isotope-labeled protein is needed, which makes it a less accessible method than the previous mentioned.

Protein Peptide Microarray Binding Assay

A peptide microarray provides a high-throughput analysis of interactions between biomolecules. The technique origins from DNA microarrays, which have been extensively used in the exploration of DNA-protein interactions.[82, 83] There is a diversity of variants of microarrays on the market; here I will describe the alternative we used which was ordered from PEPperPRINT (GmbH, Heidelberg, Germany). We ordered a customized PEPperCHIP Peptide Microarray on which a total of 5,393 x 2 peptides from 32 proteins were printed.

Designing peptides with a length of 15 amino acids and an overlap of 10 amino acids covered the full sequence of the 32 proteins. PEPperPRINT use a laser printer to synthesize, or print, the peptides directly onto the chip.[84] We contributed with recombinant His-tagged

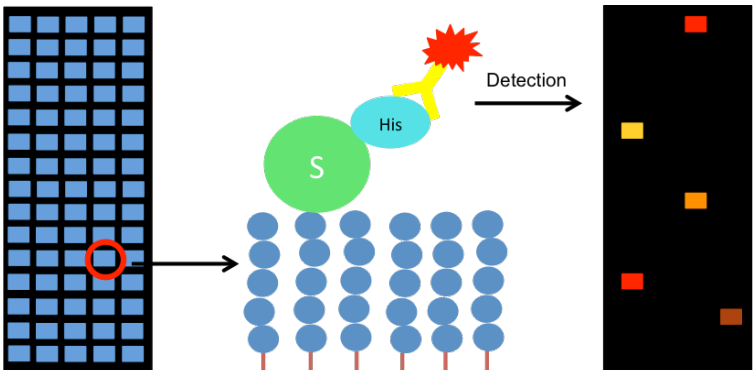


Figure 14, Schematic presentation of a protein binding peptide microarray: The size of the chip is 75.4 x 25 x 1 mm and up to 11,000 peptides can be printed onto it. The intensity of the fluorescence is scanned and reported as result.

survivin, which was added to the chip at a concentration of 1 µg/ml. A fluorescently labeled antibody with specificity for His-tag (Mouse anti-6x-His Epitope Tag DyLight680) is used for detection of bound protein.

An extensive literature search was performed to select the proteins from which the peptides were derived. Both known and assumed interaction partner protein of survivin were included in the microarray. Many positive hits were obtained and the result indicates interaction between survivin and peptides from several of the included proteins. The interactions between survivin and the proteins from the PRC2 are further elaborated in paper I.

There is a risk that false positives are detected with this method. To minimize the risk of false assumptions the peptides are usually printed in duplicates and control peptides are added. Control peptides HA (YPYDVPDYAG) and 6xHis-tag (HHHHHH) were printed in 116 spots respectively. To this microarray we also added 4 peptides from sgol1 and sgol2, which displayed strong interaction with survivin in a previous microarray, to 10 spots respectively as additional controls. It also has to be taken into account also that histidine-rich peptides may interact with the antibody and give false positive result. This method gives a really good first indication of biological interactions, however it reveals nothing about the specificity of the binding and the result need to be followed up by other interaction methods, such as ITC, MST or BLI.

Longer Peptide Design

For further analysis of the interactions between survivin and the proteins of PRC2 a number of longer peptides from the PRC2 related proteins EZH2, EED, SUZ12 and JARID2 were designed, with a size between 20 and 40 amino acids. At first we were focusing on the peptides that had the highest fluorescence intensity in the peptide microarray. We found several areas in the sequences which had high fluorescence in several constituent peptides.

Additional peptides were designed from the areas of lower fluorescence intensity, see figure xx, in this case we were also considering the biological features and the probability of binding to survivin when designing the peptides.

The interaction between these peptides and survivin were investigated with BLI and NMR, and are described in more detail in paper I.

MicroScale Thermophoresis (MST)

MST is a useful method for exploration of biomolecular interactions and dynamics. The technique is based on the physical principle of thermophoresis – the motion of molecules in a temperature gradient. The thermophoretic response of a molecule depends on its charge, size and hydration shell. This technique can be used to quantify molecular interactions, such as protein-protein interactions or the interaction between a protein and a ligand, on a very small scale. The dissociation constant, K_D, can be determined down to 1 picomolar, in short time and with very small amounts of reagents (µl). This makes it a popular method for high throughput screenings (HTS) studying the binding of different compounds to a protein. MST can also be used for monitoring dynamics in proteins, such as enzyme kinetics and protein folding.[85, 86]

Thermophoresis, or thermal diffusion, is the movement of molecules in a temperature gradient. It is based on the fact that molecules move from or to a warmer area, and the difference in concentration is calculated (Equation 3). The movement can be described as a local thermodynamic equilibrium following an exponential Boltzmann steady state distribution.[87] The Soret coefficient, S_T, is used to quantify the molecule.

Equation 12

$$\frac{c}{c_0} = e^{[-S_T(T-T_0)]}$$

c = concentration when molecules are heated

c_0 = initial concentration

T = final temperature

T_0 = initial temperature

S_T is specific for the molecule and it is based on charge, size and hydration shell of the molecule.[88] These parameters will be different for bound molecules than for unbound and hence gives a good indication of interaction between the compounds.

Fluorescence

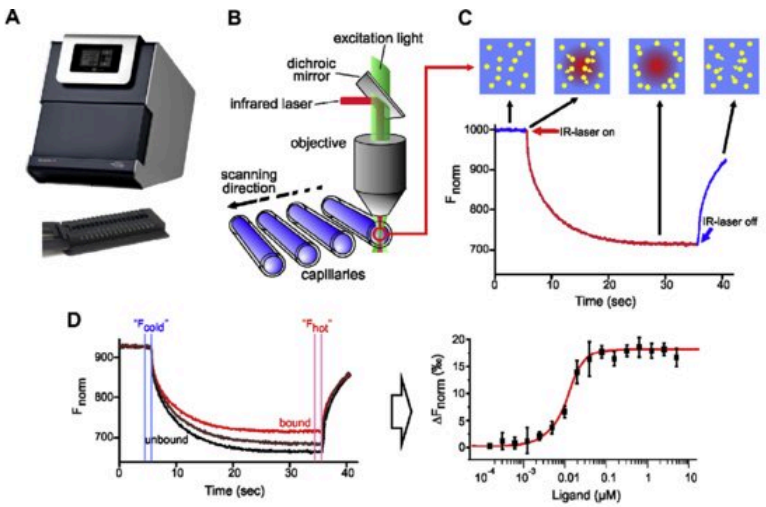


Figure 15, MST setup. (A) Monolith NT 115 and the capillary tray where up to 16 capillaries can be placed. (B) Schematics of MST optics. (C) Signal of an MST experiment. (D) MST binding experiment. F_{cold} and F_{hot} are used to calculate ΔF_{norm}, which is used for the binding curve. Each point in the binding curve represents ΔF_{norm} for a specific concentration of unlabeled ligand.[86]

Monolith NT 115 (NanoTemper Technologies) was used for the experiments in this study. The experimental procedure is relatively straightforward; a dilution series of unlabeled ligand (protein or drug) is prepared, to which a constant amount of fluorescent-labeled protein is added.[79] The samples are loaded into small capillaries; a maximum of 16 samples can be

placed into the apparatus and are measured in the given order. Concentrations and information about the samples is entered into the included software and once the measurement is started everything will be automatically ran.[89]

The measurements are initiated by a capillary scan; the fluorescence of the capillaries is quantified and the capillary positions are saved. A temperature gradient is then generated with an IR-laser, which is focused on the capillaries, one by one.

Different approaches can be used in the evaluation of the result, most common it to use the "Thermophoresis with T-jump" mode. In this mode two events are being monitored; (1) the instant temperature jump that is occurring when the IR-laser is turned on and affecting the fluorophore while the thermophoretic movement has not yet started and (2) The ratio between fluorescence from before the MST power is turned on to just before the power is turned off again is measured. Ideally, there should be none or very small difference between the samples before the MST power is turned on. However, if there is an interaction between the molecules there should be a difference between the samples with different concentration of ligand in the measured fluorescence just before the power is turned off again. The dissociation constant, K_D, is calculated by plotting the normalized fluorescence against the logarithm of the concentrations.[86]

In paper II and III Bayesian machine learning principles have been applied to assess the thermophoresis binding curves from survivin with peptides from the known interaction partner borealin[30, 90] and with peptides from the presumed interaction partners sgol1 and sgol2, which had been identified in a previous protein peptide microarray binding assay.

Protein Interactions Studied by NMR Spectroscopy

With NMR spectroscopy protein-ligand with very low affinities and high (mM-pM) can be quantified. It is one of few methods that can provide information about kinetics and structure simultaneously.[91] For these experiments usually the protein is isotopically labeled and the ligand is titrated into the sample while any changes in the spectrum of the protein are detected. Each peak of the NMR spectrum represents a residue of the protein. Interaction between the protein and ligands can be indicated by a change in the position, intensity or line width of a peak. The position of a peak depends on the chemical environment and is a clear indication of interaction. The intensity of a peak changes with a change of sample

concentration, a weakening of intensity signifies lower concentration of the original molecule, which can be the case if some is bound to the ligand. As any change in the environment of a molecule will affect the position of the peak indicate an interaction. A change in the intensity of the peak depend on the concentration in the sample, this can also indicate an interaction. The line width of the peaks is another indication of interaction, as it depends on the overall correlation time of the molecule and thus indicates change between different states. [74, 91, 92]

Already in 2005 a solution NMR structure of survivin was published by Chaohong and colleagues [33], however a truncated version (1-120) of the protein was used. In this study they assigned the peaks to the protein, investigated the dimer interphase in survivin and explored its interaction with the apoptosis related protein smac/Diablo. They demonstrated an interaction between survivin (L54, L64 and L87) and a small peptide of smac/Diablo (AVPFY) in a [$^{13}C^{1}H$]HSQC spectrum.[33]

Heteronuclear single quantum coherence (HSQC) [$^{13}C^{1}H$]HSQC spectra were recorded of ^{15}N-labeled survivin with a titration peptide. Spectra were recorded for ^{15}N-labeled survivin alone and together with a concentration series of peptide added. If the peptides are interacting with the labeled protein some peaks in the spectra are expected to change with the concentration. The peaks can be shifted or disappearing, or new peaks could appear. With this method one can know with which residues of the labeled protein the unlabeled peptide is interacting, if the peaks are assigned.

The plan was to have a good enough spectra of survivin that could be used for exploration of the interactions and hopefully locate the interacting residues of survivin. With an increasing amount of peptide, I could see changes in some of the peaks (Figure 16). However we were not able to assign these peaks.

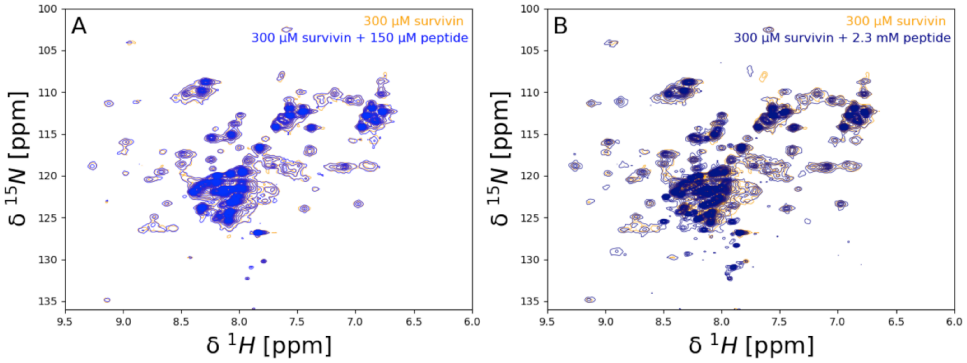


Figure 16. Overlay of 2D [^{15}N,^{1}H] NMR spectra of 300 μM survivin in the absence (orange) and presence (blue) of 0.5 μM peptide VELVNALGQYNDDDDDDDGDDPEEREEKQKDLEDHRDDKE which corresponds to residues 172-211 from EZH2. (B) Overlay of 2D [^{15}N,^{1}H] NMR spectra of 300 μM survivin in the absence (orange) and presence (dark blue) of 2.3 mM of the same peptide.

Evaluation of Experimental Data

Evaluation of the obtained experimental data and presenting is always a challenging task. A good starting point is to use a statistical approach. There are several commercial software and available, although they do not always provide the analysis that we want. We have been using Python with the Anaconda software to develop our own scripts in order to simplify and improve the evaluation of our results.

Statistical Inference

Descriptive statistics, such as mean, median and mode, can be used to describe a dataset. It summarizes and describes the characteristics of the sampled data, however it says nothing about the rest of the population.[93] Statistical inference includes drawing conclusions about a population from a sample by introducing e.g. probability distributions or hypothesis testing. By assuming that the sample is representative for the entire population one can predict how variables in a dataset would relate to one other if the entire population is considered. [93]

Probability is a measure of uncertainty. Bayes formula is is used to predict conditional probability conditional probability;

$$P(A|B) = \frac{P(B|A)B(A)}{P(B)}$$

Equation 13

Where;
P(A) = the probability that event A will occur
P(B) = the probability that event B will occur
P(A|B) = the conditional probability that event A will occur given that event B has occurred
P(B|A) = the conditional probability that event B will occur given that event A has occurred

There are two different approaches to inference statistics; frequentist (or classical) and Bayesian inference and they have different definitions of the p-value (probability value). A common frequentist approach is to define a null hypothesis (e. g. comparing the mean of two groups) and calculate a p-value. The p-value is a measure of how well the null hypothesis represents a given test and data. In a Bayesian approach the p-value defines the probability of the hypothesis, hence it's a value of how well the model fits the data. [94, 95]

In Bayesian statistics a prior probability is added to the analysis of the data in the form of a prior probability distribution. In the end the outcome will be a posterior probability distribution, which comprise all possible beliefs.[96] The prior distribution can be informative or uninformative. An informative prior distribution is derived from a known parameter or value from previous experiments (empirical Bayes). Often, there is no such information and an uninformative prior distribution is used. The choice of prior distribution is important and can be difficult.[97, 98]

The posterior distribution is proportional to the likelihood times the prior distribution, and this relation can be used to calculate the posterior distribution for the given data (Equation 15).

$$Posterior \propto Likelihood \times Prior$$

Equation 14

Bayesian approaches often include advanced calculations, algorithms and often considered as part of machine learning, in order to find the best model.

Machine Learning

It all started in 1950 with Alan Turing asking the question "Can machines think?".[99] He was far ahead of his time when this question was asked and the progress has come a long way

since then. Artificial Intelligence (AI) enables machines to simulate human behavior. Machine learning (ML) is a subset of AI, which allows a machine to automatically learn from past data without being explicitly programmed for it. Hence, AI is the bigger concept and ML is a subfield or branch of AI.[100]

The most common ML algorithms can be roughly divided into different groups; supervised and unsupervised learning. Supervised learning includes algorithms that develop predictive models on the basis on both input and output data while unsupervised learning algorithms group and interpret data based only on the input data.[101]

Classification is a group of supervised or unsupervised ML algorithms that are used to decide which category an object belongs to based on its features. There are a number of classification algorithms that can be used depending on what type of classification, binary, multi-class etc. Supervised **regression analysis** was used for analysis of MST data in paper II and paper III. This type of analysis was used in **paper II** and **paper III**, where we use a Bayesian approach to improve the interpretation of MST data.

Unsupervised classification is dominated by clustering algorithms.[102] **Clustering** is a group of unsupervised classification algorithms, with the purpose of grouping data that display a similar pattern. In **paper I** and **paper IV** we have used hierarchal clustering to identify groups with similar properties.

Concluding Remarks

The work in my thesis has been expanding over a wide range, including different projects with different goals and using different methods. **In paper I**, **II** and **III** a biological approach was used to investigate interactions and functions of survivin, which is an interesting target for therapeutics. Whereas in **paper V** and **paper IV** the focus was more directed to method development the common theme is protein dynamics. We have seen that the type of atom affect the dynamics and hence the function of a protein.

When I started this project there was already a protocol for production of survivin. I used the existing protocol and optimized it for production of isotope labeled survivin for NMR experiments. An interesting connection between survivin and PRC2 complex was found and further investigated. Although there is much more to investigate when it comes to the interaction between survivin and the proteins from the PRC2 complex. This will be an interesting challenge for someone else to take on.

Biophysical method development is a constantly ongoing procedure, although some milestones were achieved with my work. In this project the method was the focus and we used well-established, commercially available protein systems. We managed to perform X-ray crystallography diffraction experiments at a novel beamline. We were also able to detect some effect of the diffraction from trypsin crystals while probing them with THz radiation, paving the way for conducting time-resolved experiments at the FemtoMAX beamline in MAXIV.

Acknowledgments

I am so grateful for being given the opportunity to do my PhD in the Lundberg laboratory, it has really been the best time of my life! All in all I have been here for a little more than six years. There are so many people I would like to thank for this time and probably many I will forget to mention. **Majo!** I want to thank you for even being here. You took me in to do my bachelor and master thesis with you which eventually led to also doing my PhD with the group. It has been an amazing opportunity to do this and at the same time find such a good

friend! Thank you **Gergely** for seeing the potential in me. Thank you for supporting me through all of this! **Viktor and Laras**, thank you for being good colleagues and for the fun we have had in beamtimes.

Weixiao, one of the few people who has been in the lab all this time. Thank you for all your support and for being one of my best friends.

Then there are some of the oldies I would like to thank for all the fun we've had. **Elin C**, I had so much fun teaching and hanging out with you. **Rebecka**, I missed your fun comments and stories so much when you left. **Petra E**, **Petra B**, **Elin D**, **Rhawnie**, **Cecilia S**, **Rob B**, **Rob D**, **Emil** and **Oscar** thanks for all interesting talks during lunch, fika and beer club "back in the days". **Andreas**, you are still around but I include you in the oldies category. We had a lot of fun in the early beer clubs. **Cecilia W** thank you for making the lab neat when no one else cared. **Rajiv**, I'm still not pregnant. **Florian**, it has been nice to have you back in the lab now and when and to watch you prepare your tomato sandwiches. **Leo**, we have had some amazing trips together!

Ann, you are an amazing roommate and fika companion. **Amke**, I miss you at morning coffee. **Andrea,** you always have something nice to say. **Jessica**, thanks for always smiling. **Tinna**, thanks for good collaboration and fun beamtimes, and for being a good lab manager!

Darius and **Emelie** and **Damasus**, we started around the same time and did some early courses together. **Ylber** (rainbow), **Jonatan**, **Jens** and **Daniel**, I've had some nice interesting conversations in beer club with you. **Per** and **Greger**, for you I just say Lorne! **Giorgia**, too bad you didn't come with us.

Analia, we go way back to KEM350 in 2015. It's nice to have you at Lunderg.

Swagatha, thanks for teaching me to cook nice Indian food! **Ashish**, **Yosh**, **Irena**, **Laura** and **Lisa**, thanks for contributing to a better work environment.

During the years I have been in the lab many people have started which I unfortunately have not had the opportunity to get to know that well. **Lucija**, **Katharina**, **Doris**, **Leona**, **Owens**, **Adams**, **Arpitha** and **Dylan**. I'm probably forgetting someone, sorry for that!

Then I would like to thank all the other PI:s of the lab. **Richard**, my examiner who almost became my supervisor and **Gisela** who believed in me and pushed me to take on this challenge. **Johanna,** we had many interesting discussions. Thank you **Björn** for helping me with NMR in the beginning and for caring about the lab environment. You should be a role model. **Kristina**, with your teaching you inspired me to deepen my knowledge in biochemistry. **Sebastian**, enjoy Uppsala. **Michal** and **Julia,** I wish you all the best.

I would also like to thank all the collaborators I've met during the years. **Ulrika**, thanks for all the help with NMR measurements. **Martin**, **Clement** and **Andrej** at Petra III and **Manfred**, **Sia** and **Til** thanks for our beamtimes at P12. Also thanks to **Jörgen**, **David**, **Thai**, **Andrius** and **Carl** at the FemtoMAX beamline in MAXIV. Thank you **Juan** for joining several beamtimes with our group. **Ann** and **Oscar** too bad I didn't have the chance to come to BioSAXS as planned during my PhD. **Maria**, **Malin** and **Venkat** thanks for all the "fun" we had with survivin. **Sergio** and **Scott**, thanks for good collaboration in beamtimes at ESRF, and **Pablo**, thanks for good collaboration at BioMAX beamtimes. And thanks to **Göran** for being my co-supervisor.

Last of all I would like to thank my family and people around me. **Magnus,** you are the most amazing person I have ever met and you have been an amazing support to me. Thanks for always being there, I don't think I would have been finishing without you! My kids, **Hanna** and **Victor**, thanks for putting up with me all this time. My parents, **Eva** and **Henning**, thanks for supporting me and helping me so much with everything. My sisters, **Ida** and **Karin**, I hope you can read this without falling asleep. My study mates **Cicci** and **Sofie**, thanks for all the fun we had and will have. **Linn**, you always understand me. **Mari** and **Hanna** thanks for listening to all my problems during the years. **Zorica** you and me, sunset in Italy! **Jenny**, you have been my friend for so many years and still going strong.

References

1. Jaeger, G., *Are virtual particles less real?* Entropy (Basel, Switzerland), 2019. **21**(2): p. 141.
2. Smith, B.J. and M.G. Raymer, *Photon wave functions, wave-packet quantization of light, and coherence theory*. New journal of physics, 2007. **9**(11): p. 414-414.

3. Chavas, L.M.G., *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules. By Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins and Roland P. May. IUCr Texts on Crystallography, No. 19. Oxford University Press, 2013. Pp. 358. Price (hardback) £49.99. ISBN: 978-0-19-963953-3.* Acta Crystallographica Section D Biological Crystallography, 2014. **70**(4): p. 1175-1176.
4. Maxwell, J.C., *A treatise on electricity and magnetism Vol. 1*. 1873, Oxford: Oxford.
5. Dar, A. and S. Varma, *Coherent Elastic Scattering and Incoherent Inelastic Scattering in Proton-Nucleus Collisions at High Energy*. Physical Review Letters, 1966. **16**(22): p. 1003-1006.
6. Perticaroli, S., et al., *Broadband depolarized light scattering study of diluted protein aqueous solutions.* J Phys Chem B, 2010. **114**(24): p. 8262-9.
7. Henzler-Wildman, K.A., et al., *A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.* Nature, 2007. **450**(7171): p. 913-6.
8. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-72.
9. Go, N., T. Noguti, and T. Nishikawa, *Dynamics of a small globular protein in terms of low-frequency vibrational modes.* Proc Natl Acad Sci U S A, 1983. **80**(12): p. 3696-700.
10. Chapman, B. and D. Loiselle, *Perspective - life and death of a photon: an intuitive non-equilibrium thermodynamic distinction between photochemistry and thermochemistry.* Photochemical & photobiological sciences, 2020. **19**(12): p. 1623-1629.
11. Katchalsky, A., *Thermodynamics of living systems : a special issue dedicated to Aharon Katchalsky*. 1975, New York: New York.
12. Maslov, S., *Mass-Action Equilibrium, Noise, and Non-Specific Interactions in Protein Interaction Networks.* Biophysical journal, 2011. **100**(3): p. 342a-342a.
13. Lienhard, G.E., *Enzymatic Catalysis and Transition-State Theory.* Science (American Association for the Advancement of Science), 1973. **180**(4082): p. 149-154.
14. Berg, J.M., et al., *Biochemistry*. 8. ed. ed. 2015, New York, NY: New York, NY : Freeman.
15. Campbell, M.K., *Biochemistry*. 8th edition. ed, ed. S.O. Farrell and Dawsonera. 2015: Australia : Cengage Learning.
16. Srinivasan, B., *A guide to the Michaelis-Menten equation: steady state and beyond.* The FEBS journal, 2021.
17. Goutelle, S., et al., *The Hill equation: a review of its capabilities in pharmacological modelling.* Fundamental & clinical pharmacology, 2008. **22**(6): p. 633-648.
18. Basha, E., H. O'Neill, and E. Vierling, *Small heat shock proteins and alpha-crystallins: dynamic proteins with flexible functions.* Trends Biochem Sci, 2012. **37**(3): p. 106-17.
19. Markelz, A., et al., *THz time domain spectroscopy of biomolecular conformational modes.* Phys Med Biol, 2002. **47**(21): p. 3797-805.
20. Rodilla, H., et al., *Applying THz Technology in Life Science*. Swedish Microwave Days. Gigahertz Symposium, March 11-12, 2014. Gothenburg, Sweden, 2014.
21. Varvdekar, B., A. Prabhakant, and M. Krishnan, *Response of Terahertz Protein Vibrations to Ligand Binding: Calmodulin–Peptide Complexes as a Case Study.* Journal of chemical information and modeling, 2022. **62**(7): p. 1669-1679.
22. Cheatum, C.M., *Low-Frequency Protein Motions Coupled to Catalytic Sites.* Annual review of physical chemistry, 2020. **71**(1): p. 267-288.
23. Chantalat, L., et al., *Crystal structure of human survivin reveals a bow tie-shaped dimer with two unusual alpha-helical extensions.* Mol Cell, 2000. **6**(1): p. 183-9.
24. Arora, V., et al., *Degradation of survivin by the X-linked inhibitor of apoptosis (XIAP)-XAF1 complex.* J Biol Chem, 2007. **282**(36): p. 26202-9.

25. Bourhis, E., S.G. Hymowitz, and A.G. Cochran, *The mitotic regulator Survivin binds as a monomer to its functional interactor Borealin.* J Biol Chem, 2007. **282**(48): p. 35018-23.
26. Ebrahimiyan, H., et al., *Survivin and autoimmunity; the ins and outs.* Immunol Lett, 2018. **193**: p. 14-24.
27. Altieri, D.C., *Survivin and IAP proteins in cell-death mechanisms.* Biochem J, 2010. **430**(2): p. 199-205.
28. Carmena, M., et al., *The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis.* Nat Rev Mol Cell Biol, 2012. **13**(12): p. 789-803.
29. Verdecia, M.A., et al., *Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement.* Nat Struct Biol, 2000. **7**(7): p. 602-8.
30. Jeyaprakash, A.A., et al., *Structure of a Survivin-Borealin-INCENP core complex reveals how chromosomal passengers travel together.* Cell, 2007. **131**(2): p. 271-85.
31. Jeyaprakash, A.A., et al., *Structural basis for the recognition of phosphorylated histone h3 by the survivin subunit of the chromosomal passenger complex.* Structure, 2011. **19**(11): p. 1625-34.
32. Song, Z., X. Yao, and M. Wu, *Direct interaction between survivin and Smac/DIABLO is essential for the anti-apoptotic activity of survivin during taxol-induced apoptosis.* J Biol Chem, 2003. **278**(25): p. 23130-40.
33. Sun, C., et al., *Solution structure of human survivin and its binding interface with Smac/Diablo.* Biochemistry, 2005. **44**(1): p. 11-7.
34. Vallet, C., et al., *Functional Disruption of the Cancer-Relevant Interaction between Survivin and Histone H3 with a Guanidiniocarbonyl Pyrrole Ligand.* Angew Chem Int Ed Engl, 2020. **59**(14): p. 5567-5571.
35. Watanabe, Y., *Shugoshin: guardian spirit at the centromere.* Curr Opin Cell Biol, 2005. **17**(6): p. 590-5.
36. Vanoosthuyse, V., S. Prykhozhij, and K.G. Hardwick, *Shugoshin 2 regulates localization of the chromosomal passenger proteins in fission yeast mitosis.* Mol Biol Cell, 2007. **18**(5): p. 1657-69.
37. Tsukahara, T., Y. Tanno, and Y. Watanabe, *Phosphorylation of the CPC by Cdk1 promotes chromosome bi-orientation.* Nature, 2010. **467**(7316): p. 719-23.
38. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-9.
39. Wassef, M. and R. Margueron, *The Multiple Facets of PRC2 Alterations in Cancers.* J Mol Biol, 2017. **429**(13): p. 1978-1993.
40. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.
41. van Mierlo, G., et al., *The Complexity of PRC2 Subcomplexes.* Trends Cell Biol, 2019. **29**(8): p. 660-671.
42. Qi, W., et al., *An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED.* Nat Chem Biol, 2017. **13**(4): p. 381-388.
43. Dodson, G. and A. Wlodawer, *Catalytic triads and their relatives.* Trends in Biochemical Sciences, 1998. **23**(9): p. 347-352.
44. Dodson, G. and A. Wlodawer, *Catalytic triads and their relatives.* Trends in biochemical sciences (Amsterdam. Regular ed.), 1998. **23**(9): p. 347-352.
45. Kuhn, P., et al., *The 0.78 A structure of a serine protease: Bacillus lentus subtilisin.* Biochemistry, 1998. **37**(39): p. 13446-52.
46. Polgar, L., *The catalytic triad of serine peptidases.* Cell Mol Life Sci, 2005. **62**(19-20): p. 2161-72.
47. Liebschner, D., et al., *On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin.* Acta Crystallographica Section D-Biological Crystallography, 2013. **69**: p. 1447-1462.
48. Structural Genomics, C., et al., *Protein production and purification.* Nat Methods, 2008. **5**(2): p. 135-46.

49. Cai, M., et al., *A simple protocol for expression of isotope-labeled proteins in Escherichia coli grown in shaker flasks at high cell density.* J Biomol NMR, 2019. **73**(12): p. 743-748.
50. Harrison, S.T.L., *2.47 - Cell Disruption*, in *Comprehensive Biotechnology (Third Edition)*, M. Moo-Young, Editor. 2011, Pergamon: Oxford. p. 692-712.
51. Janson, J.-C., *Protein purification : principles, high resolution methods, and applications.* John Wiley & Sons, 2011.
52. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
53. Munro, I.H., *Fifty years of synchrotron radiation research in the UK.* Philos Trans A Math Phys Eng Sci, 2019. **377**(2147): p. 20180230.
54. Rhodes, G., *Crystallography Made Crystal Clear A Guide for Users of Macromolecular Models*. 3. ed. ed. 2006: Oxford : Academic Press.
55. Guss, J.M., *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology, by Bernard Rupp.* Crystallography Reviews, 2011. **17**(1): p. 65-67.
56. Schmidt, A., et al., *Trypsin revisited: crystallography AT (SUB) atomic resolution and quantum chemistry revealing details of catalysis.* J Biol Chem, 2003. **278**(44): p. 43357-62.
57. Weeratunga, S.K., et al., *Structural studies of bacterioferritin B from Pseudomonas aeruginosa suggest a gating mechanism for iron uptake via the ferroxidase center.* Biochemistry, 2010. **49**(6): p. 1160-75.
58. Rupp, B., *Biomolecular crystallography : principles, practice, and application to structural biology*. 2010: New York : Garland Science.
59. Neutze, R., et al., *Potential for biomolecular imaging with femtosecond X-ray pulses.* Nature, 2000. **406**(6797): p. 752-7.
60. Bailey, S., *The CCP4 Suite - Programs for Protein Crystallography.* Acta Crystallographica Section D-Biological Crystallography, 1994. **50**: p. 760-763.
61. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.
62. Enquist, H., et al., *FemtoMAX - an X-ray beamline for structural dynamics at the short-pulse facility of MAX IV.* J Synchrotron Radiat, 2018. **25**(Pt 2): p. 570-579.
63. Neutze, R., *Opportunities and challenges for time-resolved studies of protein structural dynamics at X-ray free-electron lasers.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1647): p. 20130318.
64. McCoy, A.J., *Solving structures of protein complexes by molecular replacement with Phaser.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 32-41.
65. Korasick, D.A. and J.J. Tanner, *Determination of protein oligomeric structure from small-angle X-ray scattering.* Protein Sci, 2018. **27**(4): p. 814-824.
66. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40**(3): p. 191-285.
67. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Lett, 2015. **589**(19 Pt A): p. 2570-7.
68. Velours, C., et al., *Macromolecular interactions in vitro, comparing classical and novel approaches.* Eur Biophys J, 2021. **50**(3-4): p. 313-330.
69. Blanchet, C.E., et al., *Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY).* Journal of applied crystallography, 2015. **48**(2): p. 431-443.
70. Jeffries, C.M., et al., *Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments.* Nat Protoc, 2016. **11**(11): p. 2122-2153.
71. Manalastas-Cantos, K., et al., *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis.* J Appl Crystallogr, 2021. **54**(Pt 1): p. 343-355.

72. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis.* Journal of Applied Crystallography, 2012. **45**: p. 342-350.
73. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Quarterly reviews of biophysics, 2007. **40**(3): p. 191-285.
74. Marion, D., *An Introduction to Biological NMR Spectroscopy.* Molecular & cellular proteomics, 2013. **12**(11): p. 3006-3025.
75. Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide Protein Data Bank.* Nature Structural & Molecular Biology, 2003. **10**(12): p. 980-980.
76. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Prog Nucl Magn Reson Spectrosc, 2013. **73**: p. 1-16.
77. Furey, T.S., *ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions.* Nat Rev Genet, 2012. **13**(12): p. 840-52.
78. Abdiche, Y., et al., *Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet.* Anal Biochem, 2008. **377**(2): p. 209-17.
79. Seidel, S.A., et al., *Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding.* Angew Chem Int Ed Engl, 2012. **51**(42): p. 10656-9.
80. Mao, Y., et al., *A novel method for the study of molecular interaction by using microscale thermophoresis.* Talanta, 2015. **132**: p. 894-901.
81. O'Connell, M.R., R. Gamsjaeger, and J.P. Mackay, *The structural analysis of protein-protein interactions by NMR spectroscopy.* Proteomics, 2009. **9**(23): p. 5224-32.
82. Chang, T.-W., *Binding of cells to matrixes of distinct antibodies coated on solid surface.* Journal of Immunoloogical Methods, 1983. **65**: p. 217-223.
83. Sheng-Ce Tao, C.-S.C.a.H.Z., *Applications of Protein Microarray Technology.* Combinatorial Chemistry & High Throughput Screening, 2007. **10**.
84. Stadler, V., et al., *Combinatorial synthesis of peptide arrays with a laser printer.* Angew Chem Int Ed Engl, 2008. **47**(37): p. 7132-5.
85. Lopez-Mendez, B., et al., *Microscale Thermophoresis and additional effects measured in NanoTemper Monolith instruments.* Eur Biophys J, 2021. **50**(3-4): p. 653-660.
86. Jerabek-Willemsen, M., et al., *MicroScale Thermophoresis: Interaction analysis and beyond.* Journal of Molecular Structure, 2014. **1077**: p. 101-113.
87. Duhr, S. and D. Braun, *Thermophoretic depletion follows Boltzmann distribution.* Phys Rev Lett, 2006. **96**(16): p. 168301.
88. Wienken, C.J., et al., *Protein-binding assays in biological liquids using microscale thermophoresis.* Nat Commun, 2010. **1**: p. 100.
89. Sedivy, A., *Standard operating procedure for NanoTemper Monolith measurements.* Eur Biophys J, 2021. **50**(3-4): p. 381-387.
90. Gassmann, R., et al., *Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle.* J Cell Biol, 2004. **166**(2): p. 179-91.
91. Teilum, K., et al., *(S)Pinning down protein interactions by NMR.* Protein Sci, 2017. **26**(3): p. 436-451.
92. Mittermaier, A. and L.E. Kay, *New tools provide new insights in NMR studies of protein dynamics.* Science, 2006. **312**(5771): p. 224-8.
93. Dennis, J., *Probability and Statistics for Engineering and the Sciences*. 2009, Alexandria: American Statistical Association: Alexandria. p. 101-101.
94. van de Schoot, R., et al., *A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research.* Child development, 2014. **85**(3): p. 842-860.
95. Held, L. and M. Ott, *On p-Values and Bayes Factors.* Annual review of statistics and its application, 2018. **5**(1): p. 393-419.
96. Detilleux, J., *Study design synopsis: Battle in the stable: Bayesianism versus Frequentism.* Equine veterinary journal, 2021. **53**(2): p. 199-204.

97. Berger, J.O., J.M. Bernardo, and D. Sun, *The Formal Definition of Reference Priors.* The Annals of statistics, 2009. **37**(2): p. 905-938.
98. Berger, J.O., D. Sun, and C. Song, *Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors.* The Annals of statistics, 2020. **48**(4): p. 2381-2403.
99. Turing, A., *Computing Machinery and Intelligence.* MInd, 1950. **LIX**(236): p. 433-460.
100. Mehlig, B., *Machine learning with neural networks : an introduction for scientists and engineers*. 2021: Cambridge, United Kingdom
New York, NY : Cambridge University Press.
101. Garreta, R., *Learning scikit-learn : machine learning in Python*, ed. G. Moncecchi. 2013, Birmingham: Birmingham : Packt Publishing.
102. Hulot, A., et al., *Fast tree aggregation for consensus hierarchical clustering.* BMC bioinformatics, 2020. **21**(1): p. 120-120.