

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 2292*

Structural studies of drug targets and a drug metabolizing enzyme

DANIELA CEDERFELT



ACTA UNIVERSITATIS
UPSALIENSIS
2023

ISSN 1651-6214
ISBN 978-91-513-1865-3
urn:nbn:se:uu:diva-508764



UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in B41, BMC, Husargatan 3, Uppsala, Tuesday, 26 September 2023 at 09:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Inari Kursula (University of Bergen, Department of Biomedicine).

Abstract

Cederfelt, D. 2023. Structural studies of drug targets and a drug metabolizing enzyme. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2292. 65 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1865-3.

The work presented in this thesis describes how structural information about a protein can be acquired, and how it can be used to answer scientific questions about proteins' function, their dynamic behaviour and their interactions with other proteins or ligands.

The catalytic function of the pyrimidine-degrading, drug metabolizing enzyme β -ureidopropionase (β UP) is dependent on the shift between oligomeric states. Substitution of amino acids H173 and H307 in the dimer-dimer interface and E207Q in the active site revealed that these are crucial for β UP activation. Inhibition studies of substrate- and product analogues allowed for a hypothesis that the ability to interact with F205 might distinguish activators from inhibitors. The first structure of the activated higher oligomer state of human β UP was determined using cryogenic electron microscopy, and confirmed that the closed entrance loop conformations and dimer-dimer interfaces are conserved between Hs β UP and Dm β UP.

Interactions between the epigenetic drug target SET and MYND domain containing protein 3 (SMYD3) and possible inhibitors were investigated. A crystal structure confirmed the covalent bond of a rationally designed, targeted inhibitor to C186 in the active site of SMYD3. A new allosteric binding site was discovered using a biosensor screen with a blocked active site. Crystal structures revealed the location of the new binding site, and the binding mode of the (S)- and (R) enantiomers of the allosteric inhibitor. Lastly, a fragment based drug discovery approach was taken, co-crystallizing and soaking SMYD3 with hits from a fragment screen. This resulted in four crystal structures with weak electron density of fragments at several locations in the enzyme.

The dynamic acetylcholine binding protein (AChBP) is a homologue of a Cys-loop type ligand gated ion channel. Hits from various biosensor screens, of which some indicated conformational changes, were co-crystallized with AChBP. Seven crystal structures of AChBP in complex with hit compounds from the biophysical screens were determined. Small conformational changes in the Cys-loop were detected in several of the crystal structures, coinciding with the results from the biosensor screens.

In these studies, we explore new strategies for the investigation of the function and regulation of proteins relevant in drug discovery and optimization.

Keywords: Biochemistry, Biophysics, Protein structure, X-ray crystallography, Cryogenic electron microscopy, Enzymology, Drug discovery, Pyrimidine degradation

Daniela Cederfelt, Department of Chemistry - BMC, Biochemistry, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

© Daniela Cederfelt 2023

ISSN 1651-6214

ISBN 978-91-513-1865-3

URN urn:nbn:se:uu:diva-508764 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-508764>)

*Science is the language in which we converse with nature,
and its truths are revealed through the language of evidence*

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I. **Cederfelt, D.**, Badjugar, D., Musse, A., Maurer, D., Lohkamp, B. and Dobritzsch, D. The allosteric regulation of the anticancer drug-metabolizing β -ureidopropionase depends on fine-tuned active-site loop and subunit interface stability (*Manuscript*)
- II. Parenti, M.D., Naldi, M., Manoni, E., Fabini, E., **Cederfelt, D.**, Talibov, V.O., Gressani, V., Guven, U., Grossi, V., Fasano, C., Sanese, P., De Marco, K., Shtil, A.A., Kurkin, A.V., Altieri, A., Danielson, U.H., Caretti, G., Simone, C., Varchi, G., Bartolini, M. and Del Rio, A. (2022) Discovery of the 4-aminopiperidine-based compound EM127 for the site-specific covalent inhibition of SMYD3. *Eur J Med Chem.* **243**, 114683.
- III. Talibov, V.O., Fabini, E., FitzGerald, E.A., Tedesco, D., **Cederfelt, D.**, Talu, M.J., Rachman, M.M., Mihalic, F., Manoni, E., Naldi, M., Sanese, P., Forte, G., Lepore Signorile, M., Barril, X., Simone, C., Bartolini, M., Dobritzsch, D., Del Rio, A. and Danielson, U.H. (2021) Discovery of an Allosteric Ligand Binding Site in SMYD3 Lysine Methyltransferase. *Chembiochem.* **22(9)**, 1597-1608.
- IV. FitzGerald, E.A., **Cederfelt, D.**, Lund, B.A., Myers, N., Zhang, H., Dobritzsch, D. and Danielson, U.H. Introduction to kinetic screening for the identification of weak fragment hits – mapping ligand binding sites in SMYD3 (*Manuscript*)
- V. FitzGerald, E.A., Butko, M.T., Boronat, P., **Cederfelt, D.**, Abramson, M., Ludviksdottir, H., van Muijlwijk-Koezen, J.E., de Esch, I.J.P., Dobritzsch, D., Young, T. and Danielson, U.H. (2021) Discovery of fragments inducing conformational effects in dynamic proteins using a second-harmonic generation biosensor. *RSC Advances.* **11(13)**, 7527-7537.

- VI. FitzGerald, E.A., **Cederfelt, D.**, Boronat, P., Lund, B.A., Dobritzsch, D., Hennig, S., de Esch, I.J.P. and Danielson, U.H. Elucidating the regulation of ligand gated ion channels via biophysical studies of ligand-induced conformational dynamics of acetylcholine binding proteins (*Manuscript*)

Reprints were made with permission from the respective publishers.

Contribution report

- Paper I **The allosteric regulation of the anticancer drug-metabolizing β -ureidopropionase depends on fine-tuned active-site loop and subunit interface stability**
Protein expression and purification, site directed mutagenesis, functional studies, X-ray crystallography, cryoEM grid preparation and data collection, drafting and revision of manuscript
- Paper II **Discovery of the 4-aminopiperidine-based compound EM127 for the site-specific covalent inhibition of SMYD3**
X-ray crystallography experiments and analysis
- Paper III **Discovery of an Allosteric Ligand Binding Site in SMYD3 Lysine Methyltransferase**
X-ray crystallography experiments and analysis
- Paper IV **Introduction to kinetic screening for the identification of weak fragment hits – mapping ligand binding sites in SMYD3**
Expression, purification, X-ray crystallography experiments and analysis, contributed to manuscript writing
- Paper V **Discovery of fragments inducing conformational effects in dynamic proteins using a second-harmonic generation biosensor**
X-ray crystallography experiments and analysis
- Paper VI **Elucidating the regulation of ligand gated ion channels via biophysical studies of ligand-induced conformational dynamics of acetylcholine binding proteins**
X-ray crystallography experiments and analysis, contributed to manuscript writing

The following publication is not included in the thesis:

- I. Cornelius, E., Bartl, M., Persson, L., Ruisheng, X., **Cederfelt, D.**, Rad, F., Norberg, T., Engel, S., Marklund, E., Dobritsch, D. and Widersten, M. (2023) Engineered Aldolases Catalyzing Stereoselective Aldol Reactions Between Aryl-Substituted Ketones and Aldehydes. *Catal. Sci. Technol.*, *Epub ahead of print*.

Contents

1. Introduction.....	13
2. Theoretical background	15
2.1. Enzymes	15
2.1.1. Enzyme kinetics.....	17
2.1.2. Pyrimidine degradation.....	20
2.1.3. Epigenetics and the enzymes' role in disease	22
2.2. Protein structure determination	23
2.2.1. X-ray crystallography	24
2.2.2. Cryogenic electron microscopy	28
2.3. Drug discovery	31
2.3.1. Ligand-gated ion channels as drug targets.....	33
3. Present investigations	35
3.1. Characterization of the allosteric regulation of a drug metabolizing enzyme (Paper I).....	36
3.1.1. Site-directed mutagenesis revealed function of crucial residues	37
3.1.2. Substrate and product analogues have effect on oligomer formation and catalytic activity	41
3.1.3. CryoEM structure provide first insights into activated oligomer complex	42
3.1.4. Conclusion.....	43
3.2. X-ray crystallography in drug discovery – An epigenetic drug target (Paper II, III and IV)	44
3.2.1. Rational design of a covalent inhibitor.....	45
3.2.2. Discovery of a new allosteric site.....	47
3.2.3. A fragment based drug discovery approach	48
3.2.4. Conclusion.....	50

3.3. Conformational changes and drug discovery – AChBP as a model of a dynamic drug target (Paper V and VI)	51
3.3.1. Biophysical screening for detection of interaction and conformational changes	52
3.3.2. Structural studies of ligand binding and conformational changes	53
3.3.3. Conclusions	55
4. Summary and future perspectives.....	56
5. Populärvetenskaplig sammanfattning	58
6. Acknowledgements.....	60
7. References.....	61

Abbreviations

5FU	5-Fluorouracil
AChBP	Acetylcholine binding protein
β -Ala	β -alanine
β UP	β -ureidopropionase
cryoEM	Cryogenic electron microscopy
CTF	Contrast transfer function
EM	Electron microscopy
FBDD	Fragment-based drug discovery
HTS	High-throughput screening
LGIC	Ligand gated ion channels
nAChR	Nicotinic acetylcholine receptor
NADH	1,4-Dihydronicotinamide adenine dinucleotide
NC β A	N-carbamoyl- β -alanine
NMR	Nuclear magnetic resonance
PEG	Polyethylene glycol
PDB	Protein data bank
SAM	S-Adenosyl methionine
SAW	Surface acoustic wave
SBDD	Structure based drug discovery
SEC	Size exclusion chromatography
SHG	Second harmonic generation
SMYD3	SET and MYND domain containing protein 3
SPR	Surface plasmon resonance
TCI	Targeted covalent inhibitor
TEM	Transmission electron microscope
XRC	X-ray crystallography

1. Introduction

The function and interactions of proteins control the biochemical processes that our lives depend on.¹ These vital functions and interactions are in turn profoundly dependent on the protein structures. In other words, proteins' structures can reveal their function, how they contribute to our health and disease, and provide essential information for the discovery of new therapeutics. Because an average protein is only up to ~10 nm in length², it is too small to study without advanced techniques. In this thesis, I explore different approaches to obtain protein structural information and use it to answer scientific questions about function and interactions.

The most common technique for the determination of a protein structure is X-ray crystallography (XRC), which has been used for structure determinations of molecules for over a century. The first small molecule structure was determined using XRC in 1913.^{3,4} Since the elucidation of the structure of myoglobin in 1958⁵, the number of protein structures determined by XRC has steadily increased.³ Due to improvements in the XRC technology, including the introduction of high brightness 3rd generation synchrotrons and pixel detectors^{6,7} and the emergence of cryogenic electron microscopy (cryoEM) and nuclear magnetic resonance (NMR), today approximately 204 000 protein structures are available in the protein data bank (PDB)⁸. Of these, 175 000 are from XRC, 13 000 are from NMR and 16 000 from electron microscopy (EM) (PDB 2023). Thus, XRC is still the predominantly used method for protein structure determination. However, recent improvements in cryoEM technology allows for structure determination of larger protein complexes, heterogeneous protein samples⁹ and proteins that cannot be crystallized.

In this thesis, the allosteric regulation of a drug metabolising enzyme in the pyrimidine degradation pathway was investigated by the combination of biochemical methods and structure determination. The complexity of the sample, as a consequence of the dynamic alternation between oligomeric states, made cryoEM the most advantageous method for structure determination.

The interactions between drug targets and possible inhibitors discovered through rational design and various compound library screens were studied through structure determination of the drug targets in complex with the potential ligands. Due to the high resolution required to structurally determine the

binding of the relatively small molecules, XRC was the most appropriate method. Two types of drug targets were explored: an epigenetic enzyme and a homologue of a ligand gated ion channel (LGIC). Enzymes are the largest class of macromolecules and they are responsible for catalysing every reaction in the body. Thus, pathologically altered activity of enzymes is a common cause of disease.^{10, 11} LGICs are receptors that control the passage of molecules and ions over membranes. Upon binding of a ligand, a large conformational change opens up the ion channel, allowing passage of an ion.^{12, 13} Disruptions in the electrochemical gradient can lead to a number of pathologies.¹⁴

2. Theoretical background

2.1. Enzymes

Macromolecules, i.e. proteins, nucleic acids, and carbohydrates, are the principal components of all living cells. Among the macromolecules, proteins are the largest group. Enzymes are the most diverse class of proteins, because practically every chemical reaction in the cell requires a specific enzyme. Enzymes have been used for practical applications since the ancient times, where they were used for making wine, vinegar and dairy products. In the eighteenth century, enzymes were systematically studied for the first time, by observation of crude extracts from plants or animals, such as peroxidase from the horseradish, α -amylase in grain and gastric juice from carnivores. In the late nineteenth century and early twentieth century, enzymes were purified which made it possible to study the complexes of enzymes with their substrates for the first time, resulting in the first mathematical model (the Michaelis-Menten model) for describing enzyme kinetics. In 1894, the "lock and key" model was proposed, which describes how the three-dimensional shapes of an enzyme and its substrate fit together (Figure 1).^{10, 15}

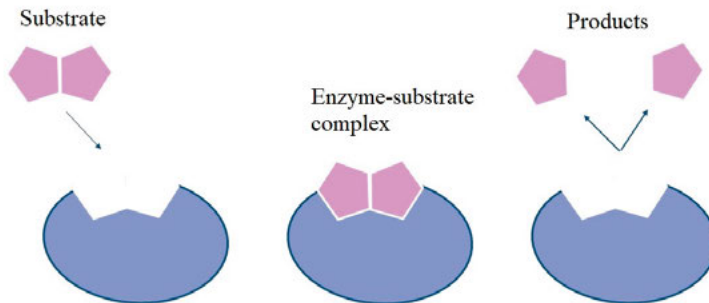


Figure 1: Illustration of the “lock and key” model, which describes how an enzyme and its substrate fit together. Here illustrated for an enzyme that splits a substrate into two products.

In 1958, a modified version of the lock and key model called the “induced fit model” was proposed, describing how the enzyme can change its shape as it interacts with the substrate. Enzymes are required to have a certain level of

flexibility, because their catalytic action is dependent on the exact positioning of the catalytic amino acid residues. The interaction with the substrate causes a change in the three dimensional arrangement of the amino acid residues in the active site, bringing the catalytic groups into a correct position for action. Interaction with other molecules than the substrate does not have the same effect.^{10, 15}

Enzymes, like other catalysts, catalyses chemical reactions by lowering the activation energy required for the reaction to occur, while not being consumed in the process (described by Figure 2).^{10, 16}

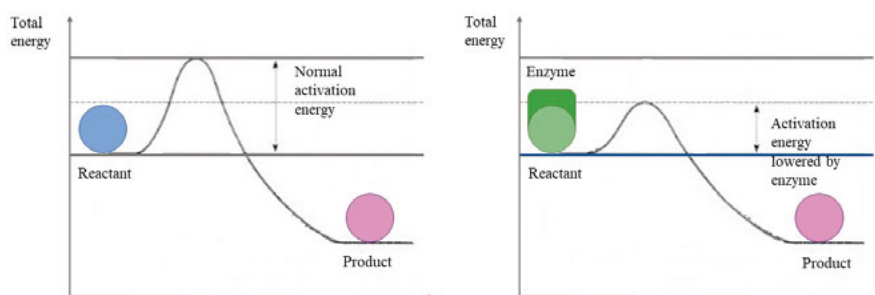


Figure 2: Figure describing the lowering of the activation energy, to enable a chemical reaction.

Enzymes are often globular proteins, containing an active site where the substrate binds. They typically contain one or more allosteric sites at locations separated from the active site, where a molecule binds and regulates the enzyme, either enhancing or inhibiting its catalytic activity. An enzyme may change its conformation upon binding an allosteric regulator. In some cases, the enzyme requires an additional molecule to react, for instance a metal ion (e.g. zinc or iron) or an organic molecule (e.g. 1,4-Dihydronicotinamide adenine dinucleotide (NADH) or S-Adenosyl methionine (SAM)). This is called a co-factor, and the enzymes that require a co-factor therefore contain a co-factor binding site.^{1, 10} Figure 3 shows an enzyme that has been studied in this thesis, highlighting the active site, an allosteric site and a co-factor binding site.

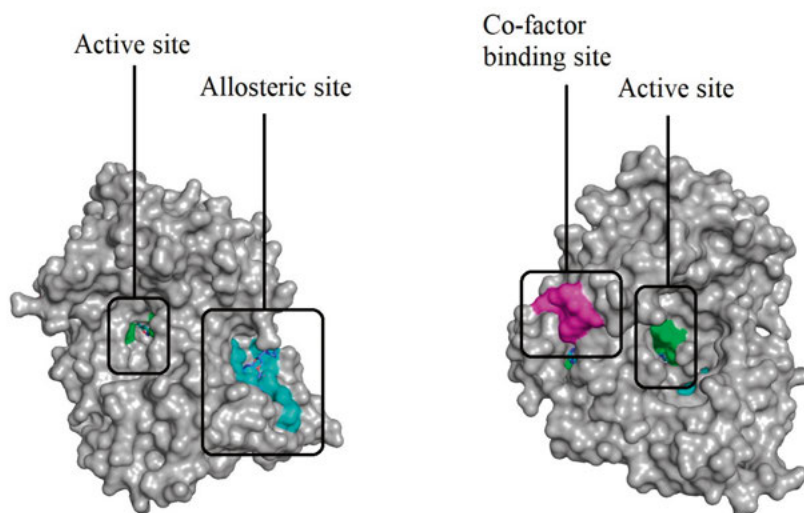


Figure 3: Figure shows two views of an enzyme that has been studied in this thesis, pointing out the active site (green surface), allosteric site (turquoise surface) and co-factor binding site (pink surface). PDB-IDs 6YUH and 6ZRB.

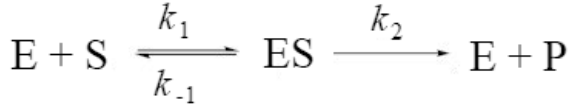
Enzymes are broadly classified into six different classes, based on the type of reaction they catalyse. Oxidoreductases catalyse the transfer an electron from one atom to another. Hydrolases cleave the substrate by taking up a water molecule. Ligases catalyse the formation of a new chemical bond, to ligate two molecules. Transferases catalyse the transfer of a chemical group from one molecule to another.^{1, 17, 18} Lyases catalyse the addition or elimination of water, carbon dioxide or ammonia molecules across a double bond or to form a new double bond. Isomerases catalyse the formation of an isomer, by the transfer of a functional group.^{1, 18}

There are numerous reasons to study enzymes. Because of the crucial role enzymes play in almost every reaction in a cell, malfunctions in enzyme activity, caused by for example underproduction or mutations, can cause a variety of diseases. Thus, enzymes are studied both as possible drugs to treat underproduction, and as drug targets, to inhibit aberrant activity.^{19, 20, 21} Additionally, enzymes have a number of industrial applications, such as in biological detergents, production of biofuels and in the food industry.²²

2.1.1. Enzyme kinetics

Reactions catalysed by enzymes can be studied with various methods, most commonly by kinetic analysis of the reaction determining the catalytic efficiency and substrate affinity with steady state kinetics. Enzymes are studied

in steady state conditions, to allow convenient interpretation of the time courses of enzyme reactions. Scheme 1 describes a catalytic scheme for a simple enzymatic reaction, where the enzyme (E) turns one substrate (S) into one product (P), involving an intermediate enzyme-substrate (ES) complex.¹⁰



Scheme 1: Catalytic scheme for a simple enzymatic reaction.

The rates for the formation (k_1) and dissociation (k_{-1}) of the enzyme-substrate complex, and the progress towards product (k_2), can be used to calculate the kinetic parameter K_M , which describes the affinity of the enzyme to the substrate (Eq. 2.1).¹⁰

$$K_M = \frac{k_2 + k_{-1}}{k_1} \quad (\text{Eq. 2.1})$$

The Michaelis-Menten equation (Eq. 2.2) describes how the reaction rate, v_0 , varies as a function of the substrate concentration, $[S]$.²³ V_{max} represents the maximal reaction rate.

$$v_0 = \frac{V_{max}[S]}{K_M + [S]} \quad (\text{Eq. 2.2})$$

The catalytic activity of an enzyme can be inhibited by interactions with molecules that block the enzymatic activity by one of four types of inhibition mechanisms: competitive, uncompetitive, non-competitive or mixed inhibition. Competitive inhibitors bind to the enzyme when it is not bound to substrate, and thereby competes with its binding of substrate. In this case, the free enzyme binds either the substrate or to the inhibitor, but not both simultaneously. Uncompetitive inhibitors bind to the enzyme when in complex with the substrate. It binds at another location than the active site, since this site is already occupied by substrate. Non-competitive inhibitors can bind both to the free enzyme and the enzyme in complex with the substrate. They bind to the

enzyme at a site distinct from the active site. Mixed inhibition is a mix of uncompetitive and non-competitive inhibition.¹⁰ Figure 4 illustrates the different inhibition modes.

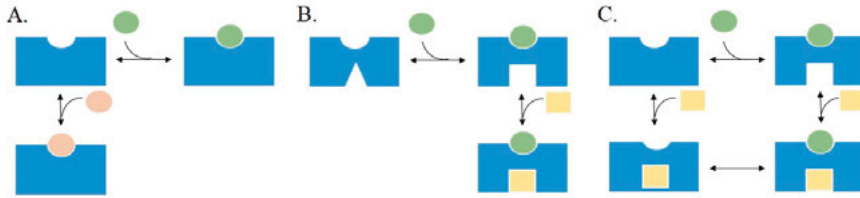


Figure 4: Cartoon representations of the three main inhibition mechanisms: competitive inhibition (A), uncompetitive inhibition (B) and non-competitive inhibition (C). Mixed inhibition is a mix of uncompetitive and non-competitive inhibition.

Michaelis-Menten equations for competitive (Eq. 2.3), uncompetitive (Eq. 2.4) and non-competitive enzyme (Eq. 2.5) inhibition are shown in the equations below.¹⁰ K_i represents the inhibition constant, which describes the affinity between the enzyme and the inhibitor, and $[I]$ the inhibitor concentration.

$$v_0 = \frac{V_{max}[S]}{[S] + K_M(1 + \frac{[I]}{K_i})} \quad (\text{Eq. 2.3})$$

$$v_0 = \frac{V_{max}[S]}{[S](1 + \frac{[I]}{\alpha K_i}) + K_M} \quad (\text{Eq. 2.4})$$

$$v_0 = \frac{V_{max}[S]}{[S](1 + \frac{[I]}{\alpha K_i}) + K_M(1 + \frac{[I]}{K_i})} \quad (\text{Eq. 2.5})$$

In some cases, other models than the Michaelis-Menten model are required to describe enzyme function. Some enzymes occur in complexes of several subunits, which all contain an active site, and these active sites may or may not act independently of one another. The binding of a ligand at one active site can result in structural changes in the enzyme, leading to either an increase or decrease in the affinity of the active sites on the rest of the subunits for ligand binding. This situation is called positive or negative cooperativity, respectively. For enzymes that display cooperativity, the Hill equation (Eq. 2.6), where h represents the Hill coefficient, can be used for the determination of kinetic parameters.¹⁰

$$v_0 = \frac{v_{max}[S]^h}{K_M^h + [S]^h} \quad (\text{Eq. 2.6})$$

2.1.2. Pyrimidine degradation

Pyrimidine degradation in animals is reductive and occurs in three different steps. The first step is the rate-limiting reaction of the pathway, where uracil and thymine are reduced to dihydrouracil and dihydrothymine, respectively, by dihydropyrimidine dehydrogenase (DPD, EC 1.3.1.2). In the second step, these are hydrolyzed to N-carbamoyl- β -alanine (NC β A, β -ureidopropionate) and N-carbamoyl β -aminoisobutyrate (NC β AIBA) by dihydropyrimidinase (DHP, EC 3.5.2.2). In the third step, these compounds are hydrolyzed to β -alanine (β -Ala) and β -aminoisobutyrate (β AIBA), respectively, by β -ureidopropionase, releasing carbon dioxide and ammonia.^{24, 25} Figure 5 is a schematic of the pyrimidine degradation pathway.

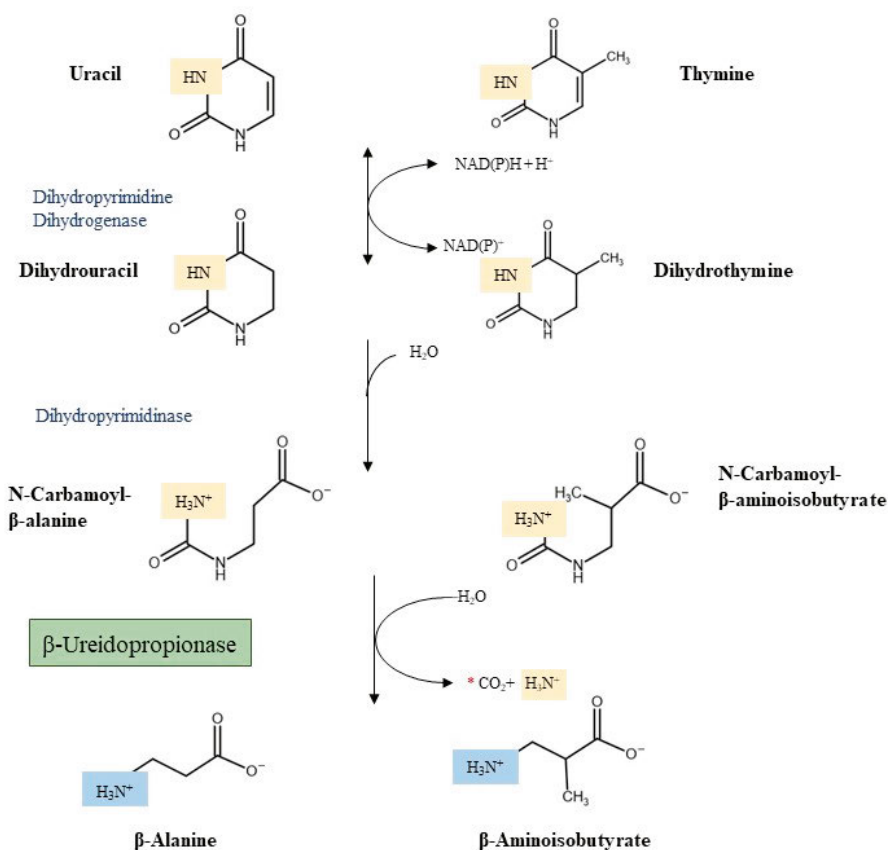


Figure 5: Schematic of the pyrimidine degradation pathway, where pyrimidines uracil and thymine are metabolized into β -alanine and β -aminoisobutyrate, respectively.²⁵

The pyrimidine degradation pathway plays a significant role for several functions in humans. Maintenance of pyrimidine homeostasis is important for the synthesis of the nucleic acids deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The uracil degradation product β -alanine is a modulator of neuropathic pain, as an agonist of γ -aminobutyrate (GABA, the main inhibitory neurotransmitter in the central nervous system) and glycine receptors in neurotransmissions. β -alanine also functions as a precursor of the peptides carnosine and anserine, which are present in high concentrations in muscle and brain tissue. In addition, the product of thymine degradation, β AIBA, has signaling functions in for example fatty acid oxidation and production of leptin.^{24, 25}

Deficiencies in the enzymes in the pyrimidine-degrading pathway caused by genetic mutations are associated with a range of neurological disorders. In addition, it poses a life-threatening risk for cancer patients receiving 5-fluorouracil (5FU)-based chemotherapy, which is used for treatment of gastrointestinal, head and neck, and breast cancer. 1% of patients treated with 5FU develop toxicity with fatal outcome, of which 30-40% are associated with DPD deficiency. β UP deficiency has been confirmed in a limited number of patients, with variable symptoms ranging from asymptomatic to severe neurological problems.^{24, 26, 27}

2.1.3. Epigenetics and the enzymes' role in disease

DNA is the hereditary material of almost all living organisms, and is made up of a double stranded helix, composed of purine and pyrimidine nucleotides paired with weak hydrogen bonds. The helixes are wrapped around proteins called histones, forming nucleosomes. In this open form, the chromatin is called euchromatin. Multiple nucleosomes pack tightly together in the compact form called heterochromatin. The complete set of genetic information in the DNA of an organism is called the genome. Epigenetic modifications are defined as inheritable changes of gene expression, without alterations of the DNA sequence, and the complete set of modifications is called the epigenome. These modifications regulate gene expression by regulating the chromatin structure, which can involve for example DNA methylation or acetylation. The openly structured euchromatin is generally connected to active transcription and the following expression of a gene, while the tightly packed heterochromatin is connected to transcriptional repression, meaning the gene is inactivated. Regulatory enzymes, such as histone methyltransferases, histone deacetylases, and DNA methyltransferases, catalyze epigenetic modification reactions.²⁸ One example of an epigenetic enzyme studied in this thesis is SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) domain containing protein 3 (SMYD3), which is a protein lysine methyltransferase, implying that it methylates lysine residues. It has a number of different targets, such as histones, mitogen activated protein kinase kinase (MAP3K2) and human epidermal growth factor (HER2).^{29, 30}

Environmental factors, such as diet, exercise, stress, and toxic chemicals, trigger activation and deactivation of genes through epigenetics, and the specific genetic and epigenetic profile of an individual affect the response to these environmental factors. This interaction between the external and the internal environments are necessary for development and health of living organisms, and influence diseases. This allows the organism to adapt to through expressing a phenotype in response to environmental factors. Thus, the epigenome is dynamic and flexible, unlike the DNA sequence which in general does not change during the lifetime of an organism.³¹ The epigenome also plays a role

in diseases, such as Alzheimer's disease, arthritis, cardiovascular disease and cancer.³² Studies have shown a correlation between dysregulation of epigenetic enzymes, which are caused by mutations in the amino acid sequence of epigenetic regulatory enzymes, and tumor onset and progression.³³ For example, aberrant activity of the lysine methyltransferase SMYD3, described above, can cause tumor onset and progression through incorrect methylation of its substrates. One case that has been described is the irregular methylation of histone H3, which causes the recruitment of the enzyme RNA polymerase II, together with its associated transcription factors, in certain regions the DNA. This activates genes that promote tumor growth. Several small molecule inhibitors targeting epigenetic enzymes, such as SMYD3, have shown promising anti-cancer efficacy.^{29, 30}

2.2. Protein structure determination

Proteins consist of amino acid residues, which arrange in three-dimensional structures. The primary structure of a protein is the sequence of amino acids in the polypeptide chain, which depends on the DNA sequence of the gene that encodes that specific protein. The secondary structure describes the content of α -helices and β -sheets, formed via hydrogen bonds in the peptide backbone.^{34, 35} The tertiary structure describes the overall three-dimensional shape of a polypeptide chain.³⁶ Soluble proteins are typically folded in an aqueous environment in a manner where the side chains of hydrophobic amino acids are positioned in the interior of the protein, while the hydrophilic side chains are mostly placed on the outside.³⁷ Charged amino acids can stabilize the structure via electrostatic interactions, while sulfur atoms of cysteine side chains can form covalent disulfide linkages.³⁸ Many proteins occur in complexes of two or more subunits, which is described by the quaternary structure. The final three-dimensional shape of the protein is important for its function, such as its interactions with chemical compounds and with other proteins, or its catalytic activity. Interactions with compounds or other proteins can in turn change a proteins' structure, for example induce it to form larger complexes or change its conformation.³⁹

The techniques most commonly used for determining a protein structure are X-ray crystallography (XRC), cryogenic electron microscopy (cryoEM) and nuclear magnetic resonance spectroscopy (NMR).⁴⁰ In XRC, the structural information comes from the X-ray diffraction pattern of protein crystals.⁴¹ In cryoEM, the structural information comes from the movies of the overall shape of the molecule, collected from the electron microscope.⁴² In NMR spectroscopy, the structural model is derived from the information on the local conformations and distances between atoms in close proximity to one another.⁴³

XRC has been the mainly used method for structure determination for a long time, with advantages including near atomic resolution and few limitations in size of the protein or protein complex. However, not all proteins can be crystallized.³ CryoEM allows for structure determination of large complexes and heterogeneous samples, typically not possible by XRC.⁹ Additionally, cryoEM can provide complementary information about a protein or protein complex for which a crystal structure already exists. For example, different structural species present in the sample, such as various oligomeric forms or conformations, can be identified using cryoEM. One of the papers included in this thesis describes this. Structure determination with NMR has the advantage that it is possible to obtain information on the dynamics of the molecule since it is in solution, at conditions close to the proteins' natural conditions.⁴⁴ However, the structure determination is limited to small proteins and complexes, generally with molecular weight <30 kDa.⁵

2.2.1. X-ray crystallography

Since the elucidation of the structures of hemoglobin and myoglobin in the 1950s, the number of protein structures determined by XRC have slowly but steadily increased. The introduction of sophisticated computer hardware and software have reduced the time required to obtain a structure, while increasing the accuracy of the results.³

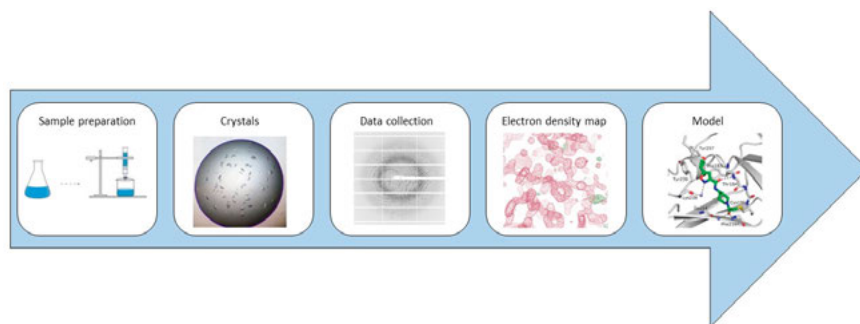


Figure 6: Flowchart showing the workflow of an XRC experiment.

The flowchart in Figure 6 shows an overview of the XRC experimental procedure. The sample requirements are generally high; the chances of growing crystals increase with sample purity and all protein molecules in the sample should share surface properties (e.g. charge). The protein crystallization, where the protein slowly precipitates from its solution, is generally the bottleneck of an XRC experiment. It is often a trial- and error process, where the protein is dissolved in a solvent including a salt or an organic compound as well as a precipitant solution such as polyethylene glycol (PEG). The solution

is then brought to supersaturation, where small aggregates, nuclei for crystal growth, form. Figure 7 illustrates the solubility of a protein as a function of the concentration of precipitant. After nuclei formation, the crystal starts growing. To optimize the crystallization, it is common to optimize sample preparation (e.g. higher sample purity), vary the concentration of precipitant such as PEG, salt or organic compound, and change the pH or temperature. These parameters are generally varied over a grid format, or using pre-made kits for automatic crystal plate setups using crystallization robots.^{3, 46, 47}

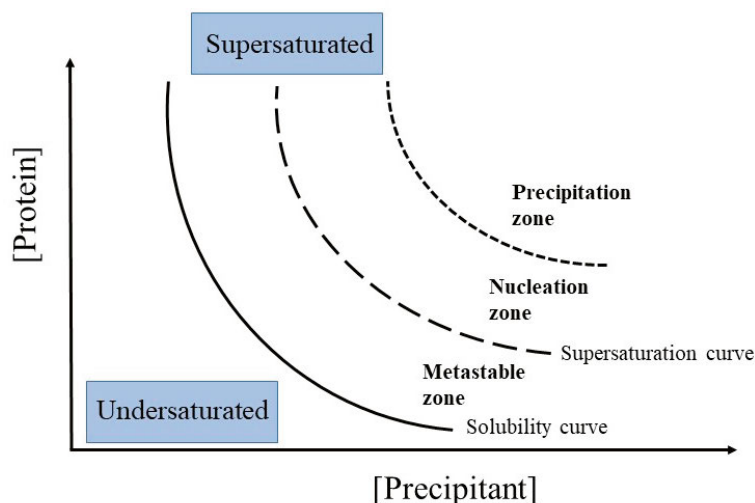


Figure 7: Phase diagram showing the solubility of a protein in solution as function of the concentration of the precipitant. In the undersaturated zone, no crystals can grow, while in the precipitation zone, supersaturation is too high and the protein precipitates without forming crystals. Below the precipitation zone is the nucleation zone, where the supersaturation is high and nucleation is observed, but the crystal growth is slow. In the metastable zone, the nucleation does not occur spontaneously, but the supersaturation is low and suitable for crystal growth.^{3, 46, 47}

In a protein crystal, the protein molecules are packed in a regular way in a three dimensional periodic arrangement. Three repeating vectors can be recognized: a , b and c , with angles α , β and γ between them. These vectors and angles define a unit cell in the crystal lattice (Figure 8). The unit cell can thus be defined as the smallest group of atoms that make up the entire crystal lattice, when repeated using translations. The smallest repeating groups of atoms that make up the unit cells when rotated and/or translated, is called an asymmetric unit, and may contain multiple copies of the protein structure. The regular packing of the protein molecules in the crystal lattice leads to a symmetric relationship between the molecules. The space group is a classification based on the symmetry of the crystal and the arrangement of the asymmetric units.

There are 230 possible space groups, but only a few of these can be adopted by protein crystals.^{3, 47}

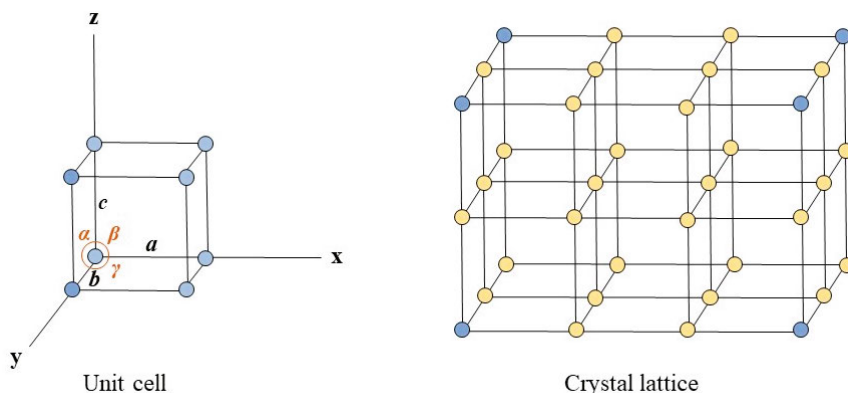


Figure 8: To the left, a unit cell, the smallest repetitive unit of a crystal. To the right, the unit cell repeated in a three dimensional pattern.^{3, 47}

Once the crystal has grown, the XRC experiment is generally carried out at a synchrotron radiation facility. Diffraction is a phenomenon that occurs when light encounters an obstacle. At the synchrotron, the X-ray beam, which has a wavelength of $10^{-7} - 10^{-11}$ m (1000-0.1 Å) hits the surface of the electron cloud surrounding the atoms of the protein crystal at an angle θ , resulting in some of the light being diffracted at that same angle away from the solid (Figure 9). The remainder of the X-ray light will continue into the crystal, where some of the light will interact with the second plane of atoms and some of the light is once again diffracted at an angle θ , the remainder continuing further into the protein crystal. The diffraction repeats for the remainder of the planes in the crystal. The X-ray beams travel different path lengths before it reaches the planes of the crystal. After diffraction, the X-ray beams interact constructively if the path length difference is equal to an integer number of wavelengths, which is described by Braggs law (Eq. 2.7). Figure 9 illustrates the diffraction of the X-ray beams at the different planes in the crystal.^{3, 47}

$$2d \sin\theta = \lambda \quad (\text{Eq. 2.7})$$

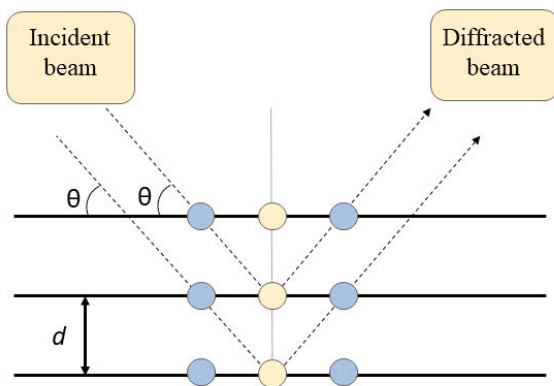


Figure 9: Reflection of the X-ray beam from the crystal lattice planes, described by Braggs law of diffraction.^{3, 47}

For data collection, an X-ray source and an X-ray detector are required. The X-ray source is most commonly a synchrotron, where X-ray radiation of exceptionally high intensity is generated when electrically charged particles circulate at a speed close to the speed of light. Data collection can also occur in a home laboratory, with a rotating anode or sealed tubes as an X-ray source. A monochromator is used to transmit X-ray radiation of a narrow band of wavelengths. The crystal is rotated in the beam, and subjected to the X-ray beam from different directions. The X-ray beams that are scattered by the crystal are recorded by a detector, resulting in a diffraction pattern. The diffraction patterns are transformed into electron density, $\rho(x y z)$, for which the fundamental equation is the Fourier transform of the structure factors $F(h k l)$ (Eq. 2.8).^{3, 47}

$$\rho(x y z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{[-2\pi i(hx+ky+lz)+i\alpha(h k l)]} \text{ (Eq. 2.8)}$$

During the experiment, the amplitude of the diffraction pattern spots are measured but not the phase information, which is required to place the electron density properly in the unit cell. This is referred to as the “phase problem” of XRC, and can be solved by a variety of techniques. Molecular replacement is the fastest and most commonly used technique, and is possible if there is already a similar structure available to use as a model. The principle is that Patterson maps (Patterson function is shown in Eq. 2.9) are generated of both the new, unknown structure and the previously determined structure and the latter is used to place the unknown structure in the proper unit cell.^{3, 47} Figure 10

illustrates molecular replacement of a previously known model of a protein in a newly obtained electron density map.

$$P(u\ v\ w) = \sum_{hkl} |F|^2 e^{-2\pi i(hu+kv+lw)} \quad (\text{Eq. 2.9})$$

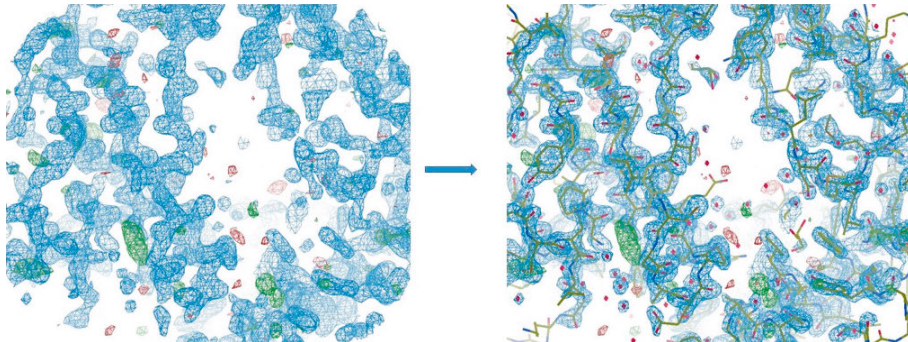


Figure 10: Figure showing the molecular replacement of a previously known protein model into electron density from an XRC experiment.

2.2.2. Cryogenic electron microscopy

CryoEM is a significantly newer technique than XRC. Though the first electron microscope was invented in the 1930s⁴⁸, the first successful application of cryoEM for structure determination was in 1984, when the structures of five vitrified viruses were determined.⁴⁹ For a long time, resolution has been low for cryoEM structures. However, recent improvements have been made to the technique.^{50, 51} Currently, the majority of the protein structures from cryoEM are at resolutions between 3 and 4 Å. However, a growing number are in the 2–3 Å range, and a few are at a resolution higher than 2.0 Å.^{52, 53} The current record is a 1.54 Å resolution structure (EMD-9865) of apoferritin^{54, 55}. Nature magazine named cryoEM as the “Method of the year” in 2015⁵⁶, and in 2017, the Nobel Prize was awarded for the development of cryoEM for high-resolution structure determination of biomolecules in solution⁵⁷. Figure 11 shows an overview of the steps of a cryoEM experiment, including the sample preparation, vitrification on grids, data collection, and the following data analysis.

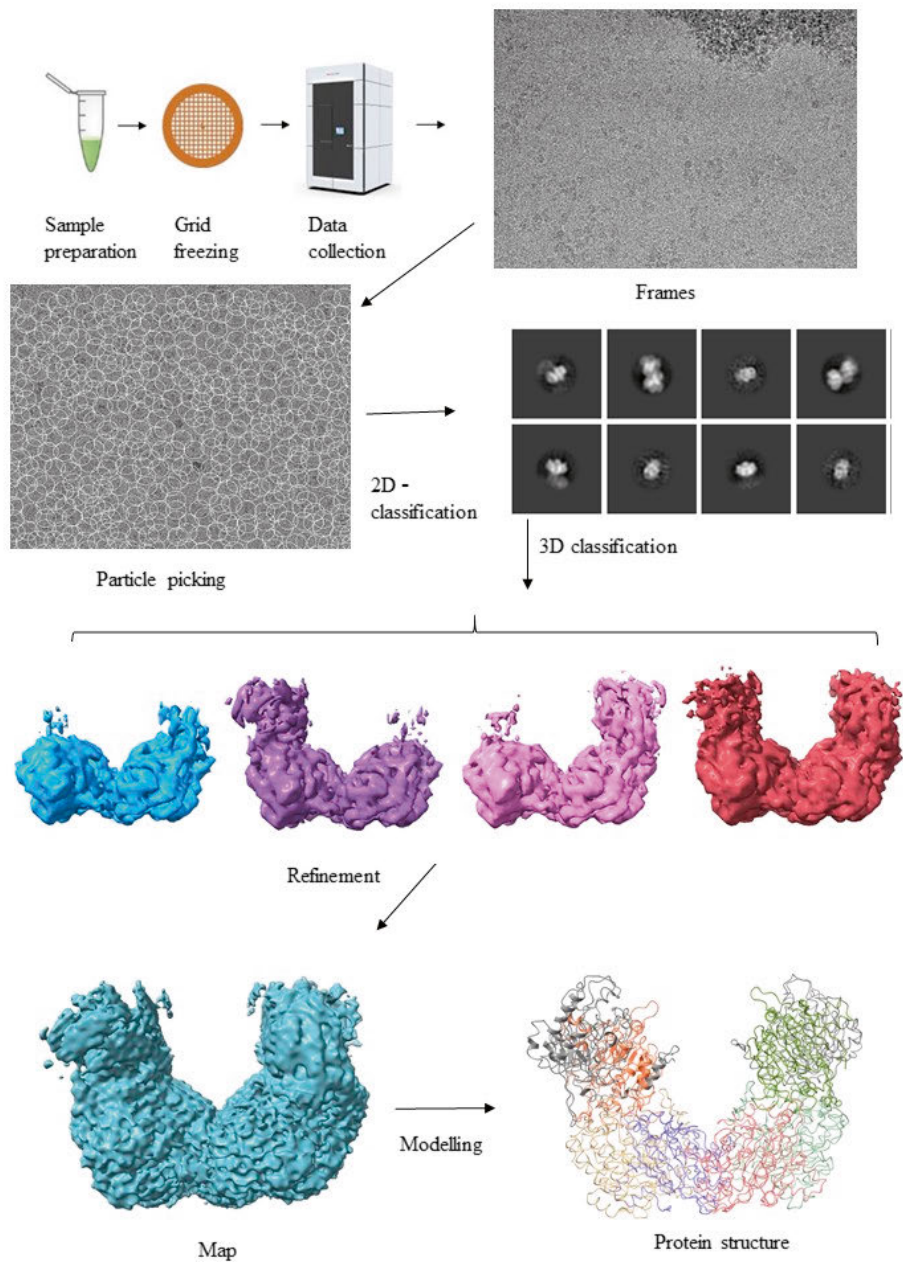


Figure 11: Schematic depicting the steps of a cryoEM experiment, including sample preparation, sample vitrification on grids, data collection in a TEM microscope, particle picking, 2D classification, 3D classification and model building.

The sample requirements for cryoEM are generally lower than for XRC and NMR, regarding both the amount and the purity⁵⁸. Recently, structures for

several proteins directly from a lysate sample were determined using cryoEM^{59, 60}. Thus, several protein structures can be generated from one single experiment.

EM uses the fact that electrons have a very small wavelength, many times shorter than the wavelength of light, making it possible to use electron beams to take images of very small objects such as proteins. The protein needs to be protected from damage from the high intensity of the electron beam. Initially, experiments with EM used dehydrated samples, such as negative staining technique, which pose a risk for introducing artefacts in the structure. By freezing samples in a thin layer of a noncrystalline form of solid water, such as vitreous ice, the sample could be preserved in a condition closer to its native state. The sample is vitrified into a glass-like state on a grid, usually made by gold or copper, by plunge freezing it in liquid ethane. Apart from protecting the sample from the electron beam, the plunge freezing also stops the movement of the molecules, making it possible to obtain clear images.^{61, 62, 63}

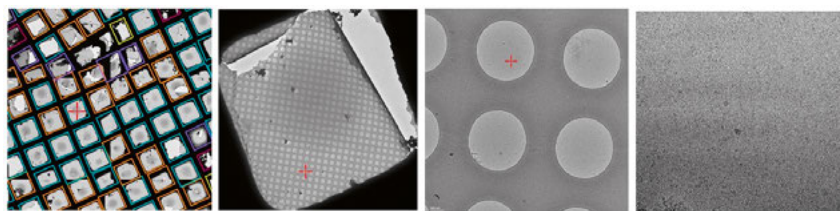


Figure 12: Figure from data screening using a TEM. The data screening was performed with the Glacios microscope at the CryoEM facility at Uppsala University.

After sample vitrification, the grids are screened and data is collected in a transmission electron microscope (TEM). An EM uses a beam of electrons as a source of illumination, and consists of an electron emission source, electromagnetic lenses and an electron detector. The grid with the vitrified protein is placed in the TEM, where the electron beam travels through the very thin sample.⁶³ Since the proteins are floating randomly in the sample solution, they are in various orientations when they are vitrified. Thus, the particles imaged by the detector are all unique.

The movies collected in the microscope are motion corrected which means correcting for the movement of the sample. This can occur both globally due to stage drift or locally due to beam-induced anisotropic sample deformation. Following this, a contrast transfer function (CTF) estimation of the images is done, to estimate and correct for aberrant modifications of the image from the TEM. Next, particles are picked automatically or manually, and the picked particles are classified into a number of 2D classes. Following this, the particles from the chosen 2D classes are classified into different 3D classes and the

final map is used for building and refining the protein structure model.^{64, 65} Figure 11 illustrates this entire process.

2.3. Drug discovery

A drug is defined as a selective chemical that is used or has potential to be used in diagnosis or treatment of disease. Plants have been consumed as a cure for various conditions for thousands of years. There is some evidence that the Neanderthals consumed medicinal plants^{66, 67}. Modern drug discovery started with the isolation of alkaloids from plants, such as morphine extracted from opium poppy plant in early 1800s. Chloral hydrate was the first synthetic drug, discovered in 1869, and was still available in some countries in the early 2000s.⁶⁸ Today's drug discovery research is a multidisciplinary field including several stages.⁶⁹ This thesis focuses on the early stages of drug discovery, described in the flowchart in Figure 13.

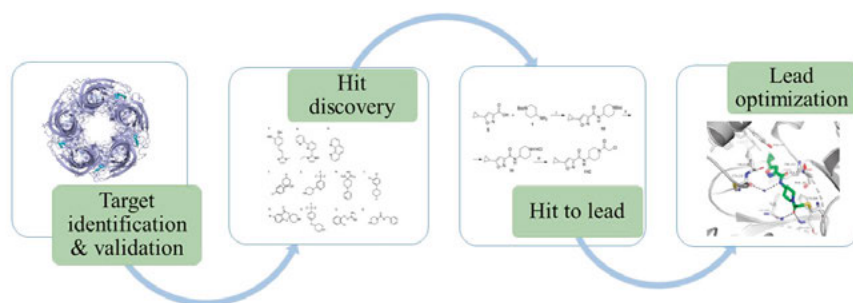


Figure 13: Scheme of the early drug discovery process.⁶⁹

The first step of the drug discovery process is to identify a drug target. The most common classes of drug targets include G-protein coupled receptors, enzymes, hormones and ion channels, of which two examples are studied in this thesis.⁷⁰ It is important that the target is “druggable”, which means that it is accessible for drug molecules and that there is a biological response that is measurable both *in vitro* and *in vivo*, so that the function of a new drug can be confirmed. Once a suitable drug target has been identified, it is validated through a variety of methods. Its involvement in disease can be confirmed by for example gene silencing using small interfering RNA (siRNA).⁶⁹

There are several approaches for identification of hit molecules, which is usually defined as molecules that fulfil the desired criteria for detection of the readout in a screen. In high-throughput screening (HTS), the focus is on efficient screening of a large library of drug-like compounds, i.e. compounds with

chemical and physical properties described by Lipinski's rule of five.⁶⁹ Fragment-based drug design (FBDD) is a relatively new approach, which was explored in two of the projects described in this thesis. In FBDD, the lead compounds are developed from smaller compounds ("fragments"), which typically have a molecular weight <300 Da and a low affinity for the target. Figure 14 illustrates the principle and process of FBDD.

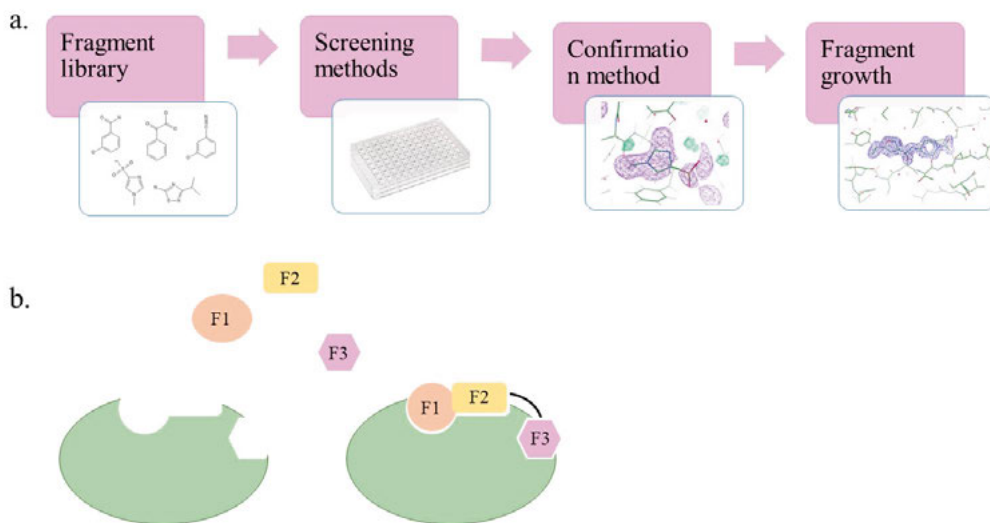


Figure 14: a. Scheme overview of fragment based drug discovery procedure. b. Illustration of the building of a drug-like molecule from fragments (F1, F2 and F3).⁷¹

The lower affinity of the fragments compared to a drug-like compound implies that a screening method with high sensitivity is required, often a biophysical method, such as XRC, NMR or a biosensor-based technique.^{71, 72} Hit identification can also be done *in silico*, using virtual screening^{69, 73}. Confirmation of the binding can be done using for example XRC, NMR, or cell-based assays, while biochemical assays can be used to confirm functional effects. Structural information is valuable for the understanding of the binding of the fragment hits, to enable the following fragment growth which can be done by for example growing one of the fragment hits or linking two or more of the fragments together. Advantages of FBDD include lower experimental costs, since a smaller number of molecules need to be screened, and the opportunity to explore drug-like compounds that may not be present in HTS libraries.⁷²

With structure-based drug design (SBDD), a structure of the drug target with a hit compound docked is used to predict which alternations could be made to the compound to improve its potency or selectivity for the target.⁷⁴ The larger

number of structures available makes it possible through machine learning to perform virtual screening of a target, which can save time and resources in comparison to an experimental screening.⁷³

Usually, few hit series result from the hit identification step. When these have been identified, they are refined into more potent and selective compounds.⁶⁹

2.3.1. Ligand-gated ion channels as drug targets

LGICs are a group of proteins located in the cell membrane, consisting of one transmembrane domain and one extracellular domain, and in some cases also an intracellular domain. Binding of a ligand to the extracellular part of the protein induces large conformational changes, which opens the ion channel to allow ions to pass through the membrane.⁷⁵ The ligand can be for example a neurotransmitter. LGICs are divided into three superfamilies: Cys-loop receptors, ionotropic glutamate receptors and ATP-gated channels.

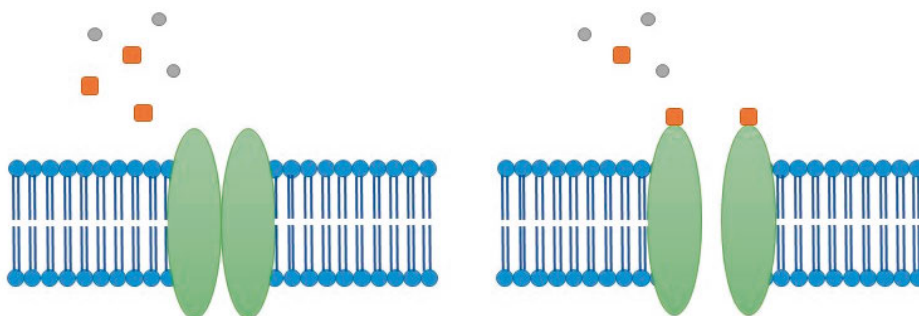


Figure 15: Illustration of the function of a ligand-gated ion channel, opening up to allow ions to pass through the cell membrane following ligand binding.

Ligand gated ion channels play an important role in controlling a large variety of physiological processes. Dysfunction can lead to pathophysiology in different processes, and because of this ligand gated ion channels are important drug targets for drugs against cancer.

In this thesis, a homologue of a Cys-loop type of LGIC is studied. Acetylcholine binding protein (AChBP) originally comes from the freshwater snail, *Lymnaea stagnalis*, where it is present in the central nervous system to modulate synaptic transmissions.¹² Its homopentameric structure is highly similar to that of the extracellular domain of the nicotinic acetylcholine receptor (nAChR) in humans, which is a Cys-loop type of LGIC. Additionally, a majority of the amino acid residues in the ligand binding sites of nAChRs are also found

in the subunit interfaces of the homopentameric AChBP. Thus, AChBP is often used as a model for the extracellular, ligand-binding domain of nAChRs.⁷⁶

3. Present investigations

The structure of a protein and its interactions with ligands or other proteins, provides an understanding of its function, which helps us hypothesize on how to affect or use the protein. Here, I describe three projects, each focusing on a specific protein with unique properties. I describe the various approaches that have been taken to acquire structural information about the protein of interest, and how this information was used to answer scientific questions.

In the first project described, cryoEM was used to determine the structure of a large oligomeric complex of a drug metabolizing enzyme in the pyrimidine degradation pathway. This information, in combination with biochemical studies of the protein in presence of substrate and product analogous compounds, and after substitution of important amino acid residues, enabled characterization of the regulation of the enzyme.

In this thesis, also two different types of drug targets have been studied: an epigenetic enzyme and a soluble homologue of the Cys-loop receptor class of LGICs. The interaction between drug targets and potential interaction partners were studied structurally, in combination with other techniques, to obtain information of the interaction between the drug target and the compounds of interest. Protein structures of the drug targets in presence of small molecules, obtained from XRC, have revealed new binding sites as well as ligand-induced conformational changes.

In all studies involving protein structures, several questions have to be asked to determine which method is best to be used for that specific project and protein of interest. What scientific questions are we looking to answer, and which technique will best do that? What are the properties of the target protein? The techniques generally used for the study of protein structures, described in previous sections, come with advantages as well as challenges, and it is important to understand the principle of the method in order to choose the right one and to obtain the information sought after.

3.1. Characterization of the allosteric regulation of a drug metabolizing enzyme (Paper I)

β -ureidopropionase (β UP, EC 3.6.1.6, also called β -alanine synthase or N-carbamoyl- β -alanine amidohydrolase) is a nitrilase like enzyme that catalyzes the third and last step in the reductive pyrimidine degradation pathway (see scheme in Figure 5). Deficiencies in β UP, and other pyrimidine-degrading enzymes, disrupts the homeostasis of pyrimidines and their metabolites, which causes neurological disorders. Additionally, it increases the risk for life-threatening side effects from the anti-cancer drug 5-fluorouracil (5FU), which is frequently prescribed for treatment of various cancer forms.^{24, 25}

In physiological conditions in the human body, β UP occurs in a mixture of different oligomer forms ranging from dimers to octamers, and possibly larger assemblies. β UP is activated by interacting with its substrate NC β A, which causes the enzyme to associate into larger oligomer complexes. The interaction with its product β -Ala causes the enzyme to dissociate into a lower oligomer form, inactivating the enzyme (Figure 16). The shifts between the various oligomer forms can also be induced by shifts in pH, indicating that enzyme activation is dependent on the charge of functional groups of certain amino acid residues. At pH 5.0, β UP occurs mainly in the activated, higher oligomer complex, while a higher pH of 9.0 prevent the association into the higher oligomer forms (Figure 16).²⁵

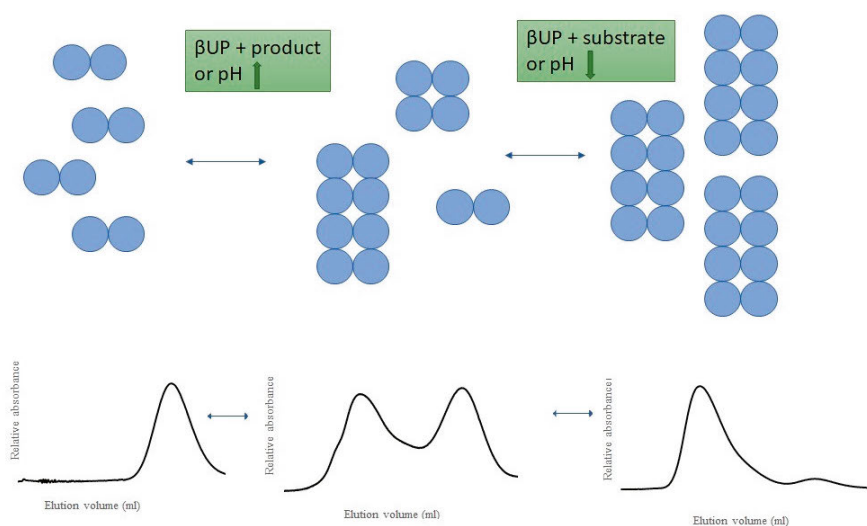
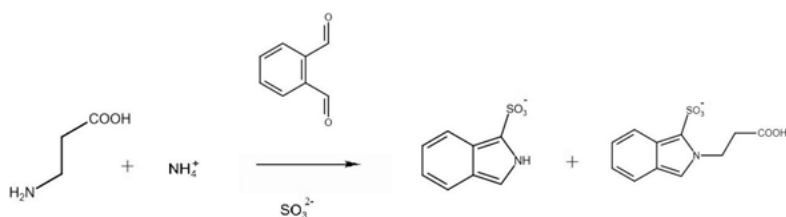


Figure 16: Schematic of the ligand and pH dependent shift between oligomer states in β UP, and the chromatograms from the oligomer forms from size exclusion chromatography (SEC).

Crystal structures of the inactivated dimeric form of Hs β UP²⁵ (PDB-ID: 6FTQ) and the activated higher oligomeric form of Dm β UP⁷⁷ (PDB-ID: 1UF7) have revealed that the enzyme structure contains three flexible loops at the entrance to the active site (EL1, EL2 and EL3). These loops are disordered and freely accessible in the inactive state of the enzyme. Activation of the enzyme causes the loops to become structured and buried in dimer-dimer interfaces of the larger oligomer complexes. EL2 contains the amino acid residue Glu207, which is also part of the conserved active site. The interaction between NC β A and Glu207 may cause the conformational changes necessary for formation of larger complexes. Glu119 aids a nucleophilic attack from C233 on the substrate carbamoyl carbon.

The reaction between the product β -Ala and O-phthalaldehyde yields isoindoles with fluorometric activity that can be measured in a spectrophotometer at 460 nm (see Scheme 2).²⁵ This reaction was used to measure the activity of β UP in the study described here.



Scheme 2: Reaction scheme of the conversion of β -Ala to O-phthalaldehyde.²⁵

3.1.1. Site-directed mutagenesis revealed function of crucial residues

Site-directed mutagenesis was used to substitute amino acids that were hypothesised to play a role in complex formation and catalytic activity. Table 1 and Figures 17 and 18 show the location and hypothesised roles of the residues. Each of the mutated variants was analysed with a thermal shift assay to confirm that the engineered enzyme was folded and stable after expression and purification.

Table 1: Variants of β UP produced, including their location in the enzyme and their hypothesised function.

Site	Location	Role
C233	Active site	Catalysis
H173	Dimer-dimer interface	Oligomerization and allosteric regulation
H307	Dimer-dimer interface and EL3	Oligomerization and allosteric regulation
E207	Active site and EL2	Catalysis and possibly oligomerization

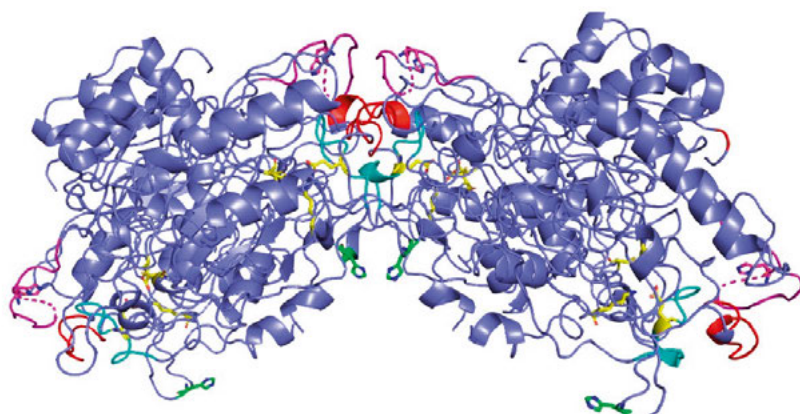


Figure 17: Illustration of the β UP tetramer (PDB-ID: 8PT4). Entrance loops are illustrated in red (EL1), turquoise (EL2) and magenta (EL3). Active site tetrad of residues are depicted in yellow. Amino acid residues targeted for site-directed mutagenesis are depicted in yellow (C233 and E207), magenta (H307) and green (H173).

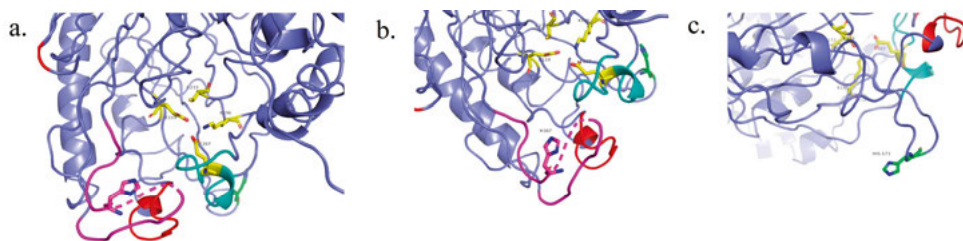
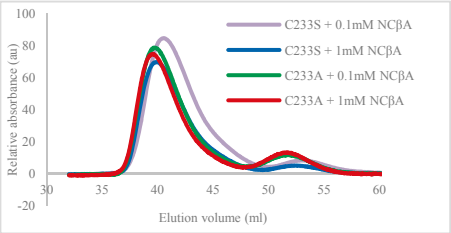
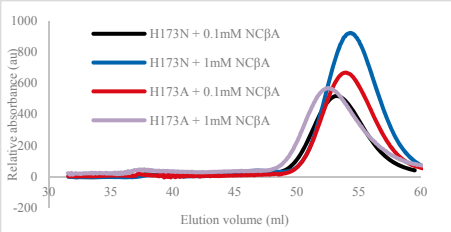
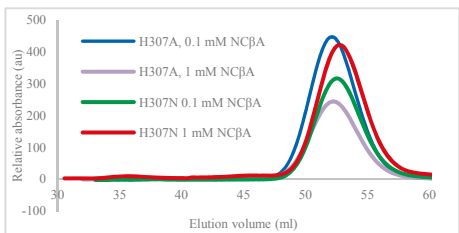
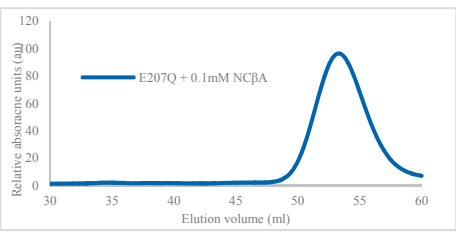


Figure 18: Close up view of active site amino acids (a), H307 (b) and H173 (c).

Size-exclusion chromatography (SEC) at physiological pH in presence of NC β A revealed that the H173, H307 and E207 variants occurred exclusively as dimers, and were thus unable to form higher oligomers. The C233 mutants occurred mainly as higher oligomers, which is an indication that they can form higher oligomers but not react, and that the enzyme was therefore saturated with the substrate. The enzyme activity assay, which utilized the reaction described in Scheme 2 to measure the reaction product, revealed that all engineered variants of β UP had lost their catalytic function. The results from the SEC and activity studies are summarized in Table 2.

Table 2: Effect of the mutations on β UP oligomer state and catalytic activity.

Site	Oligomer state	Activity
C233		Inactive
H173		Inactive
H307		Inactive
E207		Inactive

3.1.2. Substrate and product analogues have effect on oligomer formation and catalytic activity

In this part of the study, five small substrate or product analogues (isobutyric acid, 2-aminoisobutyric acid, glycyl-glycine, glutaric acid and N-carbamoylglycine) were studied to investigate their effect on complex formation and the catalytic activity of the enzyme. From the results, it could be hypothesised which functional groups in these ligands may interact with the active site, or a potential allosteric site, of β UP. Three of the investigated compounds had an effect on catalytic activity and complex formation of β UP, as seen in the activity assay and by SEC. Figure 19 summarizes the result of the analysis of oligomerization and catalytic activity of β UP in presence of the inhibitors.

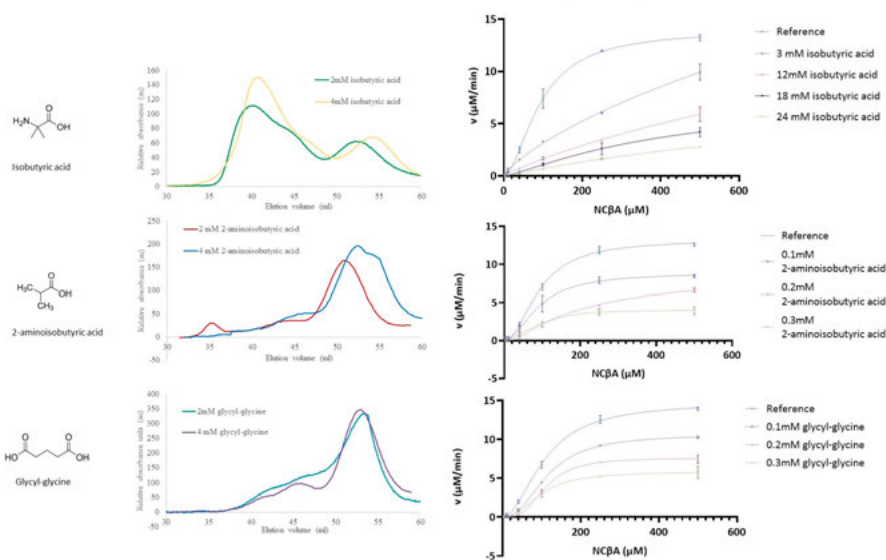


Figure 19: Analysis of the effects of substrate and product analogues on the catalytic activity and oligomeric state of β UP. Molecular structures of compounds, saturation curves and SEC chromatogram for wild-type β UP in the presence of the three compounds that showed effect on the activity and oligomerization of β UP.

Since β UP displays cooperativity, the Hill equation (Eq. 2.6) was used for non-linear regression analysis of the saturation curves and calculation of the kinetic parameters. Isobutyric acid displayed a competitive inhibition mode, while 2-aminoisobutyric acid and glycyl-glycine displayed non-competitive binding modes. Glutaric acid and N-carbamoylglycine did not have any inhibitory effect.

3.1.3. CryoEM structure provide first insights into activated oligomer complex

CryoEM was used to determine the structure of the higher oligomer, activated form of β UP. To obtain the activated, higher oligomer forms, two approaches were taken: C233S mutant in presence of NC β A in pH 7.0, and wild-type β UP in pH 5.0. Data processing from the latter resulted in 603.157 particles from 2D classification that were chosen for 3D-classification. The 3D classification shows that the sample is heterogenous, containing higher oligomer complexes ranging from tetramers to octamers (Figure 20).

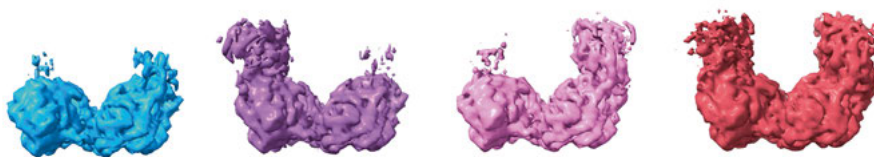


Figure 20: 3D classification shows that the β UP sample is heterogenous, containing higher oligomer complexes ranging from tetramers to octamers.

This analysis was followed by non-uniform (NU) refinement of 163.949 particles from 3D classification, which yielded a final map **1** of the octamer form of β UP, with a resolution ranging from ~ 3 - 3.5 Å in the “inner” four subunits to ~ 6 - 7 Å in the four “outer” flexible subunits. In order to attempt to model the lower resolution subunits of the octamer, a new 3D-classification and NU-refinement were done with focus masks applied to the top part of the protein complex. This resulted in a final map **2**, with a resolution ~ 4.5 Å over the entire octamer complex. Local resolution maps are depicted in Figure 21.

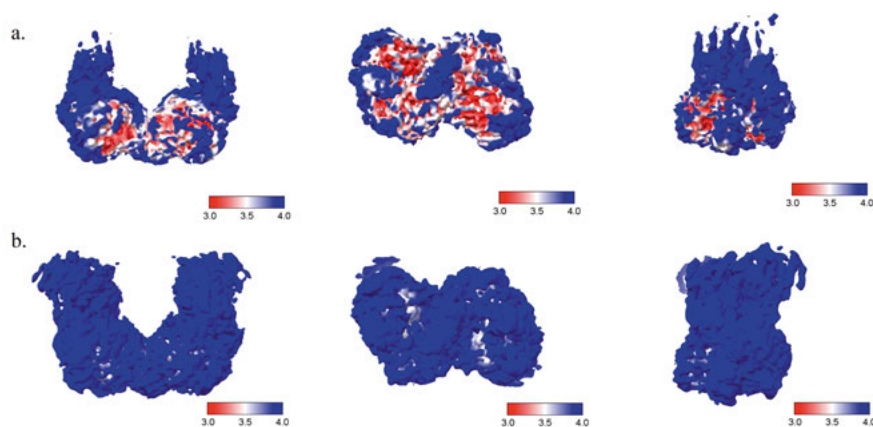


Figure 21: Local resolution maps, a) Map 1 and b) Map 2, from three different views of β UP.

Map 2 confirmed the shape of the octamer complex, but the complete octamer could not be modelled because the resolution was insufficient. Therefore, the tetramer was modelled from map 1 (Figure 22 shows the tetramer modelled in the map).

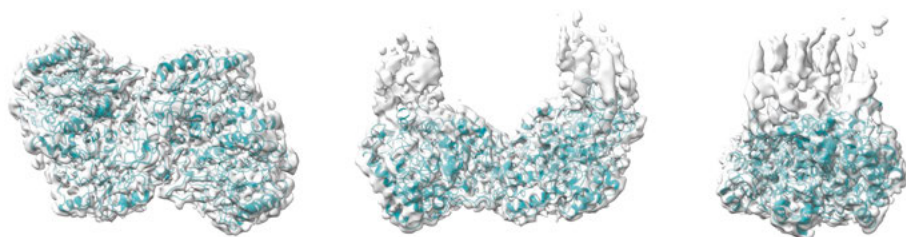


Figure 22: Map 1 (light grey map) and model (turquoise model) of the tetramer (PDB-ID: 8PT4, EMDB-ID: 17867).

3.1.4. Conclusion

In this study, we conclude that amino acid residues H173, located in the dimer-dimer interface, H307, located in the dimer-dimer interface and the EL3 loop, and E207Q, located in the active site and the EL2 loop, are all crucial for the activation of the enzyme. Upon activation, the active site loops become closed, inserting E207 into the active site. This is followed by association of the dimers into tetramers, hexamers and octamers. The association into larger oligomer complexes is in turn essential for catalytic activity, which correlates well with the result from the activity assays where none of the mutated variants showed any catalytic activity. C233 mutants were able to form higher

oligomers and occurred almost only as higher oligomers in presence of NC β A, indicating that this variant was able to form larger complexes but not react, which caused saturation with NC β A.

The study of the inhibition of β UP by substrate and product analogues compounds allowed us to hypothesize how the binding of substrate and product affects the shift between higher and lower oligomer forms. The ability of the compound to interact with the EL2 amino acid residue F205 may be what differentiates activators from inhibitors.

Lastly, we provide the first structure of the activated, higher oligomer complex of human β UP using cryoEM. The tetramer model obtained from map **1** confirmed that the closed entrance loop conformations as well as the dimer-dimer interfaces are conserved between β UP from human and *Drosophila melanogaster*.

3.2. X-ray crystallography in drug discovery – An epigenetic drug target (Paper II, III and IV)

The epigenetic enzyme SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) domain containing protein 3 (SMYD3) is a part of the RNA polymerase II complex, and was at first characterized as an enzyme catalysing methylation of lysine amino acid residue 4 of histone H3 (H3K4). Later studies have revealed other substrates for SMYD3, including lysine 5 of histone 4 (H4K5), lysine 831 in vascular endothelial growth factor receptor 1 (VEGFR1), lysine 260 in mitogen activated protein kinase kinase kinase 3 (MAP3K2), and lysine 175 in human epidermal growth factor receptor 2 (HER2). Figure 23 illustrates the structure of SMYD3, containing an active site and binding site for the co-factor SAM.

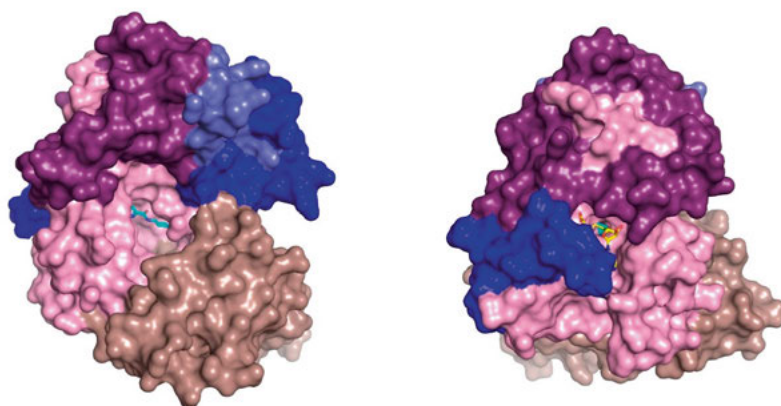


Figure 23: Illustration of SMYD3, including SET-N domain (blue surface), MYND domain (light blue surface) SET-I domain (purple surface), SET-C domain (pink surface) and C-terminal domain (brown surface), with a compound bound (turquoise sticks) in the active site and co-factor SAM (yellow sticks) in the co-factor binding site. PDB-IDs: 6ZRB.

Numerous studies have indicated that the SMYD3 gene is overexpressed in breast, lung, colorectal, hepatocellular, esophageal, and prostatic cancers. The tumor onset and progression is caused by aberrant methylation of for example histone H3 and MAP3K2. These results, in combination with studies confirming that SMYD3 can be safely targeted, have led to an interest in the development of small molecule inhibitor drugs targeting SMYD3.^{29, 30} Here, I present three studies where various approaches were used for the discovery of small molecule inhibitors for SMYD3: rational design of a targeted covalent inhibitor (TCI), discovery of an allosteric inhibitor in a new allosteric site, and FBDD.

3.2.1. Rational design of a covalent inhibitor

In this study, we used a rational design approach to develop a TCI⁷⁸ for SMYD3. An advantage with TCIs is that they generally have a longer residence time than non-covalent inhibitors, and which results in a longer time inhibition. First, a non-catalytic nucleophilic amino acid residue, non-conserved within the protein family, was identified. The amino acid selected for targeting, C186, is located in the substrate-binding pocket, adjacent to the co-factor binding site. A chloride atom for targeting this amino acid residue was introduced to a previously discovered reversible inhibitor containing an oxazole ring, which is a group commonly found in SMYD inhibitors (**11C** in Figure 24a). Additionally, the derivative without the chloride atom cysteine trap

(**11** in Figure 24a) was synthesised. Surface plasmon resonance (SPR) biosensor experiments were performed with **11C** and its non-reactive analogue **11** to confirm the binding and to determine the equilibrium dissociation constant (K_D) for the interaction (Figure 24b). The experiments confirmed that both compounds interacted with SMYD3, with **11C** having a lower K_D value because of its covalent bond to C186.

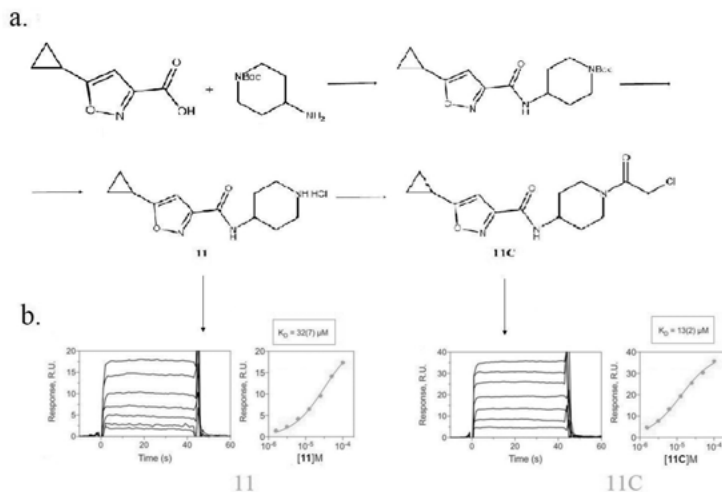


Figure 24: a. Schematic of the synthesis of the TCI **11C**, and its non-covalent analogue **11**. b. SPR sensorgrams for the interaction of compounds with immobilized SMYD3 in the presence of the co-factor SAM, for **11** (left) and **11C** (right).

Co-crystallization experiments of SMYD3 in the presence of **11C** were performed to confirm the binding mode and location of the designed covalent inhibitor. The crystals diffracted at ~ 1.8 Å, and the structure was compared with a previously determined crystal structure of apo-SMYD3 using molecular replacement. The structure confirmed that **11C** binds to SMYD3 at the bottom of the active site, across the β -strand residues 183-186 (Figure 25). It also confirmed the covalent bond between C186 and the chloride atom of the cysteine trap.

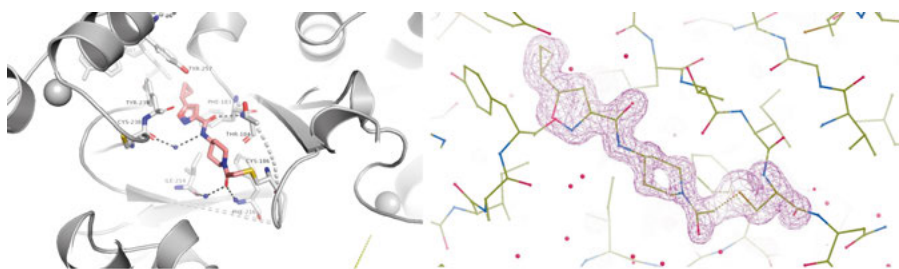


Figure 25: Crystal structure of SMYD3 in complex with 11C. a) The inhibitor is located at the bottom of the substrate binding site of SMYD3, spanning across the β -strand formed by amino acid residues 183–186 and establishing a covalent bond with Cys186. b) F_0-F_c difference density map (purple mesh), confirming electron density of the inhibitor and the covalent bond between ligand **11C** and C186.

3.2.2. Discovery of a new allosteric site

Previously, the compounds explored as SMYD3 inhibitors have interacted with either the substrate-binding site, such as the covalent compound described in the previous study, or the co-factor (SAM) binding site. SAM analogues have a high risk of causing toxic side effects, since SAM has a common role as a metabolite. In this study, we investigated the possibility of an allosteric binder. Targeting allosteric sites has the advantages that allosteric sites are in general less conserved than active sites, which makes allosteric inhibitors generally more specific. Additionally, allosteric sites often have a different structure than active sites, which are usually large and open.

By the start of this study, no allosteric site in SMYD3 had yet been identified. Consequently, there was no structural information available to allow for the rational design of an allosteric compound. In order to discover a potential allosteric site binder, a small compound library was screened in an SPR experiment, where the active site of SMYD3 was blocked to ensure that the possible binder would bind in another location than the active site. By competitive screening using SPR biosensor analysis of a small library, dipiperodone was identified as an allosteric ligand. Since dipiperodone was present in the compound library as a racemic compound, each enantiomer was isolated for validation. SPR experiments showed little difference in the affinities for the two compounds, with K_D values of 42 μM for (S)-dipiperodone and 84 μM for (R)-dipiperodone.

To determine the location of the allosteric binding site, and to compare the binding modes of the two enantiomers, SMYD3 was co-crystallized with the (S) and (R) enantiomers separately, as well as with the racemic mixture to establish if electron density was more often found for one of the forms. The

crystallographic study showed that both enantiomers of dipiperdone bound in the same allosteric site, with a similar binding mode, except for the ring that points out from the enzyme surface in different directions depending on the enantiomeric form. The crystal structure obtained from SMYD3 co-crystallized with the racemic mixture yielded a structure with poor electron density at the location of the flexible ring, indicating that both forms bind to the allosteric site to a relatively similar extent. Figure 26 shows the (R) and (S) enantiomers of dipiperdone bound to the allosteric site.

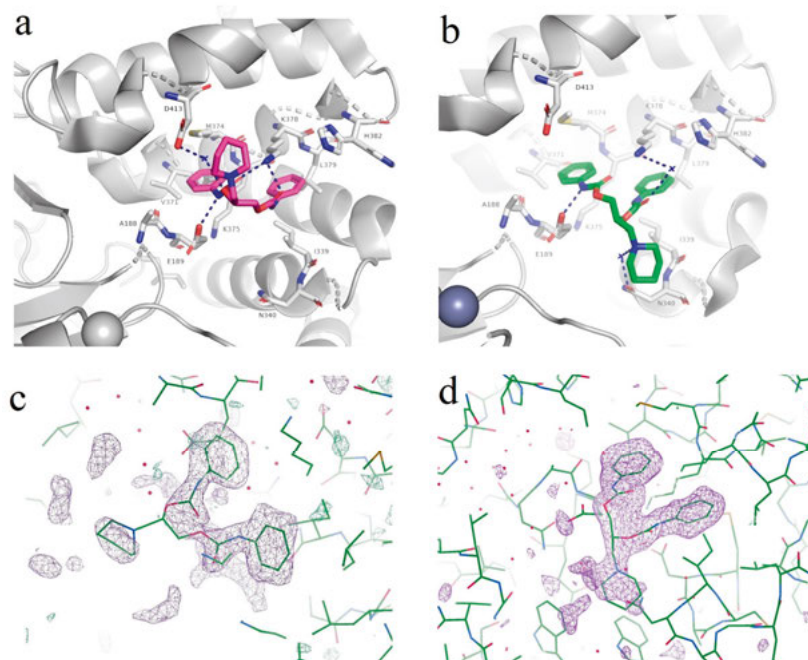


Figure 26: Illustration of SMYD3 in presence of DIP(R) (a) and DIP(S) (b) in the allosteric site. (F₀-F_c) difference density map (purple mesh) of DIP(R) (c) and DIP(S) (d).

3.2.3. A fragment based drug discovery approach

In this part of the study, a fragment based drug discovery (FBDD) approach was taken. A fragment library was screened using a grating coupled interferometry (GCI)-based biosensor to identify fragments interacting with SMYD3.

XRC experiments were done, co-crystallizing or soaking SMYD3 with the most promising fragment hits, to confirm binding and determine the location in the enzyme. Crystals diffracting to at least 1.8 Å were obtained and data was collected at a synchrotron. Using the Pipedream pipeline, weak electron density was detected for four of the fragments. This electron density was

found at several locations in the structures, including the active site and four new possible binding sites (Figures 27 and 28). Because of the weak electron density, which is indicative of weak affinities and multiple binding modes, binding in these sites could not be confirmed. However, by evolution of the fragments for higher affinities, it is possible that these ligands and allosteric sites could be exploited for design of ligands with desired properties.

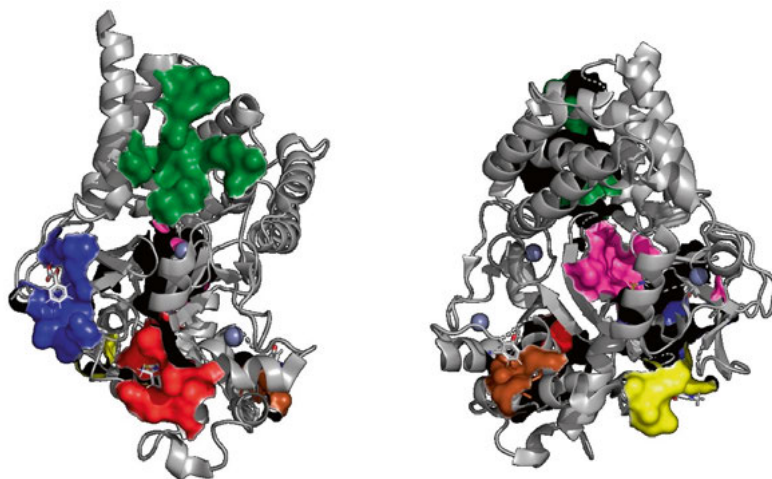


Figure 27: Illustration of the binding sites in SMYD3, including allosteric site (green surface), active site (lightmagenta surface) and four possible new binding sites, where electron density for fragments was detected (red, blue, yellow and brown surfaces).

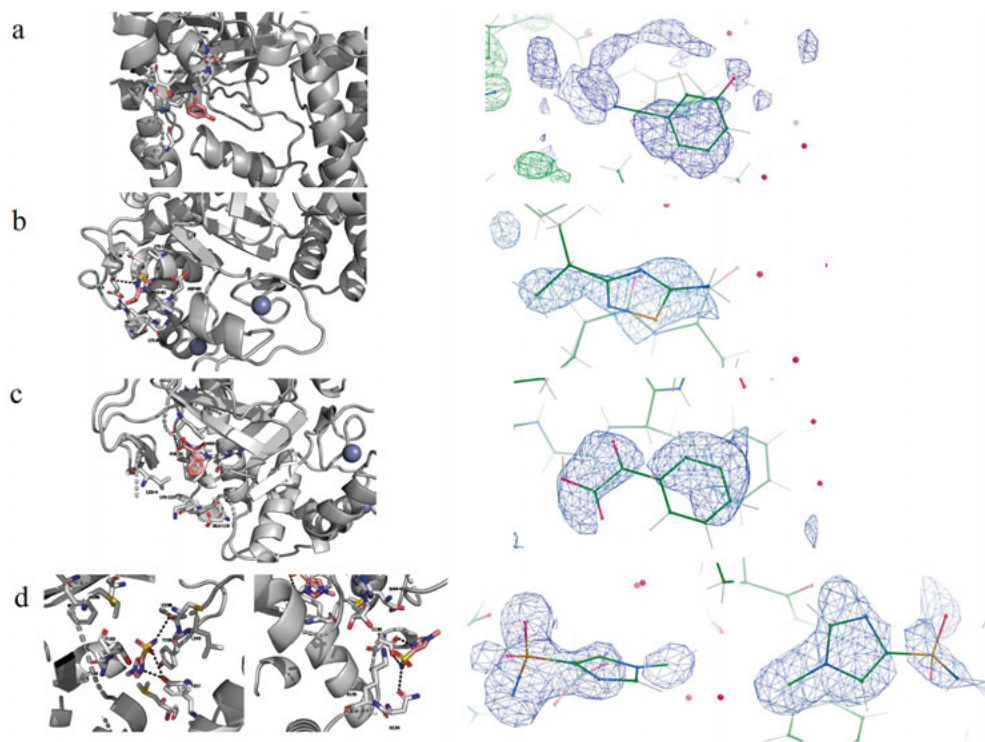


Figure 28: Binding locations and F_0 - F_c difference density map (blue mesh) for fragments (a) FL01507 (PDB-ID: 8OWO), (b) FL01791 (PDB-ID: 7QNR), (c) FL08619 (PDB-ID: 7QNU) and (d) two molecules of FL06268 (PDB-ID: 7QLB) bound to SMYD3.

3.2.4. Conclusion

The studies described in paper II, III and IV explores different approaches for discovery of possible inhibitors to the epigenetic drug target SMYD3. The first study focused on the rational design of a TCI. Advantages with TCIs include a long-lasting inhibition effect and high selectivity. Using information previously known about SMYD3 structure and inhibitors, a TCI was designed and the combination of information from SPR and XRC confirmed covalent binding and inhibition.

However, active site targeted drugs have disadvantages, such as higher risk for off-target toxicity. Therefore, the second study aimed to find an allosteric inhibitor. Allosteric inhibitors are generally more specific since the allosteric sites are often less conserved than the orthosteric sites. A hit compound was found in the SPR screen with a blocked active site, and XRC provided structures that revealed the location of the newly discovered allosteric site and the binding modes of the two enantiomer forms of the racemic compound.

In the third study, we explored FBDD for SMYD3. Several hits were found in the GCI library screen. In FBDD, structure determination of the target protein in complex with fragments is important to provide the information needed to build or combine the fragments into larger molecules that can fit the binding pocket. However, due to the weak affinities of the fragments in this study, the electron densities in the structures were weak, making identification of binding sites challenging. This problem can be solved by either identifying fragments with a slower on-off rate or using another method for the structure determination, such as NMR which provides dynamic information.

3.3. Conformational changes and drug discovery – AChBP as a model of a dynamic drug target (Paper V and VI)

The membrane spanning nAChRs are Cys-loop type LGICs which respond to the neurotransmitter acetylcholine, as well as drugs such as nicotine. nAChRs are associated with diseases of the nervous system, such as Alzheimer's disease, Parkinson's disease and schizophrenia, which makes them interesting targets for therapeutics.⁷⁹ Binding of an agonist at an orthosteric site activates the nAChR by stabilizing the open state of the receptor. In this state, ions are allowed to pass through the membrane. Competitive antagonists compete with agonists for binding, which instead results in stabilisation of the resting, closed state of the receptor. Allosteric agonists or antagonists bind to sites different from the orthosteric, resulting in either stabilizing the open state or the resting state of the receptor.^{75, 76}

Acetylcholine binding proteins (AChBPs) are homologous to the ligand-binding extracellular ligand-binding domain of the nAChR, and can be used as a model when studying the interactions and structure of nAChRs. Figure 29 shows the extracellular, intracellular and transmembrane domains of a protein structure of a Cys-loop type of LGIC in comparison with a structure of AChBP.¹²

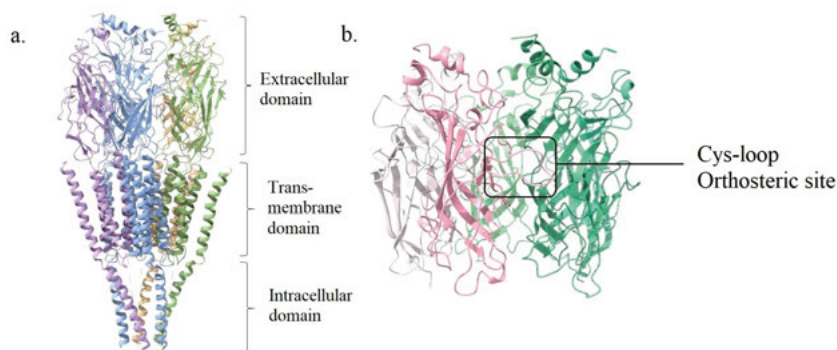


Figure 29: Structures of a Cys-loop LGIC and acetylcholine binding protein. PDB-IDs: 2BG9 and 7NDV.⁸⁰

In this study, presented in Paper V and Paper VI, a variety of biosensor techniques were combined with XRC to investigate the interactions with a set of compounds and possible conformational changes resulting from these interactions.

3.3.1. Biophysical screening for detection of interaction and conformational changes

This study explored a variety of biosensor techniques with different detection principles that can be used to monitor binding and conformational changes. They included SPR, surface acoustic wave (SAW), second harmonic generation (SHG) and GCI. In paper V, AChBP interactions with a set of fragments, hypothesised to cause conformational changes, was screened using a SHG biosensor. A subset of hits was chosen for an orthogonal analysis using a GCI biosensor, which confirmed the interactions. For two of the fragments the sensorgrams were complex and had secondary effects suggesting that conformational changes were induced upon binding of the compound.

In paper VI, three compounds that gave complex SPR sensorgrams were further investigated using SAW and SHG. Both SAW and SHG analyses resulted in signals that confirmed conformational changes. Additionally, three compounds resulting in conventional SPR sensorgrams, indicating that no significant conformational changes occur upon binding, were chosen as reference compounds for the following structural analysis.

3.3.2. Structural studies of ligand binding and conformational changes

The chosen subset of hit compounds from the various biosensor screens, in Paper V and Paper VI, were co-crystallized with AChBP. Seven of the compounds, of which three had resulted in data indicating that they caused conformational changes in the biosensor screens, could be modelled in crystal structures. Four of the compounds had resulted in conventional sensorgrams, indicating that no significant conformational changes takes place upon binding. All seven compounds bound in the ligand-binding site in the subunit interfaces, in close proximity to the Cys-loop. The ligand binding sites of the homopentamers were occupied to different extents by the various compounds. Figure 30 and 31 shown the ligands binding in the crystal structures and the F_0-F_c difference maps.

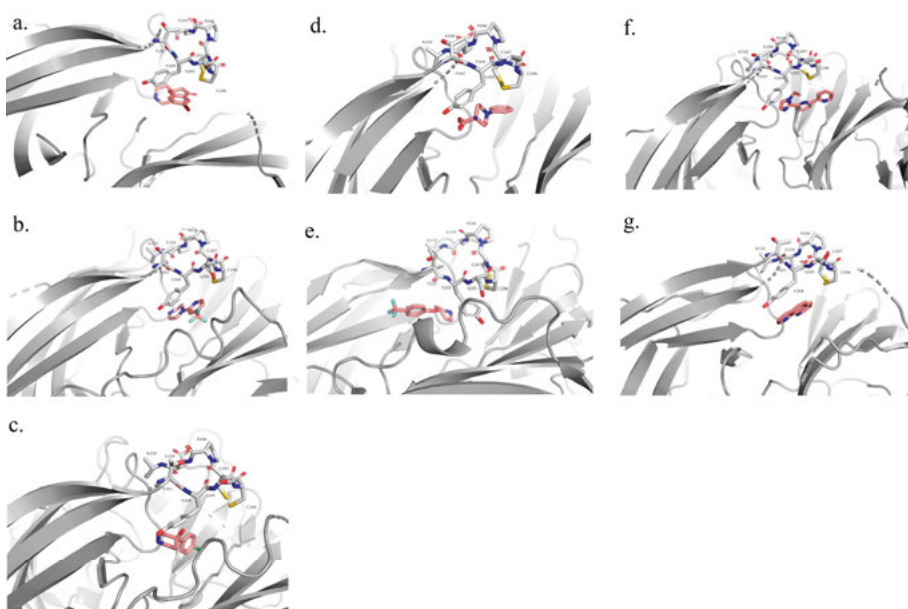


Figure 30: Compounds (a) FL001856 (PDB-ID: 7NDP), (b) FL001613 (PDB-ID: 8P1E), (c) FL003044 (PDB-ID: 8P11), (d) FL001909 (PDB-ID: 8P1F), (e) FL001888 (PDB-ID: 7NDV), (f) IOTA376 (PDB-ID: 8P22), and (g) IOTA739 (PDB-ID: 8Q1T) in the binding sites in the subunit interfaces of AChBP.

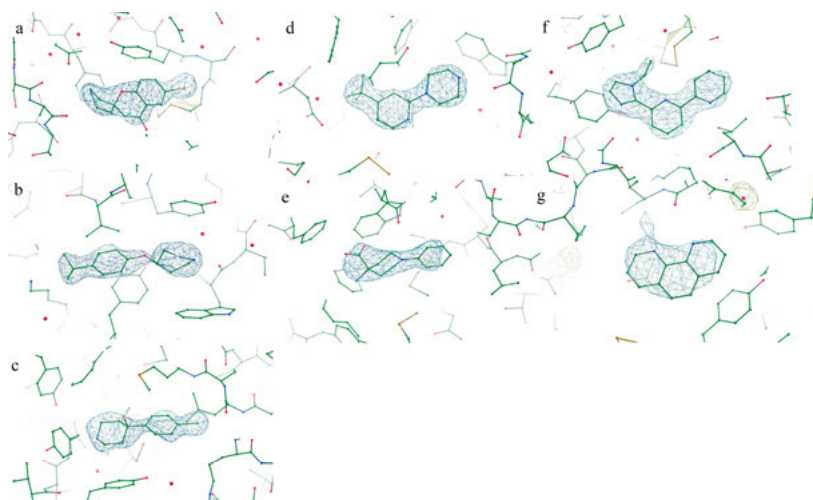


Figure 31: F_0 - F_c difference maps for compounds (a) FL001856 (PDB-ID: 7NDP), (b) FL001888 (PDB-ID: 7NDV), (c) FL003044 (PDB-ID: 8P11), (d) FL001613 (PDB-ID: 8P1E), (e) FL001909 (PDB-ID: 8P1F), (f) IOTA376 (PDB-ID: 8P22), and (g) IOTA739 (PDB-ID: 8Q1T) in the binding sites in the subunit interfaces of AChBP.

Statistical analysis of the dihedral angle over amino acid residues Y204, C206, C207 and P208 showed that local conformational changes to the Cys-loop took place following binding of FL001888, IOTA376 and IOTA739 to a larger extent than the other compounds and the reference, which coincides with the results from the biosensor screens. Figure 32 illustrates this statistical analysis, and an alignment of the ligand binding region of AChBP in presence of the compounds.

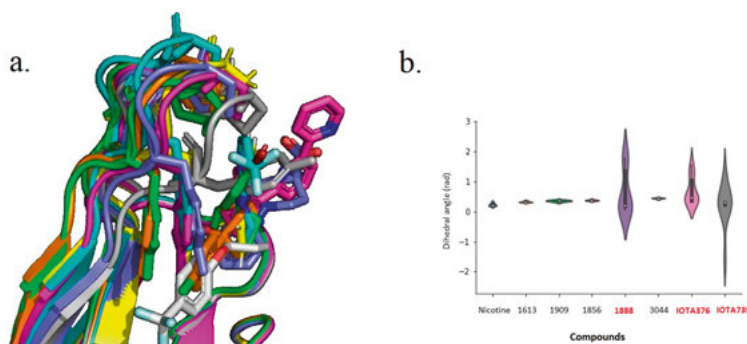


Figure 32: a. Alignment of crystal structures of AChBP with the seven compounds in close proximity to the Cys-loop. b. Statistical comparison of the dihedral angle over amino acid residues Y204, C206, C207 and P208 in close proximity to the Cys-loop and the compounds. Compounds that generated biosensor data indicating conformational changes are written in red text.

The region between amino acid residues 175 and 180 could not be modelled, which indicated that this region is flexible. Thus, it cannot be concluded whether conformational changes takes place in this region or not. For three of the compounds co-crystallized with AChBP, no electron density indicating ligand binding could be detected.

3.3.3. Conclusions

In the studies described here, we have investigated the opportunities for studying compounds binding to LGICs, including the conformational changes these ligands induce, using the model protein AChBP. A variety of biosensor techniques were explored: SHG, SPR, SAW and GCI. XRC was used to confirm ligand binding, determine binding site and investigate potential conformational changes. Comparing the crystal structures of AChBP with the different compounds, conformational changes were detected in three of the structures. These correlated to which compounds had given rise to atypical SPR sensorgrams, and indications of conformational changes in SAW and SAH. However, the conformational changes were relatively small and located only to the Cys-loop region. The flexible region between residues 175 and 180 could not be modelled due to poor electron density in this region. This could indicate that conformational changes occur at that location in the protein. It is also possible that more than one conformation of the protein occurred in the sample and that the formation of crystals selected protein species of only one conformation, resulting in crystal structures of the protein in that state. Considering this, AChBP structures with compounds present should be determined using an orthogonal method such as cryoEM. This technique could show which various conformational states are present in the sample, and could aid in modelling flexible regions. However, the higher resolution generally obtained with XRC may be required to confirm ligand binding.

4. Summary and future perspectives

Protein structures can be studied for a number of purposes. In this thesis, I have explored using XRC in drug discovery, where structural information about proteins was used to answer several scientific questions. It aided in confirming interactions, establishing binding modes and locations, and detecting ligand induced conformational changes. However, there are limitations to the technique, such as difficulties to model certain dynamic regions, and only showing a snapshot of one state of the protein when more than one state may be present in the sample. This is an indication that a complementary technique for structure elucidation may be required to answer scientific questions in similar projects.

Additionally, we have determined the first structure of the activated, larger complex of a protein where the inactive dimer state had previously been determined using XRC. The initial XRC screenings with the active, larger complex were unsuccessful, likely due to the occurrence of different complex sizes in the sample. Thus, cryoEM was employed, and the obtained model provided valuable information on the complex formation. The previously determined XRC dimer structure was used in the modelling of the new cryoEM structure, which provides an example of how the two techniques can complement each other.

The study of protein structure is complex, and every technique has its advantages and disadvantages. Care has to be taken in the choice of methodological approach for a particular purpose. For example, XRC generally results in the highest resolution and the relatively high throughput allow for applications in drug discovery, such as FBDD, where libraries of fragments can be screened in an automatic manner. However, it only captures a snapshot which needs to be considered when studying flexible proteins or proteins occurring in different oligomer states. CryoEM can provide information of several different components of a sample, for example if the sample contains the same protein in different oligomer states or even different proteins in the sample. This opens up for exciting possibilities for structure determination of samples that cannot be crystallized, for example because of sample heterogeneity. The quality and quantity of structures determined using cryoEM have drastically

increased in recent years, but the resolution is in general still lower than what can be achieved with XRC.

The drastically increased amount of structural information can allow for new possibilities for structure predictions using machine learning. This could greatly decrease the resources required in for example in drug discovery, which can allow for research projects which have not previously been prioritized. Pandemics and an ageing population will likely contribute to an increase in the demand of new drugs.

To summarize, the different techniques for determination of protein structure all have the potential to provide valuable information and it can often be useful to use more than one technique to obtain complementary structural information. The field of structure biology has had several breakthroughs in recent years. The introduction of 3rd generation synchrotrons and free electron lasers have enabled high quality XRC data collection at high throughput, enabling the use of XRC in for example SBDD and FBDD. The recent breakthrough in the quality of cryoEM data allows for determining structures of proteins that was previously not possible, as well as obtaining additional information. CryoEM is still developing and additional improvements will probably occur in both sample preparation and data processing, enabling better understanding of protein dynamics and complexes.

5. Populärvetenskaplig sammanfattning

Proteiner är den största klassen av makromolekyler, och deras funktioner och interaktioner med molekyler och andra proteiner påverkar de biokemiska processer som styr våra liv och vår hälsa. Enzymer är den största klassen av proteiner, och katalyserar alla kemiska reaktioner som sker i våra celler, genom att minska energin som krävs för att omvandla substrat till produkt.

Genom att bestämma ett proteins tredimensionella struktur kan man få information om dessa livsviktiga funktioner och interaktioner, och därmed om hur proteinerna påverkar hälsa och sjukdom. Strukturbestämning kan också ge insikter för utveckling av nya läkemedel. Proteiner är mycket små, uppskattningsvis ~10 nm långa, och därmed för små för att kunna ses med blotta ögat. Därför krävs det avancerade tekniker för att visualisera proteinernas tredimensionella struktur.

Den vanligaste tekniken för att strukturbestämma proteiner är röntgenkristallografi. Denna teknik drar nytta av den diffraktion som sker när röntgenstrålar reflekteras mot en proteinkristall. Diffraktionsmönstret översätts till elektrondensitet, och proteinstrukturen kan modelleras till elektrondensiteten. En teknik som på senare år sett stora framsteg är kryoelektronmikroskopi, där vitrifierade proteiner studeras i elektronmikroskop. Denna teknik kan användas för att strukturbestämma proteiner som av olika anledningar inte bildar proteinkristaller, men även för att erhålla kompletterande information om ett proteins struktur. Generellt ger röntgenkristallografi en högre upplöst bild på proteinstrukturen, men det kan vara utmanande att få större komplex att bilda kristaller, samt att modellera dynamiska regioner av proteinet. Metoderna kan kombineras, till exempel genom att ett större komplex strukturbestäms med kryoelektronmikroskopi, och den mer detaljerade modelleringen av subenheterna kan göras med hjälp av röntgenkristallografi.

I det första projektet jag beskriver i avhandlingen studeras ett enzym som förekommer i den reduktiva nedbrytningen av pyrimidiner, som är nukleotidbaser i DNA och RNA. Defekter hos dessa enzymer, orsakade av genetiska mutationer i DNA-sekvensen, medför flera neurologiska symptom, samt toxiska sidoeffekter av vissa typer av cancerläkemedel som bryts ned i samma process. Tidigare var endast den inaktiva dimerformen känd genom röntgenkristallografi. I denna studie fastställdes för första gången den aktiva större

komplexet av humant β UP med hjälp av kryoelektronmikroskopi. Detta visar på hur två olika tekniker för proteinstrukturbestämning kan användas för att ge komplementär information. Enzymets katalytiska aktivitet och komplexbildning efter mutation av specifika aminosyror, samt i närvaro av molekyler som påminner om enzymets substrat eller produkt, studerades för att kunna dra slutsatser om vilka mekanismer som ligger bakom enzymets katalytiska förmåga och komplexbildning. Tillsammans bidrog resultaten av dessa studier till en ökad förståelse för enzymets struktur och funktion.

En tredimensionell struktur på till exempel ett protein tillsammans med en läkemedelskandidat kan förklara hur interaktionen går till och hur molekylen kan optimeras för en starkare och mer specifik binding. Denna användning av proteinstrukturbestämning utforskades för två olika typer av målproteiner i denna avhandling.

Det första målproteinets är ett så kallat epigenetiskt enzym. Epigenetik är genetiska förändringar som sker oberoende av DNA-sekvensen, till exempel genom metylering av DNA vilket kan antingen aktivera eller de-aktivera gener. Dessa epigenetiska förändringar kan i vissa fall orsaka okontrollerbar tillväxt av tumörer, och därför är dessa epigenetiska enzymer intressanta mål för läkemedelsutveckling. Strukturen av det epigenetiska enzymet i närvaro av flera olika typer av molekyler fastställdes med hjälp av röntgenkristallografi. Denna nya information bidrog till att fastställa hur och var molekylerna binder, och att upptäcka nya bindningsställen.

I det tredje projektet studeras en homolog till en ligandstyrd jonkanal. Ligandstyrda jonkanaler är proteiner som är placerade i cellmembranet, och som efter interaktion med en ligand öppnas för att släppa igenom joner. De kan vara involverade i sjukdomar i nervsystemet. För att jonkanalen ska öppnas, behöver proteinet ändra konformation, vilket betyder att de tredimensionella positionerna av proteinets aminosyror förflyttas. Vi har studerat en homolog till den extracellulära delen av en ligandstyrd jonkanal, som ofta används som modell för dessa. Proteinets struktur bestämts med röntgenkristallografi tillsammans med en utvald grupp av molekyler, vilket medförde att vi kunde fastställa bindning samt utforska konformationsförändringar som konsekvens av interaktioner med olika typer av molekyler.

I denna avhandling presenteras forskningsprojekt som visar olika sätt att studera proteiners strukturer, samt hur strukturbestämningar kan ge information om ett proteins funktion och dess interaktioner med andra proteiner och små molekyler. Detta kan användas till exempel i syftet att bättre förstå regleringen hos proteiner som påverkar vår hälsa, eller i syfte att hitta kandidater till nya läkemedel.

6. Acknowledgements

The work presented in this thesis would not have been possible without the help of many people, and I would like to express my gratitude to all of you.

I would like to thank my supervisors, **Helena** and **Doreen**, for trusting me with your projects, supporting me and being good role models. This has been the most important and educational experience of my scientific career and would not have been possible without you. Thank you for sharing your knowledge and trusting me to work independently.

My colleagues in the Danielson and Dobritzsch groups: **Gun, Sandra, Dilip, Eldar, Helena N, Khyati, Emil, Mostafa, Guilia, Ted, Nadine, Guillermo, Marija and Daria**. Thank you, it has been a pleasure to work with all of you!

Dirk and **Vladimir**, thank you for introducing me to the lab and to my projects, and for sharing your experience with me.

Thank you to my colleagues at the Department of Chemistry, for contributing to a friendly atmosphere and a work place I am happy to go to. **Johanna, Caroline, Susanne, Max, Emily, Ylva, Leandro, Ali, Micke, Gina, Candice, Gunnar, Erik, Joana, Francoise, Sebastian, Leonardo, Jennifer and Lidija**. It has been great working together.

I would also like to thank my colleagues at ICM, specially **Annette, Terese** and **Daniel**, for being so helpful and sharing your expertise with me.

I would like to thank the project students I had the pleasure to supervise. **Ayan, He, Juliane, Adam** and **Gilles**, thank you for your hard work and contributions to my projects.

I also have people outside of work who have been important for this journey. I can only mention a few of you here. My parents **Maria** and **Daniel**, my siblings **Sebastian** and **Emilia**, my husband **Ludvig**, and last but not least **Sunniva**.

7. References

1. Nelson, D. L. and Cox, M. M. 2017. *Lehninger Principles of Biochemistry*. 7th ed. New York: W.H.Freeman and Company.
2. Erickson, H. P. (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online*. **11**, 32-51.
3. Drenth, J., 2007. *Principles of Protein X-ray Crystallography*. 3rd ed. New York: Springer.
4. Bragg, W. L. (1913) The structure of some crystals as indicated by their diffraction of X-rays. *Proc. R. Soc. Lond.* **89(610)**, 248–277.
5. Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, **181**, 662-666.
6. Powell, H. R., X-ray data processing (2017) *Biosci Rep.* **37(5)**, Article BSR20170227. Erratum in: *Biosci Rep.* **40(7)**.
7. Helliwell, J.R. New developments in crystallography: exploring its technology, methods and scope in the molecular biosciences. (2017) *Biosci Rep.* **37(4)**, Article BSR20170204.
8. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer Jr, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.* **112(3)**, 535-42.
9. Serna, M. (2019) Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes, *Front. Mol. Biosci.* **6**, 1-8.
10. Copeland, R., 2000. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. 2nd ed. New York: Wiley-VCH.
11. Seetharaman, R., Advani, M. G., Mali, S. and Pawar, S. (2020). Enzymes as targets of Drug Action: an Overview. *IJMSCR.* **3(3)**, 114-125.
12. Smit, A.B., Brejc, K., Syed, N. and Sixma, T.K. (2003) Structure and function of AChBP, homologue of the ligand-binding domain of the nicotinic acetylcholine receptor. *Ann N Y Acad Sci.* **998**, 81-92.
13. Wu, Z.S., Cheng, H., Jiang, Y., Melcher, K. and Xu, H.E. (2015) Ion channels gated by acetylcholine and serotonin: structures, biology, and drug discovery. *Acta Pharmacol Sin.* **36(8)**, 895-907.
14. Rao, R., Shah, S., Bhattacharya, D., Toukam, D.K., Cáceres, R., Pomeranz Krummel, D.A. and Sengupta, S. (2022) Ligand-Gated Ion Channels as Targets for Treatment and Management of Cancers. *Front Physiol.* **13**, Article 839437.
15. Koshland, D.E. (1995) The Key–Lock Theory and the Induced Fit Theory, *Angew Chemie Int Ed English*, **33(23-24)**, 2375-2378.

16. Benkovic, S.J., Hammes, G.G. and Hammes-Schiffer, S. (2008). Free-Energy Landscape of Enzyme Catalysis. *Biochemistry*. **47**(11), 3317–3321.
17. Alberty, R.A. (2004) Thermodynamic properties of oxidoreductase, transferase, hydrolase, and ligase reactions. *Arch Biochem Biophys*. **435**(2), 363–368.
18. McDonald, A.G. and Tipton, K.F. (2023) Enzyme nomenclature and classification: the state of the art. *FEBS J*. **290**(9), 2214–2231.
19. Rufer, A.C. (2021) Drug discovery for enzymes. *Drug Discov Today*. **26**(4), 875–886.
20. Brown, C.S. and Lichter-Konecki, U. Phenylketonuria (PKU): A problem solved? *Molecular Mol Genet Metab*. **6**, 8–12.
21. Jurecka, A. (2009) Inborn errors of purine and pyrimidine metabolism. *J Inherit Metab Dis*. **32**, 247–263.
22. Chapman, J., Ismail, A.E. and Dinu, C.Z. (2018) Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. *Catalysts*. **8**, Article 238.
23. Michaelis, L. and Menten, M.L. (1913) Die Kinetik der Invertinwirkung. *Biochem Z*. **49**, 333–369.
24. Dobritzsch, D., Meijer, J., Meinsma, R., Maurer, D., Monavari, A.A., Gummesson, A., Reims, A., Cayuela J.A., Kuklina, N., Benoist, J.F., Perrin, L., Assmann, B., Hoffmann, G.F., Bierau, J., Kaindl, A.M. and van Kuilenburg, A.B.P. β -Ureidopropionase deficiency due to novel and rare UPB1 mutations affecting pre-mRNA splicing and protein structural integrity and catalytic activity. *Mol Genet Metab*. **136**(3), 177–185.
25. Maurer, D., Lohkamp, B., Krumpel, M., Widersten, M. and Dobritzsch, D. (2018) Crystal structure and pH-dependent allosteric regulation of human β -ureidopropionase, an enzyme involved in anticancer drug metabolism. *Biochem J*. **475**(14), 2395–2416.
26. van Kuilenburg, A.B.P., Meinsma, R., Beke, E., Assmann, B., Ribes, A., Lorente, I. Busch, R., Mayatepek, E., Abeling, N.G., van Cruchten, A., Stroomer, A.E., van Lenthe, H., Zoetekouw, L., Kulik, W., Hoffmann, G.F., Voit, T., Wevers, R.A., Rutsch, F. and van Gennip, A. (2004) β -Ureidopropionase deficiency: an inborn error of pyrimidine degradation associated with neurological abnormalities. *Hum. Mol. Genet*. **13**, 2793–2801.
27. van Kuilenburg, A.B.P., Dobritzsch, D., Meijer, J., Krumpel, M., Selim, L.A., Rashed, M.S., Assmann, B., Meinsma, R., Lohkamp, B., Ito, T., Abeling, N.G.G.M., Saito, K., Eto, K., Smitka, M., Engvall, M., Zhang, C., Xu, W., Zoetekouw, L. and Hennekam, R.C.M. (2012) β -ureidopropionase deficiency: phenotype, genotype and protein structural consequences in 16 patients. *Biochim. Biophys. Acta, Mol. Basis Dis*. **1822**, 1096–1108.
28. Han, M., Jia, L., Lv, W., Wang, L. and Cui, W. (2019) Epigenetic Enzyme Mutations: Role in Tumorigenesis and Molecular Inhibitors. *Front Oncol*. **9**, Article 194.
29. Fabini, E., Talibov, V.O., Mihalic, F., Naldi, M., Bartolini, M., Bertucci, C., Del Rio, A. and Danielson, U.H. (2019) Unveiling the Biochemistry of the Epigenetic Regulator SMYD3. *Biochemistry*. **58**(35), 3634–3645.
30. Bernard, B.J., Nigam, N., Burkitt, K. and Saloura, V. (2021) SMYD3: a regulator of epigenetic and signaling pathways in cancer. *Clin Epigenet*. **13**(1), Article 45.

31. Kubota, T., Miyake, K. and Hirasawa, T. (2012) Epigenetic understanding of gene-environment interactions in psychiatric disorders: a new concept of clinical genetics. *Clin Epigenet.* **4**, Article 1.
32. Ferraccioli, G., Carbonella, A., Gremese, E. and Alivernini, S. (2012) Rheumatoid arthritis and Alzheimer's disease: genetic and epigenetic links in inflammatory regulation. *Discov Med.* **14(79)**, 379-88.
33. Egger, G., Liang, G., Aparicio, A. and Jones, P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature.* **429**, 457-63.
34. Pelton, J.T. and McLean, L.R. (2000) Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal Biochem.* **277(2)**, 167-176.
35. Eisenhaber, F., Persson, B. and Argos, P. (1995) Protein Structure Prediction: Recognition of Primary, Secondary, and Tertiary Structural Features from Amino Acid Sequence. *Crit Rev Biochem Mol Biol.* **30(1)**, 1-94.
36. Bajaj, M. and Blundell, T. (1984) Evolution and the tertiary structure of proteins. *Ann. Rev. Biophys.* **13**, 453-92.
37. Dyson, J.H., Wright, P.E. and Scheraga, H.A. (2006) The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci USA.* **103 (35)**, 13057-13061.
38. Wedemeyer, W. J., Welker, E., Narayan, M. and Scheraga, H. A. (2000) Disulfide Bonds and Protein Folding. *Biochemistry.* **39(15)**, 4207-4216.
39. Janin, J., Bahadur, R. and Chakrabarti, P. (2008). Protein-protein interaction and quaternary structure. *Q Rev Biophys.* **41(2)**, 133-180.
40. Egelman, E. H. (2016). The current revolution in cryo-EM. *Biophys J.* **110(5)**, 1008-1012.
41. Perutz, M. F. (1963). X-ray Analysis of Hemoglobin: The results suggest that a marked structural change accompanies the reaction of hemoglobin with oxygen. *Science*, **140(3569)**, 863-869.
42. Bai, X., McMullan, G. and Scheres, S.H.W (2015). How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci.* **40(1)**, 49-57.
43. Gauto, D.F., Estrozi, L.F., Schwieters, C.D., Effantin, G., Macek, P., Sounier, R., Sivertsen, A.C., Schmidt, E., Kerfah, R., Mas, G., Colletier, J.P., Güntert, P., Favier, A., Schoehn, G., Schanda, P. and Boisbouvier, J. (2019) Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat Commun*, **10(1)**, Article 2697.
44. Kempf, J.G. and Loria, J.P. (2002) Protein dynamics from solution NMR. *Cell Biochem Biophys*, **37**, 187-211.
45. Geraets, J.A., Pothula, K.R. and Schröder, G.F. (2020) Integrating cryo-EM and NMR data. *Curr Opin Struct Biol.* **61**, 173-181.
46. Russo Krauss, I., Merlino, A., Vergara, A. and Sica, F. (2013) An Overview of Biological Macromolecule Crystallization. *Int J Mol Sci.* **14(6)**, 11643-11691.
47. McPherson, A. 2009. *Introduction to Macromolecular Crystallography*. Second edition. New Jersey: John Wiley & Sons.
48. Biel, S.S. and Gelderblom, H.R. (1999) Diagnostic electron microscopy is still a timely and rewarding method. *J Clin Virol.* **13(1-2)**, 105-19.
49. Adrian, M., Dubochet, J., Lepault, J. and McDonwall, A.W. (1984) Cryo-electron microscopy of viruses. *Nature.* **308**, 32-36.

50. Xu, Y. and Dang, S. (2022) Recent Technical Advances in Sample Preparation for Single-Particle Cryo-EM. *Front Mol Biosci.* **9**, Article 892459.
51. Bai, X.C. (2021) Seeing Atoms by Single-Particle Cryo-EM. *Trends Biochem Sci.* **46(4)**, 253-254.
52. Peplow, M. Cryo-Electron Microscopy Reaches Resolution Milestone. (2020) *ACS Cent Sci.* **6(8)**, 1274-1277.
53. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehl, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E.V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S.W., Chirgadze, D.Y., Murshudov, G., Aricescu, A.R. and Scheres, S.H.W. (2020) Single-particle cryo-EM at atomic resolution. *Nature.* **587(7832)**, 152-156.
54. Yip, K.M., Fischer, N., Paknia, E., Chari, A. and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature.* **587**, 157–161.
55. Kato, T., Makino, F., Nakane, T., Terahara, N., Kaneko, T., Shimizu, Y., Motoki, S., Ishikawa, I., Yonekura, K. and Namba, K. (2019) CryoTEM with a Cold Field Emission Gun That Moves Structural Biology into a New Stage. *Microscopy and Microanalysis.* **25(S2)**, 998–999.
56. Method of the Year 2015. (2016) *Nat Methods.* **13**, 1.
57. Cressey, D. and Callaway, E. (2017) Cryo-electron microscopy wins chemistry Nobel. *Nature.* **550**, Article 167.
58. Kim, H. and Jung, H.S. (2021) Cryo-EM as a powerful tool for drug discovery: recent structural based studies of SARS-CoV-2. *Appl. Microsc.* **51**, Article 13.
59. Tüting, C., Kyriulis, F.L., Müller, J. et al. (2021) Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nat Commun.* **12**, Article 6933.
60. Zhong, E.D., Bepler, T., Berger, B. and Davis, J.H. (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods.* **18**, 176–185.
61. Weissenberger, G., Henderikx, R.J.M. and Peters, P.J. (2021) Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods.* **18**, 463–471.
62. Dillard, R.S., Hampton, C.M., Strauss, J.D., Ke, Z., Altomara, D., Guerrero-Ferreira, R.C., Kiss, G. and Wright, E.R. (2018). Biological Applications at the Cutting Edge of Cryo-Electron Microscopy. *Microscopy and Microanalysis*, **24(4)**, 406–419.
63. Franken, L.E., Grünewald, K., Boekema, E.J., Stuart, M.C.A. (2020) A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small.* **16(14)**, Article 1906198.
64. Dhakal, A., Gyawali, R., Wang, L. and Cheng, J. (2023) A large expert-curated cryo-EM image dataset for machine learning protein particle picking. *Sci Data.* **10(1)**, Article 392.
65. Sigworth, F.J. (2016) Principles of cryo-EM single-particle image processing. *Microscopy (Oxf).* **65(1)**, 57-67.
66. Atanasov, A.G., Waltenberger, B., Pferschy-Wenzig, E.M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E.H., Rollinger, J.M., Schuster, D., Breuss, J.M., Bochkov, V., Mihovilovic, M.D.,

- Kopp, B., Bauer, R., Dirsch, V.M. and Stuppner, H. (2015) Discovery and re-supply of pharmacologically active plant-derived natural products: A review. *Biotechnol Adv.* **33**(8), 1582-1614.
67. Weyrich, L.S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A.G., Alt, K.W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A.G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E.C., Kaidonis, J., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Semal, P., Soltysiak, A., Townsend, G., Usai, D., Wahl, J., Huson, D.H., Dobney, K. and Cooper, A. (2017) Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*. **544**, 357–361.
 68. Jones, A.W. (2011) Early drug discovery and the rise of pharmaceutical chemistry. *Drug Test Anal.* **3**(6), 337-44.
 69. Hughes, J.P., Rees, S., Kalindjian, S.B. and Philpott, K.L. (2011) Principles of early drug discovery. *Br J Pharmacol.* **162**(6), 1239-49.
 70. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I. and Overington, J.P. (2017) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* **16**(1), 19-34.
 71. Bon, M., Bilsland, A., Bower, J. and McAulay, K. (2022) Fragment-based drug discovery-the importance of high-quality molecule libraries. *Mol Oncol.* **16**(21), 3761-3777.
 72. Li, Q. (2020) Application of Fragment-Based Drug Discovery to Versatile Targets. *Front Mol Biosci.* **5**(7), Article 180.
 73. Maia, E.H.B., Assis, L.C., de Oliveira, T.A., da Silva, A.M. and Taranto, A.G. (2020) Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front Chem.* **8**, Article 343.
 74. Batool, M., Ahmad, B. and Choi, S. (2019) A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **20**(11), Article 2783.
 75. Nys, M., Kesters, D. and Ulens C. (2013) Structural insights into Cys-loop receptor function and ligand recognition. *Biochem Pharmacol.* **86**(8), 1042-53.
 76. Stober, S.T. and Abrams, C.F. (2012) Enhanced meta-analysis of acetylcholine binding protein structures reveals conformational signatures of agonism in nicotinic receptors. *Protein Sci.* **21**(3), 307-17.
 77. Lundgren, S., Lohkamp, B., Andersen, B., Piškur, J. and Dobritsch, D. (2008) The crystal structure of β -alanine synthase from *Drosophila melanogaster* reveals a homooctameric helical turn-like assembly. *J. Mol. Biol.* **377**, 1544–1559.
 78. Baillie, T.A. (2016) Targeted Covalent Inhibitors for Drug Design. *Angew Chem Int Ed Engl.* **55**(43), 13408-13421.
 79. Xiao, Y., Hammond, P.S., Mazurov, A.A. and Yohannes, D. (2012) Multiple interaction regions in the orthosteric ligand binding domain of the $\alpha 7$ neuronal nicotinic acetylcholine receptor. *J Chem Inf Model.* **52**(11), 3064-73.
 80. Unwin, N. (2005) Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. *J Mol Biol.* **346**(4), 967-89.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2292

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)



Distribution: publications.uu.se
urn:nbn:se:uu:diva-508764

ACTA UNIVERSITATIS
UPSALIENSIS
2023