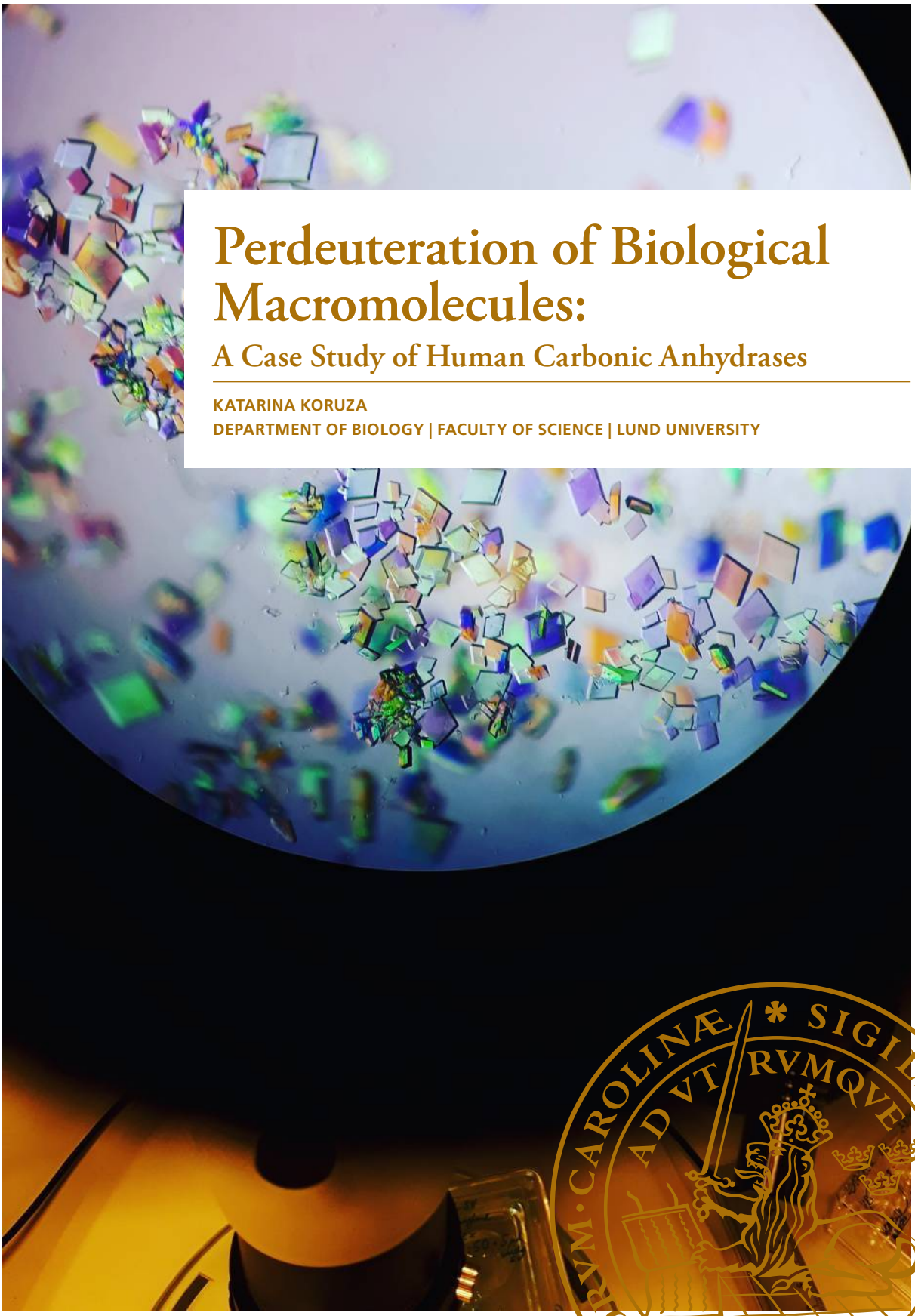
Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Koruza, Katarina

2019

Document Version:
Publisher's PDF, also known as Version of record



Perdeuteration of Biological Macromolecules:

A Case Study of Human Carbonic Anhydrases

KATARINA KORUZA
DEPARTMENT OF BIOLOGY | FACULTY OF SCIENCE | LUND UNIVERSITY

Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Katarina Koruza

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at the Biology Lecture Hall, Sölvegatan 35, Lund on the 24th of May 2019, at 13:00.

Faculty opponent
Tomas Lundqvist, RISE Research Institutes of Sweden

Organization LUND UNIVERSITY	**Document name** Doctoral dissertation in Biology
	Date of issue: 2019-05-24
Author(s) Katarina Koruza	Sponsoring organization

Title and subtitle:
Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Abstract

Deuterated biomolecules - where hydrogen atoms (H) are exchanged to its isotope deuterium (D) - are essential for biological experiments in neutron scattering, but they are not readily available. In neutron scattering, D has a strong positive coherent scattering length (6.67 fm) in comparison with the negative scattering length of H (-3.74 fm). These different characteristics of neutrons to D and H are used in neutron protein crystallography (NPX) where the use of deuterated proteins maximizes signal-to-noise ratio, reduces background and gives better nuclear scattering length density maps. At the same time, NPX allows at resolutions of ~2.5 Å or better to unambiguously determine the positions of the exchanged D. This is important because H atoms constitute nearly half of all the atoms in a protein molecule. Accurate structural information about H positions and hydrogen bonds that are involved in most aspects of protein structure and function remains challenging by using X-ray crystallography alone. Therefore, NPX offers a unique and complementary approach to X-rays for locating H atoms.

My thesis focuses on production of recombinant proteins for NPX with human carbonic anhydrase IX (hCA IX) as the case study protein. The long-term ambition of my work is to show that joint X-ray crystallography and NPX can provide insight into protein–ligand interactions and can give valuable input into structure-guided drug design and computational chemistry. As an example, for this, an emerging cancer target hCA IX was chosen, that is in need of the development of isoform-specific inhibitors.

In paper I, we demonstrate the improvement of binding affinity of the inhibitor acetazolamide to hCA IX when conjugated to a polypeptide. In paper IV, we discuss crystallisation optimisation for large crystal growth, that is often a bottleneck for NPX experiments. Finally, in paper IV we describe jointly refined X-ray and neutron crystal structures of hCA IX_{mimic} (hCA II with mutations that mimic the active site of hCA IX) with two selective inhibitors. This study shows how neutron studies can contribute unique information on inhibitor binding.

I was particularly interested in improvement of (per)deuteration in living organisms in order to maximize recombinant protein production and expression of hCA IX under deuterium stress, which are the main discussion points of paper II and III. In paper II, we address some of the technical difficulties of protein deuteration *in vivo* in bacteria, and study the biophysical effects of deuteration on the protein product. Further on we explore the possibilities improvement of deuteration of proteins by supplementing a minimal growth media with a fully perdeuterated extract from algae, as described in paper III.

Key words
Neutron protein crystallography, deuteration, protein production, cancer, inhibitor, human carbonic anhydrase IX

Classification system and/or index terms (if any)

Supplementary bibliographical information	**Language** English
ISSN and key title	**ISBN** (print) 978-91-7895-068-3 (pdf) 978-91-7895-069-0

Recipient's notes	**Number of pages** 74	Price
	Security classification	

Signature Katarina Koruza Date 2019-04-09

Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Katarina Koruza

LUND
UNIVERSITY

Cover photo by Katarina Koruza

Faculty of Science
Department of Biology

ISBN
(print) 978-91-7895-068-3
(pdf) 978-91-7895-069-0

Printed in Sweden by Media-Tryck, Lund University
Lund 2019

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To my family

Table of Contents

List of papers

The thesis is based on the following papers,

I. Yang J, **Koruza K**, Fisher SZ, Knecht W, Baltzer L. (2017) Improved molecular recognition of Carbonic Anhydrase IX by polypeptide conjugation to acetazolamide. Bioorganic and Medicinal Chemistry, 25, 5838-5848.

II. **Koruza K**, Lafumat B, Végvári Á, Knecht W, Fisher SZ. (2018) Deuteration of human carbonic anhydrase for neutron crystallography: Cell culture media, protein thermostability, and crystallization behavior. Archives of Biochemistry and Biophysics, 645, 26-33.

III. **Koruza K**, Adachi M, Akutsu K, Fisher SZ, Fujiwara S, Knecht W, Á. Végvári, von Wachenfeldt C. (2019) Development of new methodologies to express deuterated proteins in algal-derived media. (*Manuscript*).

IV. **Koruza K**, Lafumat B, Nyblom M, Knecht W, Fisher SZ. (2018) From initial hit to crystal optimization with microseeding of human carbonic anhydrase IX—A case study for neutron protein crystallography. Crystals, 8, 434.

V. **Koruza K**, Mahon BP, Blakeley MP, Ostermann A, Schrader TE, McKenna R, Knecht W, Fisher SZ. (2019) Using neutron crystallography to elucidate the basis of selective inhibition of carbonic anhydrase by saccharin and a derivative. Journal of Structural Biology, 205, 147-154.

List of contributions:

My contributions to the here presented studies were as follows:

I. I prepared proteins used in the study and wrote the relevant parts of the manuscript.

II. I participated in study conception and design. I designed and performed the experiments and analyzed the data. I contributed to interpretation of the data and to manuscript writing.

III. I led the study conception and design. I designed and performed all experiments, analyzed the data and wrote the manuscript.

IV. I participated in study conception and design. I designed and performed the experiments and analyzed the data. I contributed to interpretation of the data. I contributed to manuscript writing and I am corresponding author.

V. I participated in study conception and design. I planned and designed experiments, I performed the experiments. I participated in structure refinement and analysis. I prepared the draft of the paper.

Introduction

The work presented in this thesis focuses on production of recombinant proteins for neutron crystallography with human carbonic anhydrase IX (hCA IX) as the model protein used in my work. In the first section, the background of the study and methods are presented. In the second section, the outcomes of the individual projects are discussed, followed by conclusions and future perspectives.

The long-term ambition of the work presented here is to show that joint X-ray and neutron crystallography can provide insight into protein–ligand interactions that goes beyond what can be achieved by X-ray crystallography alone, and can give valuable input into structure-guided drug design and computational chemistry. As an example for this, hCA IX was chosen, an emerging cancer target in need of development of isoform-specific inhibitors.

The outline of the thesis follows the workflow necessary for many protein structural biology projects: from first protein production to optimization of protein expression including protein (per)deuteration, to growing large crystals for neutron diffraction experiments, and finally preparing complexes of hCA IX with promising new enzyme inhibitors and exploring their co-structures. The papers presented in this thesis are ordered in accordance with this workflow.

Deuteration is crucial and a big challenge for neutron experiments. After establishing the protein production of hCA IX (presented for the first time in paper I), I first address method developments for deuteration of proteins (paper II and III). Having neutron diffraction in mind, another requirement is to be able to grow large single crystals – therefore strategies for large crystal growth were explored in paper IV. And finally, in paper V the final outcome of joint x-ray and neutron structures of hCA IX alone and in complex with two inhibitors is shown. The importance of structural studies of hCA IX and the possible use of valuable structural information in cancer research is discussed. The data helps us to determine the specific interactions between the compounds and the enzyme that are responsible for binding, and gives input into which groups we could modify for improved binding.

Section I: Background

Biological macromolecules

Organic molecules and water are the main constituents of the cells. Water as a key indicator for life represents more than 70% of the total mass of the cell. It is not just a passive liquid solvent but it also interacts in the formation of the structure of proteins, nucleic acids and their complexes. It contributes to the stability and flexibility required for proper function of the living cell (Privalov and Crane-Robinson, 2017). Water has polar properties caused by a slightly positive charge of hydrogen atom and the slightly negative charge of the oxygen atom. Polarity enables formation of intermolecular or intramolecular hydrogen bonds with polar molecules and interaction with positively or negatively charged ions. This results in solubility of ions and polar molecules in water (hydrophilic). On the other side, nonpolar molecules that cannot interact with water are poorly soluble in an aqueous environment (hydrophobic). Biological macromolecules belong to four classes: proteins, carbohydrates, lipids, and nucleic acids. Proteins, nucleic acids and most of carbohydrates (the polysaccharides) are macromolecules or polymers formed by joining of large number of organic molecules or monomers: amino acids, nucleotides, and sugars, respectively. The rest of the cell mass is composed of lipids and a variety of small organic molecules (Cooper, 2000).

Biological macromolecule is therefore a broad term used to describe molecules in living organisms with essential functions for biological processes. Probably the most diverse of all the biological macromolecules are proteins. They are the foundation of a cell's structure and function. One of the most prominent properties of proteins is that most of them act as enzymes and catalyse nearly all the chemical reactions in biological systems.

Proteins are polymers of 20 unique amino acids, that have evolved over billions of years of evolution. Every amino acid consists of a carbon atom (α carbon) bonded to a carboxyl group (COO^-), an amino group (NH_3^+), a hydrogen atom, and a distinctive side chain that determines the role of each amino acid in protein structure and function. Proteins are composed of amino acids that are joined together by peptide bonds between the α amino group of one amino acid and the α carboxyl group of a second (Cooper, 2000). The amino acid sequence and specific chemical properties of the amino acid side chains determine the three-dimensional structure

or conformation of the proteins and its function, shown in (Figure 1). The current techniques for determining the atomic and molecular structures of proteins and biological macromolecules are discussed later in this thesis.

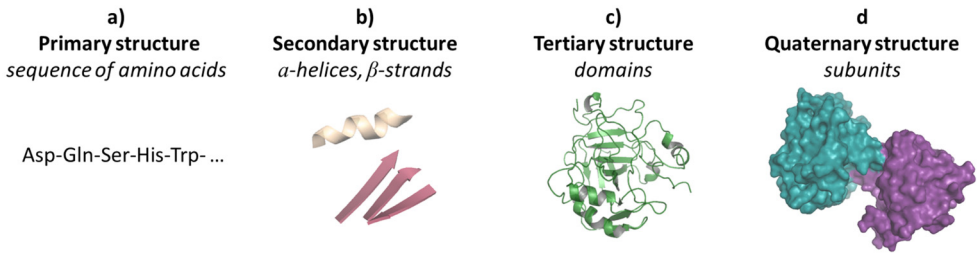


Figure 1: The three-dimensional structure of a protein can be described by four levels of structure. (a) Primary structure is defined by the linear sequence of amino acids residues. (b) Secondary structure is built up by parts of regularly repeating conformations of a peptide chain, such as α-helices and β-strands, stabilized by hydrogen bonds between different amino acids (c) Tertiary structure is the shape of a folded polypeptide chain, stabilized by the interaction of amino acid side chains in non-neighbouring regions of the polypeptide chain. (d) Quaternary structure is the arrangement of polypeptide chains that interact with each other, forming for example a dimer. Shown here is an example of a dimer formed by the two catalytic domains of hCA IX. All figures were made using *Pymol* (Schrödinger, 2015).

The three-dimensional shape enables proteins to be functional as enzymes or selective biological catalysts that increase the rate of chemical reactions within the cells. There is a large variety of reactions catalysed by enzymes, each with a special reaction mechanism that often involves the breaking and/or creating of a covalent bond. The catalytic mechanism of an enzyme is directly related to the molecular interaction between a substrate molecule and the active site of the enzyme. The exact nature and timing of these interactions define the mechanism and kinetics of catalysis (Cooper, 2000).

Structure determination

Starting from the primary structure of a protein, we can often predict the structural elements of the secondary structures such as membrane-spanning α helices or β-strands. The three-dimensional folded structure can be predicted from the amino acid sequence to some extent but with less certainty, unless the structure of a very similar protein is already known. There are several possibilities for determining the structure of molecules, including proteins at atomic resolution.

The structure determination methods mentioned in this section have advantages and disadvantages, but the basic principle is the same: the derived structure is based on pieces of complex information against which an atomic model is proposed or modelled. Using complementary information from more than one technique can help to build a more accurate model that is consistent with both the experimental

data and the expected composition and geometry of the molecule (Berman et al., 2000).

X-ray diffraction

Protein crystallography began in the early 1930s with pioneering studies of pepsin crystals with X-rays, however the structure of pepsin was not determined for several more decades. The first protein crystal structures were determined in 1960s and include myoglobin and hemoglobin (Kendrew et al., 1958; Perutz et al., 1960). Interest in protein structure and function continued to grow and structural biologists understood the importance of archiving primary data. The Protein Data Bank (PDB) was established in 1971 with just seven protein structures. Since then, the PDB archive is ensuring that experimental data and metadata are expertly curated, validated, and made freely available (Berman and Burley, 2016). The number of structures in PDB has grown to almost 150 000 to date and X-ray diffraction is the dominant method for solving protein structure. According to the PDB archive around 90% of the 150000deposited structures (https://www.rcsb.org/pdb; accessed 2019-03-01) were determined by X-ray crystallography (Figure 2).

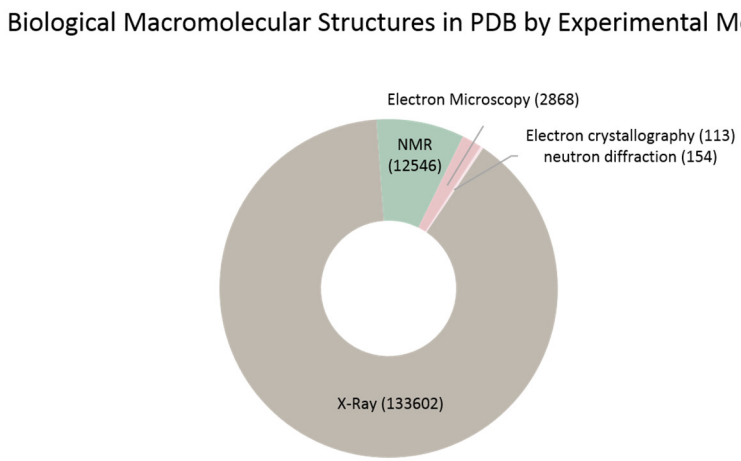


Figure 2: Number of biological macromolecular structures deposited in PDB by experimental method. Adapted from PDB; http://www.rcsb.org/pdb/, accessed on 2019-03-01.

X-ray diffraction has been historically the most important technique for advancement of our understanding of protein structure and function. X-ray crystallography exploits the fact that molecules organised in a crystal diffract the X-rays by generating a distinct diffraction pattern recorded by detectors. By measuring the angles and intensities of diffracted beams, it is possible to produce the three-

dimensional model of electron density in the crystal. Prerequisite for studies of proteins with X-rays is to obtain well-ordered crystals of biological molecules, which is demanding and not always possible (Blundell, 2017). Furthermore, exposing the crystal to the X-ray irradiation produces global and radiation specific damages that can lead to artefacts in the final derived model. X-rays belong to the class of ionising radiation that can generate free radicals that damage biological molecule through a variety of mechanisms.

The magnitude of X-ray scattering increases with the number of electrons, therefore hydrogen (H) atoms and protons (H^+) are intrinsically invisible with this technique. Exceptions are ultra-high resolutions structures (1 Å or better) where limited number of H atom positions in well-ordered regions can be experimentally determined. Dynamic H atoms, such as those in the active sites of an enzyme usually remain invisible even at ultra-high resolutions (Langan and Chen, 2013). Taking a protein as an example of a biological macromolecule, H atoms account for approximately half of the atoms and play a crucial role in protein folding, H bonding, enzyme catalysis, and ligand (e.g., drug) binding interactions.

Beside X-rays, there are currently several stand-alone or complementary experimental methods for determining protein structures at the atomic level, like neutron protein crystallography (NPX), nuclear magnetic resonance spectroscopy (NMR), electron diffraction, and electron microscopy. These other techniques all together represent the remaining 10% of the structures in PDB database.

Neutron diffraction

In 1935, James Chadwick was awarded The Nobel prize in physics for the discovery of the neutrons (Chadwick, 1932). Since then, neutrons produced in nuclear reactors and spallation sources with wavelengths on the order of Ångstroms, offer an attractive probe for the study of biological samples due to their particular properties. Neutrons are neutral subatomic particles with greater mass than electrons. With no charge, neutrons interact weakly with matter and can penetrate deep into materials in a non-destructive way. The interactions between the neutron and atomic nuclei involve complex nuclear interactions between nuclear spin and magnetic moment. In contrast to X-rays, neutrons interact only with nuclei and not with the cloud of electrons around the nuclei. Each element and isotopes of the same element have specific scattering magnitudes not related to the atomic (Z) number. This property is exploited in the concept of "contrast variation" where the differences in scattering amplitude can be used to discriminate between the contributions from different isotopes of the same element (Blakeley, 2016).

With respect to neutron scattering, H atoms have a strong negative scattering length (bc= -3.74 fm), while the isotope of H, deuterium (D), has a positive scattering

length (bc= 6.67 fm) that is comparable to the scattering lengths of heavier atoms (https://www.ncnr.nist.gov/resources/n-lengths/, accessed on 2019-03-19). However, H atoms have a large incoherent scattering cross section that contributes to the background and reduces signal-to-noise in scattering experiments. Unlike with X-rays, the coherent scattering lengths of H and D in neutron crystallographic studies are of similar magnitude as they are for C, N, O, or S atoms. Negative coherent scattering length of H and large incoherent scattering cross section (80.27 barn) contribute to high background that eventually lowers the accuracy of collected crystallographic data. When H are replaced for D this problem is avoided since D has much lower incoherent scattering cross section (2.05 barn) and a positive coherent scattering length (Yee et al., 2017). The larger magnitude of the D coherent scattering length compared to H results in improved signal-to-noise for measured reflections. At the typical resolutions (up to ~2.5 Å), substitution of H for D facilitates the analysis of nuclear density maps by limiting the number of sites where the negative scattering of H atoms leads to nuclear density cancelation effects. Due to this, exchanging H with D is often done to minimize these effects (Langan and Chen, 2013). Neutron scattering studies of D-labelled macromolecules are part of a powerful strategy to determine the H positions. The enhanced scattering signal of biomolecules when H atoms are replaced by D gives an insight into structure-function relationship of the molecule with respect to the location of its H atoms (Fisher et al., 2009; Katz and Crespi, 1966; Yee et al., 2017; Kossiakoff and Spencer, 1981).

Experimental determination of H atom positions in protein and water provides important information for biochemical structure-function studies of proteins and other biological macromolecules. It also allows the direct determination of the protonation states of catalytic groups, amino acid residues, the charge of bound ligands, ligand binding interactions, water molecule orientation, and the nature of bonds involving H. The parameters (distance, angle) of H-bonds can be used in the calculation of hydrogen bond energies, which in turn can be used in molecular-dynamics simulations.

In all neutron scattering techniques, strategic isotopic-labeling (H/D exchange) can greatly increase the amount of useful information obtained from an experiment and allows us to study the parts of molecules that would be otherwise “invisible” (Haertlein et al., 2016; Dunne et al., 2017). Diverse neutron scattering and spectroscopy methods are used for structural studies in biology and nearly all of them use some deuteration, either partial in the solvent (labile H only, H/D exchange) or as incorporated into chemical groups (all H changed for D, perdeuterated) (Blakeley et al., 2015; Langan and Chen, 2013). Methods for perdeuteration of biological macromolecules have therefore a much broader impact than just being useful for NPX and are discussed further in section 2.

The main limitation of neutron experiments results from the inherent low flux of neutron beams, orders of magnitude less than X-ray beams. In practice it means that data collection times are much longer (days or weeks) than for X-rays and large crystals are required (> 0.5 mm^3). Recent technical advances in design of the instruments now allow collecting neutron diffraction data from protein crystals larger than 40 kDa with lattice constants between 10 and 300 Å. To meet the demand for access to neutron crystallography beamlines, new neutron sources are being optimized and planned for the near future, where studies of increasingly complex biological systems and smaller samples will be possible (Langan and Chen, 2013; Meilleur et al., 2018; Fisher et al., 2017). As of today, there are 154 structures determined by NPX available in PDB (https://www.rcsb.org/, accessed 2019.03.01).

Other methods for structure determination

Cryo-electron microscopy

Modern cryo-electron microscopy (cryoEM) is a rapidly developing type of electron microscopy that is benefiting enormously from recent improvements in better microscopes, detectors, and advanced software. Advances in the last years makes it possible to obtain near-atomic resolution 3D electron density maps of larger macromolecular assemblies (ideally these should be larger than 250 kDa) using single particle cryo-EM without the need for crystals (Henderson, 2015). The breakthrough in electron microscopy happened in the 1990s with the technique for rapid cryopreserving of individual biomolecules in a thin layer of vitrified ice (Dubochet and Mcdowall, 1981). This enabled imaging of biological molecules in vacuum at cryo temperatures that limit the beam damage to the sample. The final 3D maps are based on several thousand 2D images of different particles collected at different angles and orientations. Modern detectors and improved software for 3D reconstruction have led to improvements in the achievable resolution of structures (Henderson, 2015; Murata and Wolf, 2018). There are significant technical challenges to the use of cryo-EM, including complicated analysis of multiple conformations of molecules in their native environment, electron beam-induced movement during exposure, molecular weight limitations, radiation damage of the sample, and low signal-to-noise ratio. However, today there are already 7640 entries deposited in the Electron Microscopy Data Bank (https://www.emdataresource.org) and they account for 2868 PDB coordinate entries, several reaching as high in resolution as 1.8 Å (https://www.rcsb.org/, accessed 2019-02-20).

NMR spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is based on absorption of electromagnetic radiation by atomic nuclei placed in a strong magnetic field. A distinctive set of an electromagnetic signal with a frequency characteristic of the

magnetic field at the nucleus is produced and used to build a model of the protein (Mittermaier and Kay, 2009). Like all the other techniques described in this section, NMR also recently experienced big improvements, making it more suitable to study larger proteins, despite sensitivity losses and increased spectral complexity. In contrast to previously described techniques, NMR provides information of the protein in solution that can better mimic the physiological or non-physiologically solutions (e.g. denaturation), or even the inside of the cell at room temperature. Therefore, it is the method of choice for studying the dynamic features of the molecular structures, thermodynamics, interactions between the proteins and other molecules (e.g. macromolecules, ligands).

Uniform and selective isotopic labelling with ^{1}H, D, ^{13}C and ^{15}N extends the use of NMR for studies of larger proteins (> 35 kDa) and protein complexes (Verardi et al., 2012; Mittermaier and Kay, 2009; Katz and Crespi, 1966). Replacement of atoms with stable isotopes is used as spectral editing since it limits the relaxation rate and gives narrower, non-overlapping peaks. It is isotopic labelling where methods used in NMR overlap with NPX, additionally the two techniques provide unique insight into the dynamics of hydrogen bonding. The PDB contains more than 12000 NMR structures (https://www.rcsb.org/, accessed 2019-03-01).

Electron crystallography

Electron diffraction is a rather new and emerging method for high resolution structure determination from protein crystals. Electrons are accelerated by an electrostatic potential to gain certain energy level and wavelength before they interact with the small, thin crystals. Electrons interact with material stronger than X-rays and can yield meaningful data from a relatively small sample. The 3D electron diffraction data and a structure-refinement algorithm, resembles the standard refinement procedure used in X-ray diffraction. Structure determination is possible by merging data from hundreds of individual crystals. A particular feature of this method is a possibility of locating hydrogen atom positions directly in sub-micrometre-sized crystals based on the scattering power of an atom (Palatinus et al., 2017; Clabbers et al., 2017; Duyvesteyn et al., 2018). Electron diffraction can be done in common transmission electron microscopes. However, the technique does have certain requirements in sample preparation. Even if much smaller crystals are required in comparison with X-ray crystallography, the right shape of very thin crystals (100 to 500 nm size) need to be consistently produced. In the PDB there are currently 113 structures determined by electron crystallography (https://www.rcsb.org/, accessed 2019-03-01).

X-ray Free Electron Lasers

Serial femto-second crystallography at X-ray Free Electron Lasers (XFEL) is proposed to revolutionize the methods for X-ray crystallography. An XFEL creates

femtoseconds pulses of radiation at extreme brightness. This beam is then crossed by a stream of micro-crystals (micrometres in size) and millions of diffraction patterns from the crystals are merged together to yield complete data sets (Chapman et al., 2006).

As mentioned before X-ray irradiation produces global and radiation specific damages. The first one is seen in gradual fading of Bragg spot intensities or diffraction limit. Other effects are evident in the processed data where with increasing dose there is an increase in the unit-cell parameters, crystal mosaicity, Wilson B-factors, a decrease in I/(I) and a worsening of merging R-factors. Specific damages in protein crystals emerge as changes in electron density maps from disulfide bond breakage, elongation, and decarboxylation of amino acid residues (Garman and Weik, 2017; Ravelli and McSweeney, 2000).

In XFEL diffraction occurs from a single crystal before specific chemical and structural damage within the crystal has time to develop, this is called “diffraction-before-destruction” (Neutze et al., 2000). The method is extremely powerful since it allows the study of molecular processes that occur over very short time scales, such as the absorption of light by biological chromophores. Performing experiments using XFEL has many challenges, including oversubscription and needing very large quantities of optimal crystals, but it has led to multiple breakthroughs in structural biology, like the recording of molecular movies. Rapid development of tools, protocols, instrumentation and data processing in this technique is promising and will certainly answer more important biological questions in future when XFEL development stage will terminate and the instruments will become more accessible (Nango et al., 2016). 116 structures determined using an XFEL have been deposited in the PDB (https://www.rcsb.org/, accessed 2019-02-19).

Neutron protein crystallography (NPX)

Neutron diffraction or elastic neutron scattering is a method that can be used to generate diffraction datasets from protein crystals to subsequently determine and refine protein structures. It is very similar to X-ray crystallography in concept, but neutrons are scattered from the atomic nuclei. When using X-rays alone, it might be very challenging to assign positions of H atoms even if sub-atomic resolution data can be collected to a very high resolution, such as crambin (PDB ID 3nir), reported to 0.48 Å (Schmidt et al., 2011). Even in this extreme case, it is practically possible to locate only about two-thirds of H atoms, usually in well-ordered regions of the protein. In protein crystal structures H atoms are found in solvent molecules or on the protein itself and due to movement are normally invisible, even if involved in crucial functional roles (Chen and Unkefer, 2017). NPX offers a complementary

approach for locating H atoms at medium resolution by using the intrinsic neutron scattering properties of D atoms (Blakeley et al., 2015).

In 1969, neutron diffraction of protein crystals was first applied for studying myoglobin (Schoenborn, 1969) and in 1981 a deuterated protein crystal was used for neutron diffraction studies of monoisopropylphosphoryl-trypsin (Kossiakoff and Spencer, 1981).

Due to the strong neutron scattering of D, it is possible to explicitly observe numerous features involved with ligand binding to proteins: protonation states of catalytic residues and H-bonding of amino acids; water networks and how water molecules interact with each other and the protein; the charged state of the drug and how that modulates binding; the possibility to discern solvent species (H^+ vs OH^- vs H_2O vs H_3O^+). In summary, NPX can provide an added layer of structural and chemical information that is not otherwise obtainable (Langan and Chen, 2013).

There are currently five operating instruments available for macromolecular crystallography. The Oak Ridge National Laboratory (ORNL) in the USA hosts IMAGINE instrument at the HFIR neutron source and MaNDi at the Spallation Neutron Source. In Japan there is the iBIX instrument (Figure 3b) at the Japan Proton Accelerator Research Complex (J-PARC) and in Europe there are two instruments operating at reactor neutron sources: LADI-III (ILL, Grenoble, France) and BIODIFF (MLZ, Munich, Germany). The macromolecular diffractometer NMX at the European Spallation Source in Lund will be available for users in the near future.

In a typical NPX experiment, a single protein crystal is exposed to a neutron beam. The crystal scatters the neutrons which are then recorded as intensities or spots on the detector (Figure 3a). From the diffraction pattern (Figure 3c) structural information can be extracted. Due to weak interactions of neutrons with matter, much larger crystals (> 0.5 mm^3) and much longer data acquisition times (days to weeks) are necessary compared to X-ray experiments (< 0.01 mm^3, seconds to minutes) (Fisher et al., 2017; Lakey, 2009).

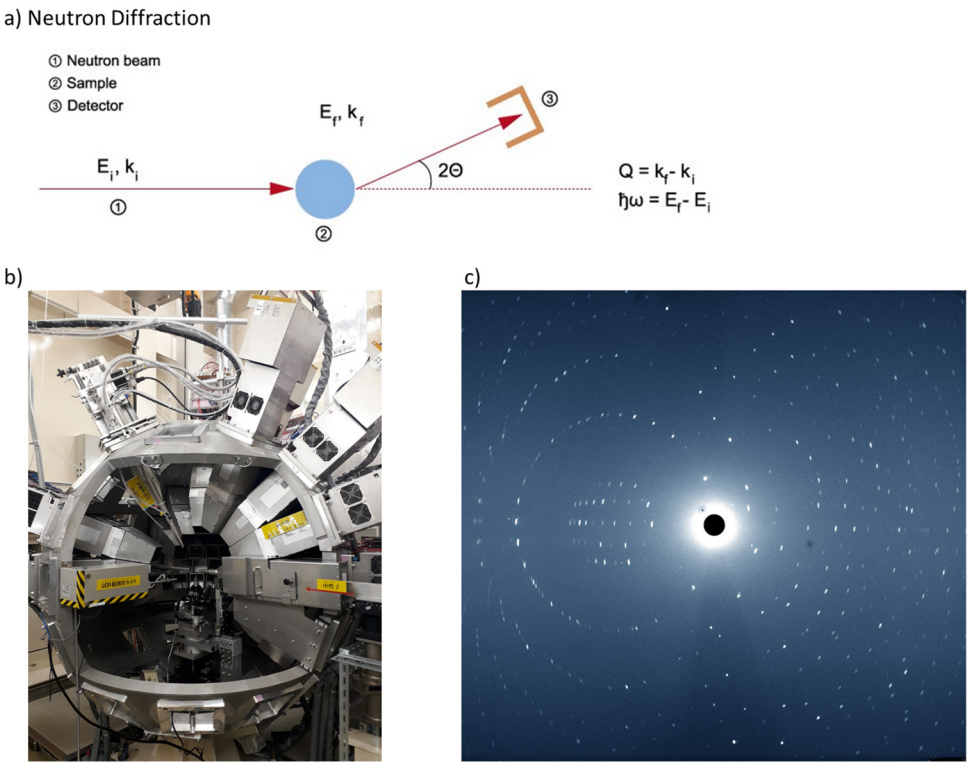


Figure 3: Neutron diffraction. (a) Neutron diffraction principle (adapted from (Crespo, 2012)). (b) Neutron protein crystallography instrumentation - Photograph of the iBIX instrument at the Japan Proton Accelerator Research Complex (J-PARC, Tokai, Japan). Photograph by Katarina Koruza. (c) Representative diffraction of the hCA IX_{mimic}:saccharin crystal tested at the LADI instrument at ILL, Grenoble (Koruza et al., 2019).

One advantage of NPX is the collection of diffraction data at room temperature (RT) without cryo-cooling effects and without radiation damage issues (Blakeley et al., 2015). Refinement of crystal structures can either be done in joint X-ray/neutron mode or against neutron data alone with programs such as phenix.refine (Adams et al., 2011). The main limitation of NPX remains the need of growing large D-labeled crystals, even though the performance of neutron sources, instruments, and data analysis have improved during the last decade. Studies on relatively smaller crystals and of larger and more complex systems are now possible with shorter data collection times (Fisher et al., 2017; Blakeley et al., 2015; Meilleur et al., 2018).

Some examples of perdeuterated proteins in literature are myoglobin (Shu et al., 2000), human CA II (Budayova-Spano et al., 2006), human transthyretin (Haupt et al., 2014), type III antifreeze protein (Petit-Haertlein et al., 2009), cholesterol oxidase (Golden et al., 2015), *Pyrococcus furiosus* rubredoxin (Gardberg et al., 2010), human aldose reductase (Hazemann et al., 2005), HIV protease (Weber et

al., 2013), human Mn superoxide dismutase (Azadmanesh et al., 2017), haloalkane dehalogenase (Liu et al., 2007) and cytochrome P450cam (Meilleur et al., 2005).

Isotope effect

Interest in studying the effect of deuterium on various organisms emerged soon after the discovery of the stable H isotope in 1932 by Urey et.al. (Urey, 1932). Early reports of deuteration of yeast (Richards, 1934), algae (Daboll et al., 1962; Crespi et al., 1960), plants (Bhatia and Smith, 1968) and bacteria (De Giovanni, 1960; Crespi et al., 1959) started to emerge in the sixties. Effects of H/D exchange on biophysical properties of organic compounds, such as dynamics, denaturation temperature, aggregation, kinetics, and solubility became known as isotope effect and lead the interest in deuteration and experimental techniques (Hattori et al., 1965).

Kinetic isotope effects are observed as a change in chemical processes where the reaction rate depends on the masses of the reacting molecules. Given the same amount of kinetic energy, molecules of lighter mass m_1 will react at a rate $\sqrt{(m_2/m_1)}$ faster than the heavier molecules of mass m_2 (Wieser and Brand, 1999). Introducing D (atomic weight of D=2.014), instead of H (atomic weight of H=1.008) in labile, exchangeable positions, helps to stabilize the protein, while incorporating D in non-labile positions has the opposite effect by decreasing non-polar amino acid side chains interactions. There are many plausible explanations for the observed changes in protein properties. C—H and C—D bonds have fundamentally different, albeit subtle, properties. C—D bonds have lower vibrational amplitudes and frequencies, leading to slightly shorter, stronger bonds compared to C—H bonds. The additive effect over an entire protein will not be insignificant where perdeuterated side chains are smaller. Smaller side chains, especially for non-polar residues in the hydrophobic core of the protein, lead to weaker hydrophobic interactions (Hattori et al., 1965; Kholodar et al., 2017; Kuchitsu and Bartell, 1962).

The stable isotope (e.g. D) forms stronger bonds that require higher energy to break and slow down the chemical reaction rate. Due to ubiquitous presence of H transfer in various biological processes, this subtle change in energy and structure may alter the ground state and transition state of the active site and might also change the kinetics of enzymatic reactions (Kholodar et al., 2017). Slower reaction rates on the molecular level also reflect in the growth rate of organisms under deuterated conditions, where changes in metabolic fluxes, decrease in growth and development have been observed in addition to morphological changes (Andriukonis and Gorokhova, 2017). Despite the profound effect that replacing H with has on living organisms, the molecular response mechanisms underlying adaptation are unknown.

Additionally, an isotope effect is observed trough alteration of the pH optimum of the protein and a compensatory adjustment of 0.4 units for $\Delta pK_a = pK_a D - pK_a H$ was suggested by Covington et al., (Covington et al., 1968). However, it is rarely observed that the difference is that dramatic in H_2O solutions containing protein, polysaccharide or nucleic acid (Rubinson, 2017; Kholodar et al., 2017). In practical terms it means that deuteration affects acidities of individual residues and might change the isoelectric point of the protein, local charges, and charge distribution across the protein (Kholodar et al., 2017). The same effect can be taken into consideration for biological buffers. In a recent article, Rubinson empirically suggest guidelines for corrections in most commonly used biological buffers: no correction is needed for any D_2O content, where p(H,D) values are in the range $4 < p(H,D) < 8$ and when D_2O content is less than 50% in the range $8 < p(H,D) < 10$. Corrections depend on the specific conditions and the identity of the buffer, when D_2O content is greater than 50% if the $4 < p(H,D) < 10$, however the correction was always lower than 0.4 (Rubinson, 2017).

Specifically, for neutron structural studies, researchers have to demonstrate that perdeuterated and hydrogenated biomolecules do not show different structural properties (Yee et al., 2016; Hattori et al., 1965; Brockwell et al., 2001). In the case of incomplete H/D exchange, different regions of biological macromolecules may exchange at different rates, depending on solvent accessibility, *p*Ka, and stability (Bennett et al., 2008). Highly dynamic and solvent-exposed regions (i.e. loops connecting the alpha helices) will exchange rapidly while less dynamic regions involved in hydrogen bonding networks or buried within the interior of the protein (i.e. β-sheets or α-helices) will exchange slower (Marcsisin and Engen, 2010). Isotope effects and the possibility of back-exchange during the neutron experiment should be considered since it may limit the quality of neutron scattering length density maps and complicate overall analysis (Yee et al., 2017).

Deuteration of proteins

Complete deuteration (perdeuteration) of biological molecules is only feasible when they are expressed under deuterated conditions *in vivo.* The production of recombinant proteins utilizes the machinery of living cells. The most commonly used expression hosts for recombinant protein production are prokaryotic cells (e.g. *Escherichia coli*) and eukaryotic cells like yeast (*Pichia pastoris*), insect cells (Baculovirus Expression Vector System) or mammalian cells (e.g., HEK293, CHO). An overview about their use so far to produce proteins for structure determination is presented in Figure 4.

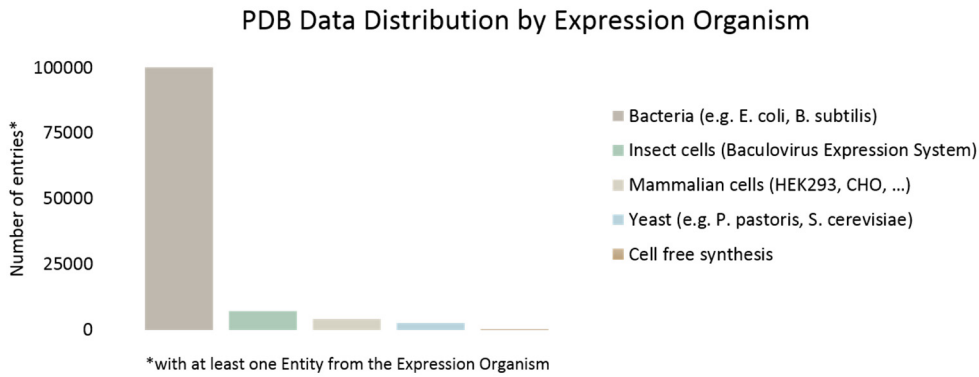


Figure 4: Distribution of PDB entries by expression organism. (Adapted from PDB; http://www.rcsb.org/pdb/, accessed on 2019-03-01).

Deuteration in bacterial systems is reasonably mature from a practical standpoint. However, very little is known about the response mechanisms by which bacterial organisms react to and possibly adapt to D-induced cellular toxicity and stress. Therefore, perdeuteration of recombinant proteins in bacterial cells might be far from what could optimally be achieved. Understanding these mechanisms would allow us to mitigate the negative effects of D and maximize recombinant protein production and expression of proteins under deuterated conditions.

Growth of eukaryotic and prokaryotic cells in deuterated media is slower than in hydrogenated media and yields of biomass or expressed recombinant proteins are usually lower (Meilleur et al., 2009; Katz and Crespi, 1966). The large amounts of highly pure recombinant deuterated protein that is needed for a single NPX project makes sample preparation challenging, repetitive, and expensive (Fisher et al., 2017). The amount of required protein depends on the protein and the neutron experimental technique and these in turn impact the design of the protein production protocol used to achieve the yields and level of deuteration. For example, in SANS experiments of biological complexes match-out of labelled proteins is very effective at 40% deuteration levels (Dunne et al., 2017), therefore expression can follow a protocol in minimal medium based on 85% D_2O and unlabelled (protiated) carbon source.

Eukaryotic proteins, membrane proteins, large protein complexes, and post-translationally modified proteins, are at the centre of interest for current drug discovery research. For example, G protein-coupled receptors are an attractive target class and the target of 33% of all marketed small molecule drugs (Santos et al., 2016). Complex proteins can often not be expressed in *E. coli* in an active form, due to improper folding or lack of the appropriate secretion machinery. Membrane proteins pose a special challenge due to low protein production yield, limited

solubility, and the need to maintain a membrane-like environment. Therefore, deuteration and partial deuteration efforts are made towards the use of eukaryotic expression systems like yeast *P. pastoris* (Skora et al., 2015), insect cells (Skora et al., 2015; Opitz et al., 2015) or even mammalian cells and cell-free systems (LaGuerre et al., 2015; Ozawa et al., 2005). These few reports, compared to the prevalence of deuterated protein production in *E. coli,* indicate the research and deuteration trends that are emerging with availability of new neutron sources around the world and especially by advances of exclusive site-specific labeling approaches for NMR.

In general, deuteration of proteins in eukaryotic systems is still very demanding and results in low yields and incomplete deuteration. In addition to this, the prices of perdeuterated chemicals needed for eukaryotes are often prohibitively high. D toxicity also increases with the complexity of host organism, meaning that it is more toxic to higher eukaryotes then it is to prokaryotes. Consequently, *E. coli* remains the most popular host for perdeuteration of biological macromolecules.

Most perdeuteration protocols are based on high cell-density growth of *E. coli* in order to obtain higher yields, usually in a fermentor. Initial stepwise adaptation of cell culture from optimized minimal medium to the deuterated minimal medium is recommended in order to avoid a prolonged lag phase (Golden et al., 2015; Duff et al., 2015; Paliy et al., 2003; Sivashanmugam et al., 2009; Michalczyk et al., 2015; Cai et al., 2016). Cells are grown either in flasks to an OD_{600} up to 20 (Cai et al., 2016) or in a fermentor where OD_{600} above 30 is reached prior to the induction of gene expression with IPTG (Golden et al., 2015). Higher densities are reached with fed-batch culturing in fermentor where the carbon source is fed to the culture during the linear growth phase and better control of dissolved oxygen, temperature and pH regulation is available in comparison with growth in flasks. There is also one report available where deuterated proteins (Sivashanmugam et al., 2009) were expressed in auto-induction medium (Studier, 2005), or where alternatives to glycerol as a carbon source were studied for *E. coli* growth (Paliy and Gunasekera, 2007). It appears however that optimization of growth medium is case specific and depends on a variety of factors, starting from expression system, laboratory equipment, to the availability of deuterated chemicals, and the desired resulting level of deuteration. Different research groups and deuteration facilities report the use of perdeuterated media with commercially available or lab-made algal extract as a supplement for minimal media which has a strong positive effect on cell growth and protein yields, while maintaining a high deuteration level (Katz and Crespi, 1966; Fiaux et al., 2004; O'Brien et al., 2018; Klopp et al., 2018; Franke et al., 2018; Opitz et al., 2015). Photosynthetic algae grow autotrophically in minimal deuterated media by using only CO_2 as inorganic carbon and light energy (Daboll et al., 1962). In this regard algal extract is an easily available source of deuterated carbohydrates, amino acids, vitamins and trace metals that form rich media for culturing of

eukaryotic cells. Additionally, deuterated algal secondary metabolites and cell components like accumulated fatty acids, lipids or carbohydrates are of interest for studies using neutron reflectometry and SANS (Fisher et al., 2017). In the same way as algae, labeled yeast extract was used as a supplement for complex insect cell growth media for selective labeling in NMR (Opitz et al., 2015).

Practical aspects of protein deuteration

Production of deuterated proteins in the case of NPX is the same as for standard protein production for other methods. Typically, it begins with selection of an expression system, cloning the target gene into a suitable expression vector, transformation of cells, growing transformed cells, protein production, crystallization, and finally structural studies as illustrated in Figure 5.

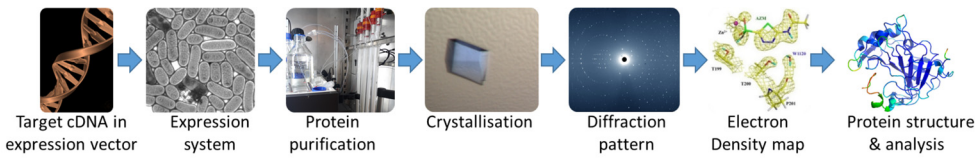


Figure 5: Workflow for protein structure determination using X-ray or Neutron crystallography.

A major concern in the production of deuterated proteins is the high price of deuterium and deuterated reagents (e.g. D_2O, d-glycerol, d-glucose). To minimize the cost of deuteration many labs routinely recycle the D_2O of used media by rotary evaporators and then re-use it to make media for cell culture. This saves money but of course the D content decreases over time and the resulting proteins are < 98% deuterated.

Plasmid Choice and Selection

Commonly used *E. coli* protein production strains, like BL21(DE3) can produce good yields of deuterated proteins when the culture is adapted to deuterated medium (Paliy et al., 2003; Duff et al., 2015). Shorter lag phase and better yields are observed when cells are adapted directly after transformation with plasmid. However, it is the adaptation step where unknown mechanisms enhance D tolerance and it can be speculated that these are metabolic adaptations and maybe even non-specific mutations. Extended growth times may cause plasmid loss, therefore kanamycin-resistant plasmids with T7 promotor system and lac*I* repressor are recommended to reduce expression before induction and to secure the plasmid stability (Duff et al., 2015). In the case of using ampicillin as a selection marker, the β-lactamase enzyme that confers bacterial resistance is located in the periplasmic space and can easily leak into the culture medium. There it degrades ampicillin and can either cause

bacteria to lose their plasmid or can lead contaminant cells to overgrowth. In contrast, kanamycin interferes to some extent with protein synthesis in the cytosol, therefore it is less likely to leak into the culture media (Meilleur et al., 2009).

Partial deuteration - H/D exchange

Using heavy water as a solvent for biological molecules will result in replacement of labile H atoms with D atoms in solution, but non-labile H atoms such as those in C–H groups will not exchange (Blakeley et al., 2004; Bennett et al., 2008). In the case of NPX, the protein is commonly expressed, purified, and crystallized in protiated conditions and then subjected to H/D exchange prior to neutron data collection. This can be achieved in different ways, such as soaking the crystal in deuterated buffer, setting up a crystallization drop in deuterated buffers, or through vapor-exchange once the crystal is mounted in the quartz capillary (Blakeley et al., 2015; O'Dell et al., 2016; Bennett et al., 2008). In general most of labile H would exchange within 30 days (Bennett et al., 2008). Exchange rates vary depending on hydrogen bonding, solvent exclusion by protein folding, solvent pH, and incubation temperature (Yee et al., 2017; O'Dell et al., 2016).

Partial deuteration - using unlabelled carbon source

Labile H constitutes around 25% of the total H atoms of a protein and can be exchanged by one of the methods described above. To exchange the remaining H, proteins can undergo partial deuteration by using D_2O and unlabelled C source (e.g., glycerol, glucose etc.) during synthesis *in vivo* (Bennett et al., 2008). Longer generation times of the expression host are one of the consequences of the resulting isotope effect. Lower yields of the cells can be offset by a series of adaptation steps or by pre-growing cells to high density in a protiated rich medium, followed by fast reconstitution into deuterated medium just prior to induction of protein expression (Duff et al., 2015; Michalczyk et al., 2015).

Perdeuteration - using a deuterated carbon source

A typical perdeuteration experiment to produce proteins for NPX includes progressive adaptation of microorganisms to increasing concentrations of D in the growth medium. An *E. coli* cell culture is grown in minimal media prepared with D_2O, D-labeled glycerol, and H/D-exchanged salts. Improvement of microorganism growth can be achieved by introducing isotopically labelled metabolites extracted from yeast or algae. Perdeuteration with a deuterated C source significantly benefits NPX and other types of neutron experiments, since ~99% of H are replaced by D (Paliy et al., 2003; Paliy and Gunasekera, 2007).

Determining the deuteration level

Mass spectrometry (MS) is commonly used to monitor the level of deuterium exchange in a protein. The two primary methods for ionization of whole proteins are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF). One of the main differences, between MALDI and ESI is the state in which the sample is introduced to the ion source. Sample preparation for ESI includes quenching of deuterated protein by acid and subsequently, an acid-tolerant protease such as pepsin is added to induce proteolysis. The resulting quenched peptides are desalted and separated by reversed phase liquid chromatography before sprayed directly into a mass spectrometer to determine the mass of the intact protein (Konermann et al., 2011). In the case of MALDI-TOF MS, the ions are desorbed from the solid phase. The sample preparation requires mixing the protein with a matrix and drying, which introduces another possibility for back-exchange with protons from the matrix (Villanueva et al., 2000). Measuring the mass of the whole protein gives global information about the protein but does not locate where the deuterium is within the protein (Marcsisin and Engen, 2010). A major problem for the measurements of deuterium exchange by MS is occurrence of amide back exchange (D → H) during sample preparation and analysis (Konermann et al., 2011). Use of deuterated solvents, inert atmosphere and control of temperatures and pH limits the back-exchange (Marcsisin and Engen, 2010; Villanueva et al., 2000).

Protein crystallization

Protein crystals are solids composed of highly ordered, structurally uniform protein molecules. They are usually limited in size, can be quite brittle, are temperature sensitive, affected by radiation damage, and easily disintegrate if allowed to dehydrate. Additionally, due to the small size and great number of atoms they exhibit weak optical properties and diffract X-rays and neutrons poorly (McPherson and Gavira, 2014). Crystallization of the first enzyme (urease) was reported by Sumner in 1926 (Sumner, 1926). This work proved that proteins can be crystalized and opened possibilities of studying biological macromolecules with diffraction. In 1934 the crystals of pepsin, grown in Uppsala were studied by X-rays by Bernal and Crowfoot and this was the first reported diffraction experiment from a crystal of a protein (Bernal and Crowfoot, 1934). The application of X-rays to biological macromolecules brought interest to the field and eventually lead to its great expansions towards the end of the 20^{th} century. Macromolecules that are interesting for structural studies nowadays are large and complex proteins, nucleic acid complexes, viruses and ribosomes. In case of neutron studies, these molecules result in big unit cells that diffract weakly at low-flux neutron sources. Optimization of existing instruments, addition of new instruments, and advances in refinement

software have led to increasing numbers of deposited NPX structures in the PDB. However, the crystallization step still presents a major bottleneck for NPX since most available instruments require large crystals (in the 0.5 - 1 mm^3 range). Growing large crystals is not trivial and can consume hundreds of milligrams of pure protein (Fisher et al., 2017). A good approach for obtaining high quality neutron diffraction data is to start with pure, homogenous and monodispersed protein that has been characterized through a crystallization phase diagram (Figure 6) (Gavira, 2016).

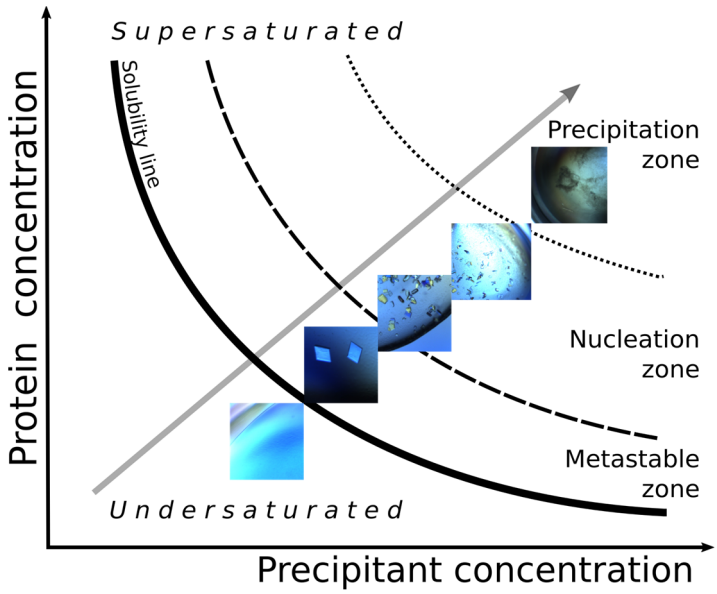


Figure 6: A schematic phase diagram showing the solubility of a protein in solution as a function of the concentration of the precipitant. The grey arrow shows the expected path of crystallization that leads from the undersaturated zone below the solubility curve, where no crystals grow, to the precipitation zone, where supersaturation is too high and 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. Ideal zone for crystallization is the metastable zone, where the nucleation does not occur spontaneously, but the supersaturation is low (Asherie, 2004). The scheme is adopted from paper IV in this thesis (Koruza et al., 2018a).

Protein crystallization is based on creating a supersaturated solution of a macromolecule with the addition of precipitants (neutral salts, high molecular weight polymers, organic solvents, polyalcohols) and by manipulation of crystallization conditions (pH, incubation temperature, ionic strength and chemical/biochemical modification of proteins). Different approaches have been developed to promote crystallization, and among the most widely used are vapor diffusion, batch, and dialysis (Figure 7) (McPherson and Gavira, 2014). The main idea behind the specific set ups is to control the thermodynamic and kinetic contribution to the supersaturated state where the protein nucleates.

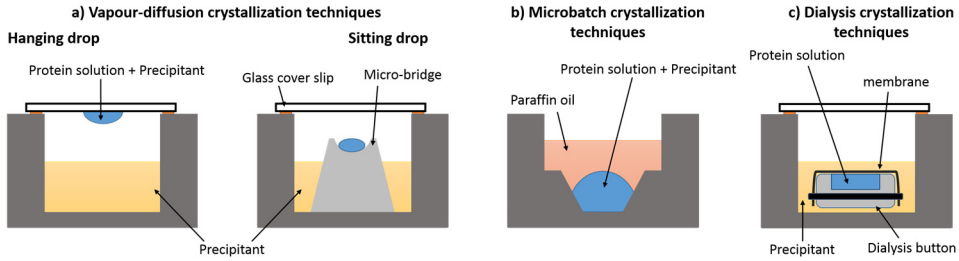


Figure 7: Crystallization techniques. (a) Schematic presentation of hanging and sitting drop vapour-diffusion technique; (b) Schematic presentation of microbatch crystallization technique; (c) Schematic presentation of dialysis crystallization technique. Adapted from McPherson and Gavira, 2014.

Large crystal growth

Vapour diffusion (Figure 7a) is a method of choice for NPX where larger volumes (up to 1000 µL drops) are required. A drop of protein mixed with precipitant equilibrates against the reservoir in a sealed environment with usually double the concentration of precipitant as a reservoir. Under these conditions, water leaves the drop to equilibrate with the reservoir solution, effectively increasing the protein concentration until the supersaturated region of protein solubility is reached. Solubility of the protein in solution when using vapour diffusion techniques can be controlled with the surface area of the drop and its size. Therefore, we use the hanging drop technique (Figure 7a - left) and sitting drop (Figure 7a - right).

In a batch crystallization technique (Figure 7b), the protein is mixed with precipitant and can be covered with a layer of oil (paraffin, silicon, or a mixture of two - Al's Oil). Conditions are chosen so that the supersaturation point is reached at the moment of drop preparation, meaning the process does not rely on evaporation. The advantage of batch crystallization is that it is easy to scale-up for large crystal growth.

For the dialysis crystallization technique (Figure 7c) the protein solution is separated from the precipitant with a semipermeable dialysis membrane that allows small molecules such as salts, additives, and other crystallization reagents to diffuse trough the membrane while preventing the protein from diffusing out. Dialysis is the method of choice where the macromolecule crystallizes at low ionic strength (McPherson and Gavira, 2014).

Seeding

Seeding is a technique where crystals are crushed in mother liquor and diluted to create a seed stock that is added to crystallization solutions to aid with nucleation (D'Arcy et al., 2003). It allows optimization of crystal growth conditions independently from conditions needed for initial nucleation.

Spontaneous nucleation cannot occur in undersaturated solution (Figure 6), therefore seeds introduced in this phase will dissolve. The same applies along the saturation curve, but crystal seeds added will not dissolve or increase in size. Spontaneous homogeneous nucleation occurs above the saturation curve - in supersaturation zone but it is a lasting process. This zone is ideal for seeding and crystal growth; consequently, seeding can remarkably speed up the crystallization and provide more consistent results (Luft and DeTitta, 1999). In the nucleation and precipitation zone where spontaneous nucleation occurs, seeding often results in excessive nucleation and formation of amorphous precipitate (Asherie, 2004). Additionally, the size and number of crystals can be manipulated by using a dilution series of seed stock (Luft and DeTitta, 1999). This technique can generate better-diffracting crystals since crystals are more likely to grow in the metastable zone (Asherie, 2004). Seeding from the seed stock can be applied at any stage of a crystallization experiment and automated with a pipetting robot.

Strategies for co-crystallization with ligands

Several methods are established for preparing protein:ligand complexes for crystallization, but they are often destructive for crystals and often conducted manually. The method of choice is co-crystallization where a protein and ligand are mixed and incubated prior to drop set-up. This often results in prolonged crystallization times, small crystals, and low success rate. A second method is to soak the crystals by adding the ligand to an existing crystallization drop. This method can be destructive and require optimization. In many cases crystals dissolve or shatter upon the addition of ligand solution. With “dry” co-crystallization some of the abovementioned issues are removed, especially when working with water insoluble drugs or when performing in situ characterization (Gelin et al., 2015). The ligand solution is dispensed into the crystallization well and solvent is evaporated. Afterwards the entire drop, including mother liquor, seeds and protein, are moved into the same well. This procedure is risky as it involves mechanical manipulation of the entire drop, including crystals. However, it is quite effective as the mother liquor then dissolves the ligand directly into the drops where the crystals are, without causing osmotic shock or dissolution issues.

Contributions of neutron crystallography to inhibitor design

Enzymes are catalysing almost every biological reaction in the cell therefore; impairment of an enzyme in the metabolic network can sometimes be directly related to a disease. An effective drug may target the enzyme to recover the equilibrium in the metabolism. Thus, it is crucial for drug discovery to identify the

drugs specific for a certain target protein. Knowledge of protein structure has contributed to drug discovery and development and is broadly used in industry and academia (Thomas et al., 2017).

An early example of using protein crystallography in structure-guided drug discovery is the human immunodeficiency virus (HIV) protease and the design of AIDS antivirals in the 1990s (Thomas et al., 2017). Neutrons studies followed up on the idea of structure-guided drug discovery almost 20 years later with the example of HIV-1 protease, a dimeric aspartic protease that cleaves the nascent polyproteins of HIV-1 and has an essential role in replication of the virus. Neutron structural studies of the inhibitor KNI-272 revealed direct experimental evidence for the catalytic mechanism of HIV-1 protease (Adachi et al., 2009). Further on, the example of the clinical drug amprenavir (Weber et al., 2013) and darunavir (Gerlits et al., 2017) in complex with HIV-1 protease have been studied by room-temperature joint X-ray/neutron crystallography to better understand the contribution of solvent and H-bonds to ligand binding, as well as the main mechanism behind inhibition of HIV-1 protease.

Bennet et al. provided one of the first examples of using NPX to study ligand binding, through identification of protonation states directly from neutron density maps and to analyse solvent structure involved with inhibitor binding. The subject of their study was *E. coli* dihydrofolate reductase in complex with the well-known chemotherapeutic agent methotrexate (Bennett et al., 2006).

In recent years, a growing number of neutron structures have been deposited in the PDB, including human CA II, where the neutron structures of hCA II in complex with clinically used sulphonamide inhibitors acetazolamide (Fisher et al., 2012), methazolamide (Aggarwal et al., 2016), brinzolamide, dorzolamide, and ethoxzolamide were determined (Kovalevsky et al., 2018).

Another example is the neutron structure of farnesyl pyrophosphate synthase with the osteoporosis drug risedronate bound (Huang et al., 2014).

Human galectin-3 is an emerging pharmaceutical target involved in various diseases, including inflammation, diabetes, cancer proliferation and metastasis. A recent study by Manzoni et al. presented neutron studies of galectin-3 in complex with lactose and glycerol and the idea is that these structures inform drug design efforts against galectin-3 (Manzoni et al., 2018).

Recently Schiebel et al. reported a model study for better defining ligand binding through NPX. They used high resolution neutron structures, X-ray, thermodynamic data and computer simulations to reveal the geometry of H-bonds between trypsin and the inhibitors of trypsin along with the dynamics of the residual solvation pattern (Schiebel et al., 2018).

Despite the overall number of neutron structures is still relatively small, there is a growing number of examples for which neutron crystallography revealed the molecular details of small molecule inhibitor binding. Neutron studies can give useful information to guide structure-based drug design by revealing unfilled pockets in the active site, the role of water-mediated binding, and areas where H-bonding can be introduced, or removed, by chemical modification of new ligands (Oksanen et al., 2017). In such a way NPX has provided the answers to question that have remained elusive using other techniques and may contribute to the development of more specific inhibitors for therapeutic applications (Blakeley, 2016). However, most reports mentioned above are relatively new and/or aimed at showing, or solving open questions about the mechanisms of already marketed drugs. It is therefore not surprising that reports of new inhibitors developed with input from NPX that reached the clinic are not yet available.

Carbonic Anhydrases

The model protein used in my thesis research is an isoform of the human carbonic anhydrases (CAs). CAs are mostly zinc-containing metalloenzymes that catalyse the reversible hydration of CO_2 that produce HCO_3^- and a proton (reaction 1).

$$CO_2 + H_2O \rightarrow HCO_3^- + H^+ \quad (1)$$

This reaction can occur spontaneously without a catalyst at physiological pH, however the rate of the reaction was calculated to be significantly lower than the observed physiological rate of CO_2 respiration (Meldrum and Roughton, 1933). Almost all the organisms contain the enzymes to catalyse this reaction since poorly soluble CO_2 could damage the cellular components if accumulated inside the cells.

Since the discovery of two distinctive forms of mammalian CAs in 1932 there were 7 genetically unique families of CAs identified. They all catalyse the rapid inter-conversion between H^+ and HCO_3^- and have evolved independently at least 7 times (Silverman and McKenna, 2007). These enzymes represent 7 families of CAs, the α-, β-, γ-, δ-, ζ-and η-, and θ-CA class, and are classified by differences in structural fold, some of them presented in Figure 8.

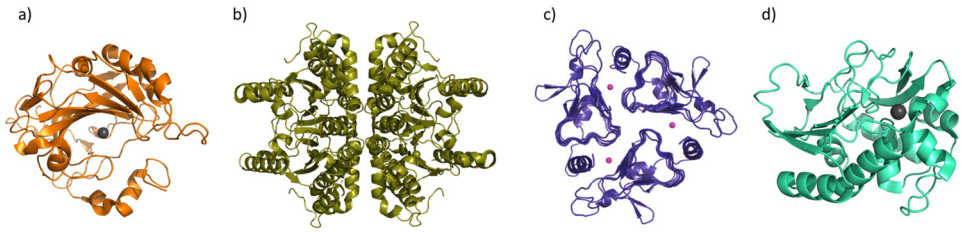


Figure 8: Structural diversity of CA families. Examples of CAs from: (a) α family – monomeric α-CA, hCA II (PDB ID 3kkx), (Fisher et al., 2010); (b) β family - tetrameric β-CA, VchCA from *Vibrio cholerae* (PDB ID 5cxk), (Ferraroni et al., 2015); (c) γ family - trimer of γ-CA from *Methanosarcina thermophila* (PDB ID 3OTM), (Iverson et al., 2000); (d) ζ family - monomeric ζ-CA R3 domain of Cd(II) bound *Thalassiosira weissflogii* (PDB ID 3uk8) (Alterio et al., 2012). All figures were made using *Pymol* (Schrödinger, 2015).

All CAs are metalloenzymes, where the metal ion (Zn(II), Fe(II), Co(II) or Cd(II)) is crucial for the activity of an enzyme. The tetrahedral coordination of a metal ion to three active site residues and a water molecule is conserved between the classes. The CAs from α class (Figure 8a) that includes all the mammalian CAs are the most studied ones. They have Zn that is coordinated by three His residues and the catalytic reaction follows a two-step catalytic mechanism. The first step of catalytic reaction is conversion of CO_2 into bicarbonate through the nucleophilic attack of the reactive Zn-bound hydroxide, and the resulting bicarbonate is subsequently displaced from the zinc by H_2O molecule. The second step of the reaction is the transfer of a proton from the Zn-bound H_2O through an intervening network of hydrogen bonded water molecules to the bulk solvent for the regeneration of the Zn-bound hydroxide. The final acceptor of the proton is bulk water that is moved from the active site water network via a proton shutting residue (His64 in hCA II) (Domsic et al., 2008; Kim et al., 2016; Silverman and McKenna, 2007).

β-CAs were isolated soon after α-CAs in the 1940s from plant chloroplast and have since been identified in eubacteria, algae, and archaea. β-CAs are dimeric (Figure 8b) and they exhibit allosteric regulation. Their catalytic activity is essential for bacteria and they are promising targets for the design of new antibiotics (Supuran, 2011). γ-CAs represent class of CAs found in archaea and bacteria. They are trimers with the metal ion coordinated between the monomers (Figure 8c). δ and ζ (Figure 8d) classes of CA have been recently identified in marine diatoms and their structure is hypothesized to be monomers, similar to the α-CAs while the active site with Cd or Zn bound resembles the one of β-CAs. The η-, and θ- CAs are the most recently identified families (Supuran, 2016). The η-class has been identified in malaria pathogen *Plasmodium falciparum* (Del Prete et al., 2014) while the second one was identified in *Phaeodactylum tricornutum* (Kikutani et al., 2016). Many representatives of all these enzyme classes have been crystallized and characterized in detail, except the δ-CAs. There are almost 1000 CA structures deposited in PDB

and many of these structures represent the α-class, often in complex with various inhibitors (https://www.rcsb.org/pdb; accessed 2019-03-22).

There are 15 mammalian (α-class) CA isoforms expressed and they vary in molecular features, quaternary structure, organ and tissue distribution, subcellular localization, catalytic properties and capability to interact with different ligands. All enzymes of the α-class share the same tertiary structure, but alter in sequence. Overall there is a variable sequence conservation among the 15 isoforms (30 to 80% sequence identity) but they all have a highly conserved active site core. Most of the hCA isoforms (except hCA IV, IX and XII) are monomers and only the transmembrane spanning hCA IX and hCA XII are implied to be involved in cancer pathology (Whittington et al., 2001; Alterio et al., 2009). Kinetic studies show a large range in catalytic efficiency among CA isoforms, due to the rate limiting step in catalysis (Silverman and McKenna, 2007; Supuran et al., 2018; Aggarwal et al., 2013).

Human carbonic anhydrase IX

The analysis of available structures of hCA IX shows a characteristic ten stranded β-sheet surrounded by additional β-strands and seven α-helices. The catalytic Zn sits ~15 Å deep, on the bottom of a conically shaped active site which extends from the surface to the centre of the enzyme. The Zn is tetrahedrally coordinated to three His residues and a Zn-bound water molecule (Liljas et al., 1972). The cavity itself is similarly to other CAs, divided into a hydrophobic and hydrophilic region. One side of the hCA IX active site is lined with hydrophobic amino acids (Leu223, Val253, Val262, Leu266, Leu272, Val274, Leu231, Pro235) while the surface on the other side of the active site leads towards the bulk solvent and is lined with hydrophilic amino acids (Tyr7, Asn62, His64, Asn67, Arg195, Arg197, Asn198, Thr199, Thr200, His201, Ser202, Gln204, Thr206 and Gln225) (Figure 9).

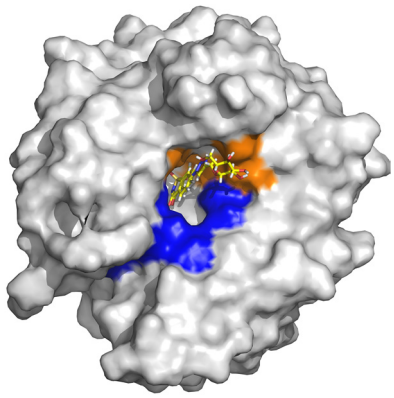


Figure 9: Surface representation of catalytic domain of hCA IX with bound inhibitors saccharin (SAC) and saccharin-glucose conjugate (SGC). hCA IX is shown in a grey surface representation, inhibitors in yellow ball-and-stick, hydrophobic and hydrophilic areas of the active site shaded in orange and blue, respectively. Figure was made using *Pymol* (Schrödinger, 2015) and is adopted from paper V in this thesis (Koruza et al., 2019).

It has been suggested that the hydrophobic region assists the entry of CO_2 in nucleophilic attack by the Zn-bound hydroxide ion in the active site (Domsic et al., 2008). On the other side, the hydrophilic environment facilitates the displacement of HCO_3^- and protons along a H-bonded water network that stretches to the proton shuttle residue His64, positioned on the edge of the active site. This histidine, in the protonated form, rotates and flips conformation towards the bulk solvent, releasing the proton into bulk solvent. The speed of proton shuttling determines the catalytic efficiency among hCAs (Mikulski et al., 2013; Michalczyk et al., 2015).

The wild type hCA IX crystalizes as a dimer. Two other dimeric CA isoforms are hCA VI and hCA XII, but they show different dimeric interfaces (Supuran et al., 2018). The availability of structural studies allows hypothesizing about the model of the full-length protein on the cell membrane (Figure 10).

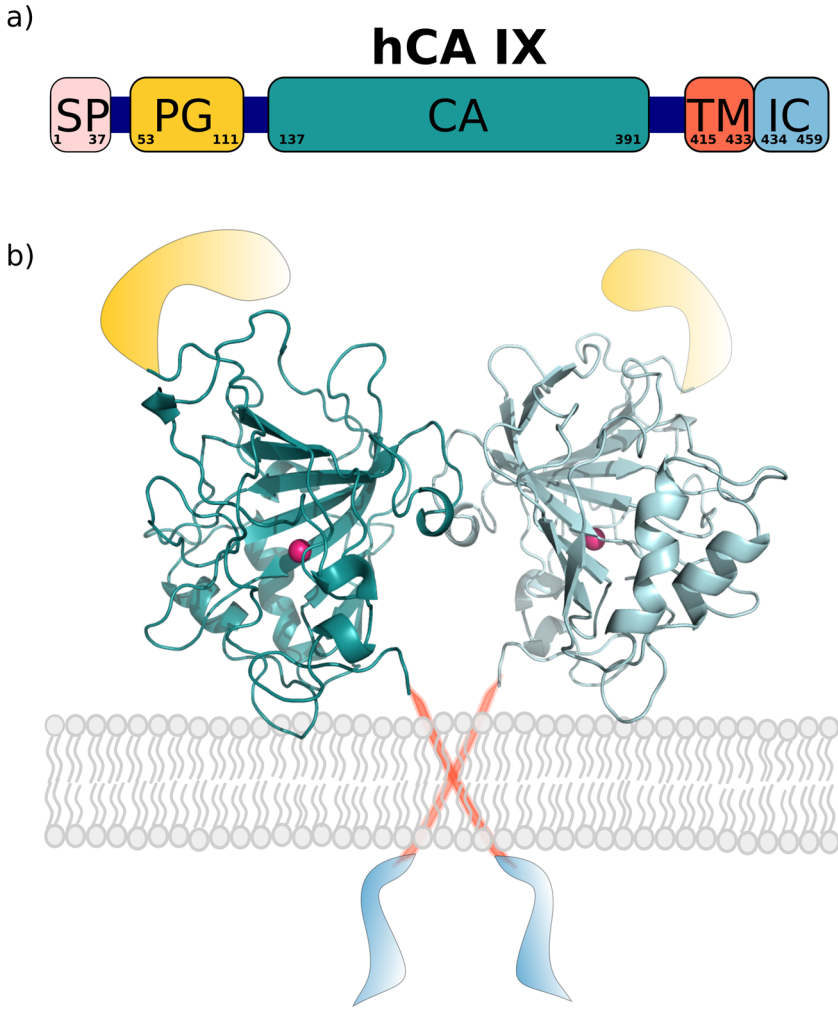


Figure 10: Representation of domain organization of hCA IX and structural arrangement on the cell membrane. (a) Domain organisation of hCA IX with the signal peptide (SP), the proteoglycan domain (PG), the catalytic domain (CA), the transmembrane domain (TM), and the intracellular segment (IC). (b) Schematic model of the structural arrangement of the full-length hCA IX (hCA IX FL) on the cell membrane. Proteoglycan (PG, in yellow), dimer of catalytic domain (hCA, cyan and carmine), a transmembrane anchor domain (orange) and intracellular domain (blue). The Zn ion in the active site of hCA IX is represented as a sphere. The schematic models are adapted from Supuran et al. (Supuran et al., 2018). Strucutre representation of hCA IX FL was made using *Pymol* (Schrödinger, 2015).

The native enzyme is a dimeric trans-membrane protein as modelled in Figure 10 with the catalytic (CA) domain and proteoglycan (PG) domain located on the extracellular face, transmembrane (TM) domain and an intracellular domain (Alterio et al., 2009). Other cytosolic and membrane CAs lack the PG domain which is in hCA IX anchored on the side of the active site cavity with the C-terminal and

presumably allows hCA IX to be metabolically active even at very low pH. Both, the PG and CA domains are glycosylated (Hilvo et al., 2008).

One feature of hCA IX is that it is difficult to express large amounts of soluble protein that would form well-ordered crystals. So far only the catalytic domain of the wild type enzyme was structurally characterized by X-rays (PDB codes 3iai, 5dvx, 5fl4, 5fl6) (Leitans et al., 2015; Alterio et al., 2009; Mahon et al., 2016). Different approaches have been used to overcome the issue of poor crystal formation and among them are mutants that are supposed to facilitate the crystal growth, providing higher yields and more stable protein such as hCA IX_{mimic} (Pinard et al., 2013) and hCA IX SV (Mahon et al., 2016), both also used in this present work.

Pathophysiology of hCA IX in cancer

hCA IX is a very interesting target as a cancer biomarker and potentially for treatment since it is overexpressed under hypoxia in many solid tumours and shows a very limited expression in normal, healthy tissues/organs. In normal human tissues a strong expression of hCA IX is limited to the basolateral surface of proliferating crypt enterocytes of the duodenum, jejunem and ileal mucosa (Supuran et al., 2018). Weak expression has been reported in epithelia of human male efferent ducts, foci in the pancreatic acini and gall bladder epithelia. To date strong expression has been identified in > 30 types of solid tumour, including lung, colon, breast, cervix, bladder, ovarian, brain, head and neck, oral cavity cancers as well as astrocytoma (McDonald et al., 2012). hCA IX presence in solid tumours is associated with metastasis and indicates a poor prognosis (van Kuijk et al., 2016). Besides being an attractive biomarker e.g. for cancer staging, hCA IX is also interesting for cancer treatment. A study in a mouse model for breast cancer focused on the role of hypoxia and CA IX in primary tumour growth and metastasis. The results indicate that under CA IX knock-down, both human and mouse breast tumours display regression and inhibited lung metastasis. Similar to CA IX knock-down, treatment of mice carrying tumours with either sulfonamide- or coumarin-conjugated CA inhibitors resulted in diminished tumour volume, reduced metastasis, and improved significantly the overall survival of mice (Lou et al., 2011).

It is thought that hCA IX helps cancer cells to metastasize through several mechanisms, including changes in cell-cell adhesion by its interaction with β-catenin (Švastová et al., 2003) and the acidification and degradation of the extracellular matrix (McDonald et al., 2012).

Clear cell renal cell cancers that have in 80% of cases a defective Von Hippel–Lindau (VHL) gene, express high levels of hCA IX, also under normoxic conditions (Kats-Ugurlu et al., 2014; Oosterwijk-Wakka et al., 2015). Hypoxia or VHL-loss related pseudohypoxia cause a series of genes to be activated, such as glucose and

ion transporters, growth factors, and hCA IX. These changes cause a shift to a glycolytic metabolism, and in conjunction with the activity of hCA IX, ultimately leads to an acidic extracellular pH, creating a toxic intertumoral microenvironment. In such environment tumour cells have a selective advantage over non-tumour cells (Pastorek and Pastorekova, 2015). External pH is important for maintaining a balanced intracellular pH and proton gradients across the membrane to support several normal cellular functions. Disturbed external pH also causes degradation of the extracellular matrix through protease activation and has documented effects on uptake and efficiency of chemotherapy drugs, resulting in chemoresistance (Vukovic and Tannock, 1997). It has been shown that tumours that express high levels of hCA IX are less responsive to chemo- and radiotherapy and that sensitivity can be restored when the external pH is normalized again. This has been shown in patients who were made to drink large quantities of sodium bicarbonate solution to reduce acidity or when the classic hCA II inhibitor acetazolamide (AZM) is administered (Ibrahim Hashim et al., 2011; Teicher et al., 1993).

All this together is strong evidence that identifying patients with high hCA IX expression and subsequently treating them with hCA IX inhibitors could be beneficial. However, treatment with currently available drugs inhibiting CAs, like AZM, inhibits many of the other isoforms involved in different physiological functions and is therefore causing off-target binding and side-effects. To specifically target only cancers that are hCA IX-positive with isoform-specific ligands, there is a need for new approaches, including structure-guided drug design for small molecules (Pastorek and Pastorekova, 2015; Moeker et al., 2014b) and new concepts of specific binders as for example an acetazolamide peptide conjugate that was selected from a set of polypeptides (Figure 11a) to bind isomers of hCA, described in paper I (Yang et al., 2017; Tegler et al., 2011).

Inhibitors of hCA IX

hCA IX is a promising marker for targeted diagnostic imaging and for systemic anticancer therapy due to its overexpression in tumour tissue and limited presence in normal tissue. In this section diagnostic and therapeutic inhibitors of hCA IX are briefly described.

Diagnostic cancer imaging tools are valuable for cancer diagnosis and staging since hCA IX has broad expression among aggressive and late stage tumours. For this purpose, small molecules and antibodies (Ab) have been developed for *ex vivo* and *in vivo* targeting of tumours, sometimes also as radiolabelled or fluorescent versions of already existing therapeutic modalities (Ahlskog et al., 2009; McDonald et al., 2012). An example of an imaging tool is the highly specific mAb M75 (Oosterwdk et al., 1986), that targets the PG domain of hCA IX and is used for

immunohistochemical detection of hCA IX while its radiolabelled derivatives are used for imaging hCA IX in hypoxic tumours (Chrastina et al., 2003).

So far three major tools have been explored for their efficacy as hCA IX specific therapeutic modalities in anticancer/antimetastatic treatment: small molecule inhibitors, monoclonal Ab (biologics) and small interfering RNA.

Small molecule inhibitors

Similarities in both amino acid sequence and structure, especially in the active site of the α-class CAs leads to challenges in design of isoform selective CA inhibitors (CAIs). Inhibitors that are not highly specific are targeting multiple isoforms, and decrease inhibitor effectiveness or even cause side effects. Therefore, the design of new inhibitors explores limited differences in active sites, interactions with residues out of the active site and reduced cell permeability to limit targeting of cytosolic CAs (Pinard et al., 2015). A typical small molecule inhibitor has a Zn-binding group, a linker region and a variable "tail" region (Supuran, 2012). Based on the Zn-binding coordination, the inhibitors can be divided to those that form tetrahedral adducts and interact directly with the zinc (e.g., sulfonamides and bisulfites), and those that form trigonal-bipyramidal adducts through binding to the zinc-bound hydroxyl/water (e.g., cyanates and formates) (Supuran, 2012; Supuran, 2017; Supuran et al., 2018). Small molecule inhibitors are a topic of great interest which has been reviewed extensively in the last years (Supuran et al., 2018; Lomelino et al., 2016; Mboge et al., 2017; Mboge et al., 2015; McKenna and Supuran, 2014).

Sulfonamides based inhibitors are the most studied CAI, also referred to as "classical" inhibitors and include drugs such as acetazolamide, brinzolamide, dorzolamide etc. Sulfonamides and their isosters bind in deprotonated forms to the Zn, displacing the Zn-bound OH/H_2O while maintaining a tetrahedral coordination on the active site. These first generation inhibitors, that bind deep within the active site, have been used for treating glaucoma and epilepsy and recently cancer, but they rarely exhibit selective inhibition for a single isoform (McKenna and Supuran, 2014; Supuran, 2008).

However, some CAI show higher affinity for CA IX over CA XII or CA II, for example the cyclic secondary sulfonamide saccharin (SAC) (Figure 9 and Figure 11) (Mahon et al., 2015b; Koruza et al., 2019). SAC is more known as artificial sweetener, but it binds to CA IX with nanomolar affinity. It uses both the sulfonamide and lactame carbonyl group to bind to Zn (Köhler et al., 2007). Based on this observation, a series of SAC-based derivatives was tested for CA IX selective inhibition and the best one exhibited > 1000-fold selectivity over other CA isoforms (Moeker et al., 2014b). A SAC-derivative that was produced by addition of a triazole linker and sugar to the SAC (SAC-glucose conjugate or SGC) showed remarkable properties in terms of hCA IX inhibition and selectivity. Beside the

interaction with Zn, the SGC interacts with active site cleft which contributes to isoform-specificity ((Moeker et al., 2014a), and also discussed in paper V (Koruza et al., 2019)).

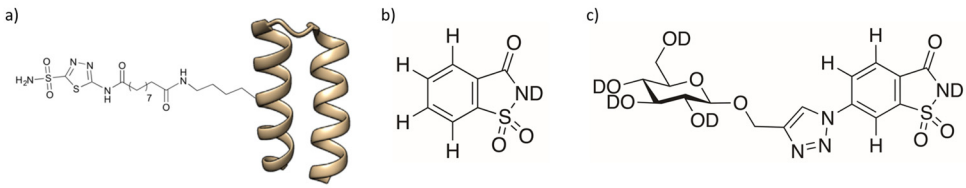


Figure 11: Structure formulae of inhibitors used in this study. (a) Acetazolamide-polypeptide conjugate. (b) Saccharin (SAC) and (c) Saccharing-glucose conjugate (SGC). Adapted from paper V and I (Koruza et al., 2019; Yang et al., 2017).

SAC shows 60-fold specificity for hCA IX (K_i = 100 nM) over hCA II (K_i = 60000 nM), but based on previous studies it was not possible to ascertain which interaction provide isoform selectivity. Better binding selectivity of SAC-glucose conjugate (SGC) was reported for hCA IX (K_i = 50 nM) where addition of sugar also decreased inhibition against hCA II almost 10-fold (SGC K_i > 50000 nM and SAC K_i = 6000 nM) (Moeker et al., 2014a; Moeker et al., 2014b)

The "non-classical" CAI include all the other small molecule inhibitors and biologics, including Ab, RNAi and peptide conjugates, like before mentioned polypeptide conjugate (Tegler et al., 2011). The small molecule inhibitors that do not bind directly to the Zn include thiocarbonates, phenols, coumarins, and polyamines (Supuran, 2008). Coumarins and their derivatives bind at the entrance of the active site cavity, while some carbohydrate based sulfonamide derivatives (Figure 9 and Figure 11c) bind to the Zn but reach out with additional mono- or disaccharide moieties (Mboge et al., 2015; Supuran et al., 2018). Some secondary sulfonamides, ethers and other compound classes inhibit CAs by still unknown mechanisms (Supuran et al., 2018).

The carbohydrate-based sulfonamide inhibitors show high affinity for hCA IX/hCA XII and their bulky sugar moieties cause reduced membrane permeability, which is favourable for targeting the extracellular membrane associated isoforms. Additionally, the sugar moieties retain water solubility and good bioavailability (Lomelino et al., 2018; Lomelino et al., 2016; Mahon et al., 2015c).

Monoclonal antibodies (biologics)

Therapeutic efficacy of hCA IX-specific monoclonal Ab (mAbs) derives either from (a) antibody-mediated cell cytotoxicity (ADCC), (b) direct binding of mAb to the active site or (c) as receptor mediated internalization for targeted delivery of

therapeutic agents (cytotoxins, radionuclides) (McDonald et al., 2012). mAb G250 and its derivatives (one of them is chimeric version cG250) have been studied extensively in clinical trials for use in cancer therapy. The highly selective mAb G250 acts trough ADCC, it targets PG domain and does not cross-react with hCA I, II and XII isoforms (Oosterwdk et al., 1986; Zatovicova et al., 2014). Additionally, it was used for Ab-based targeting systems to deliver cytotoxic agents or radiolabeled Ab to tumour cells in many preclinical targeting studies (Mboge et al., 2015; Supuran et al., 2018). The chimeric cG250 Ab (RENCAREX) entered phase III of clinical trials as an adjuvant therapy but the results showed that the Ab did not meet its primary end point, but a more detailed analysis showed prolonged disease-free survival. Another cG250 derivative that was investigated in preclinical and clinical studies for the treatment of metastatic clear cell renal cell carcinoma is Leutetium-^{177}-cG250/Girentuximab (Muselaers et al., 2016).

Ab that are targeting the CA domain of hCA IX have an advantage of possibly disrupting the catalytic activity of the enzyme and induce endocytosis. An example of such an Ab is Mab VII/20 that was shown to induce a good receptor-mediated internalization and cause significant anticancer effect in mouse xenograft model of colorectal cancer (Zatovicova et al., 2010). Phage display technology has been utilized for the development of novel fully human mAb (A3 and CC7) with high affinity to extracellular domain of hCA IX (Ahlskog et al., 2009).

Small interfering RNA

Researchers have utilized small interfering RNA (RNAi) technology to modulate hCA IX expression in cancers (Lou et al., 2011). The mechanism uses the endogenous miRNA pathway to regulate protein expression by gene silencing through post-transcriptional mechanisms. Ultimately, this activates translational repression and induces mRNA cleavage or degradation (Banerjee and Slack, 2002). The efficient siRNA is complementary to mRNA and induce complete degradation of CA IX expression. In a study by Lou *et al.* the knockdown of CA IX expression by shRNA in the 4T1 mouse metastatic breast cancer cells resulted in regression of tumour size and reduced metastasis formation. The same study also shows stable depletion of CA IX, inhibited tumour growth and metastasis progression in human breast cancer cells when implanted into mouse mammary tissue (Lou et al., 2011). Successful knockdown of CA IX by siRNA or shRNA mediation has been shown to enhance the effects of commonly used inhibitors in several types of cancers (Mahon et al., 2015a; Hilvo et al., 2008), but more research is needed to evaluate the use of RNAi technology for therapy (Mboge et al., 2015).

What is the best approach to inhibit CAs? Over the last years, there has been a significant increase in focus on the development CAIs for the treatment of multiple diseases. The classes of classical and non-classical inhibitors show strong potential as lead compounds for isoform-specific drug design. However the use of biologics

like proteins and peptides in medical treatments of various diseases is increasing and between 2015 and 2018, 30% of all drugs approved by the US Food and Drug Administration were biologics (Anselmo et al., 2019). With several clinical trials ongoing a similar trend is expected in case of CA IX. However, poor bioavailability via the oral route, susceptibility to degradation, large molecular size and variable dose range, the biologics require invasive injection, while small-molecule inhibitors can be administrated via oral delivery in a non-invasive way. Oral delivery of medicines is the most widely used route, it is convenient and well accepted in comparison with intravenous administration that requires medical personnel and is in general painful (Anselmo et al., 2019; Hamman et al., 2005). However, at the end of the day, only the outcome of clinical trials can and will decide what type of CA inhibitors will prove to be successful.

Constructs used in this study

hCA IX$_{mimic}$

As a multi-domain, post-translationally modified membrane protein, hCA IX is very difficult to obtain and to crystallize and there are only few reports of medium resolution X-ray crystal structures of the catalytic domain of hCA IX. We are working with a hCA IX$_{mimic}$ (Figure 12c) to avoid these difficulties for neutron studies. hCA IX$_{mimic}$ was designed by introduction of seven mutations (Ala65Ser, Asn67Gln, Glu69Thr, Ile91Leu, Phe131Val, Lys170Glu, and Leu204Ala) into hCA II (Figure 12b), to exactly match the active site residues in hCA IX, while retaining positive hCA II properties such as high solubility, ease of expression in *E. coli,* easy purification, and well-defined crystallization conditions (Pinard et al., 2013). This gives an enzyme with an active site that looks similar to the native hCA IX and is the basis for the determination of neutron structures in complex with small molecular inhibitors and was used in papers II and V in this thesis.

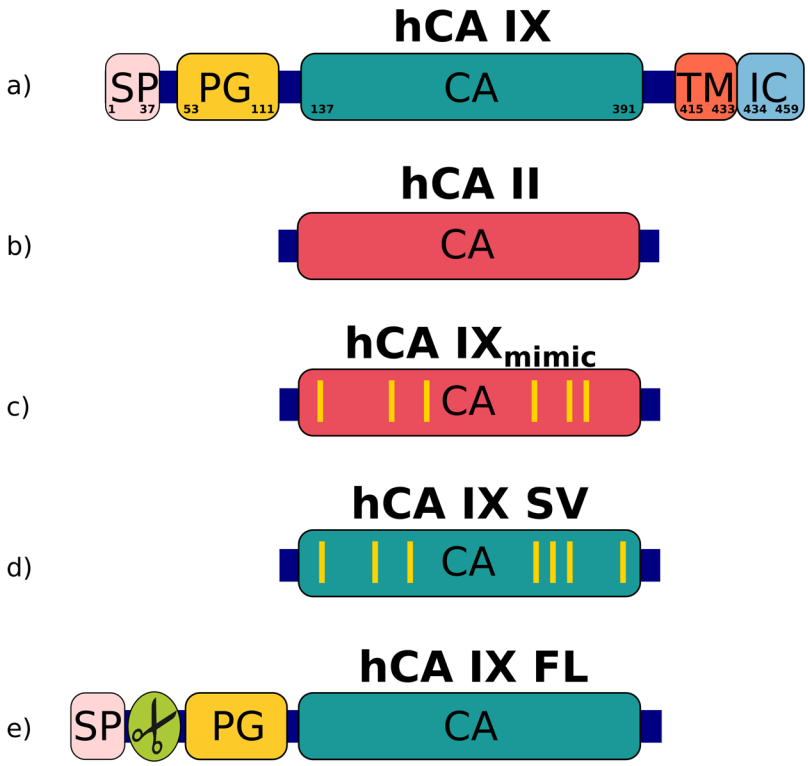


Figure 12: Schematic representation of domain organization of CA constructs used in this study. (a) Schematic representation of the full length native hCA IX with the signal peptide (SP), the proteoglycan domain (PG), the catalytic domain (CA), the transmembrane domain (TM), and the intracellular segment (IC). The numbers of the first and last amino acid of each domain are shown. The schematic model is adapted from Supuran et al. (Supuran et al., 2018). (b) Schematic representation of hCA II. (c) Schematic representation of hCA IX_{mimic} with seven mutations in the active site. (d) Schematic representation of hCA IX SV, with six surface mutations. (e) Schematic representation of hCA IX FL, with a His-tag and thrombin cleavage site presented as green circle with a scissor.

hCA IX SV

The "surface variant" of hCA IX was first described by Mahon et al. (Mahon et al., 2016). It is a wild-type version of the catalytic hCA IX domain that was engineered to have 6 surface mutations (Figure 12d), that provide a more soluble and stable enzyme suitable for expression in *E. coli*, and also to be more amenable to crystallization. It is important to note that hCA IX SV originates from the catalytic domain (CA) of native hCA IX (Figure 12a). The mutated residues are Cys174Ser, Leu180Ser, Ala210Lys, Ala258Lys, Phe259Tyr, and Met360Ser. Part of an effort to grow large, deuterated crystals of hCA IX SV for future neutron experiments is summarized in paper II and IV, attached to this thesis (Koruza et al., 2018a; Koruza et al., 2018b).

hCA IX FL

"hCA IX FL" is a short name for full extracellular part of hCA IX (Figure 12e). Expression of hCA IX FL using the baculovirus expression vector system (BEVS) was modified after the methods described previously (Alterio et al., 2009; Hilvo et al., 2008). The expression construct for the full extracellular domain of hCA IX FL (signal peptide, PG domain and CA catalytic domain) with an 8xHis-tag and thrombin-cleavage site that was inserted between the signal peptide and PG domain was subcloned into pVL1393. hCA IX FL expression was done in Sf9 insect cells, as described in paper I (Yang et al., 2017).

Section II: Outcome and discussion of the study

In manuscript I "*Improved molecular recognition of Carbonic Anhydrase IX by polypeptide conjugation to acetazolamide*" we evaluate the improvement of acetazolamide (AZM) inhibition of hCA IX when conjugated to a set of polypeptides. AZM is a clinically used inhibitor that does not discriminate strongly between CAs. Conjugating AZM to a polypeptide is an intermediate approach between biologics and small molecular inhibitors. Taking into account that biologics are very expensive to develop, this approach features a design that is easier when it comes to chemical synthesis and testing of potential targets. In this article, I present the expression of hCA IX FL, hCA IX SV and hCA IX_{mimic}. Of these constructs we used hCA IX SV and hCA IX_{mimic} in the later studies.

Meantime, we managed to crystallize hCA IX FL and even obtain well diffracting crystals with the best inhibitor of this study bound. The project described in manuscript I is therefore the foundation for an early stage manuscript (not included in this thesis) that will reveal the third structure of the extracellular part of hCA IX, but with details about novel inhibitor binding. While further evaluation of the co-structure is ongoing, we expect it to deliver insight into the molecular background of the increased affinity and specificity that coupling of AZM to the selected peptide yielded.

In manuscript II "*Deuteration of human carbonic anhydrase for neutron crystallography: cell culture media, protein thermostability, and crystallization behavior"* we address the challenges of expressing proteins in bacteria under deuterated conditions. Technical difficulties, such as cellular toxicity that results in reduced yields of recombinant protein and the high cost of deuterated growth media for the production, makes it challenging to produce enough deuterated protein for NPX studies. Furthermore, there are effects of deuteration on protein stability and solubility. In almost all reported cases deuteration causes a reduction in thermal stability and solubility and consistently leads to smaller crystal sizes, compared to what is possible with the protiated versions.

Our study was designed to address some of the technical difficulties of protein deuteration *in vivo* in bacteria and to study the biophysical effects with regards to thermal stability and crystallization behaviour of the products. We started with a survey of the literature and the global protein structure database (https://www.rcsb.org/) to reveal the most common approaches used for deuteration of proteins in NPX. Our goal was to prepare partially deuterated samples, in a short time and with the best possible yield-to-cost ratio.

We combined approaches that still gave very good deuterium incorporation with optimal protein yields, but also significantly reduced the time and cost for producing the proteins compared to established protocols. After producing and isolating the deuterated proteins, we measured the D incorporation with mass spectrometry and tested thermal unfolding as a function of pH to characterize protein stability and solubility. Finally, we tried to grow crystals of deuterated and non-deuterated protein side-by-side to assess the impact of deuteration on crystallization behaviour. The result of the study was a cost-effective optimization of the cell culturing and expression protocols used for production of three constructs of hCAs (hCA II, hCA IX_{mimic} and hCA SV). Our analysis showed that the deuterated proteins were on average 1 – 4 °C less stable than the protiated versions and this effect was strongly pH-dependent. Interestingly, deuterated proteins dissolved in D_2O were more stable than deuterated proteins in H_2O. However, our partial deuteration strategy did not seem to affect protein solubility in the concentrations necessary for crystallization. Comparing crystallization trials of deuterated and non-deuterated proteins at the same pH indicated that deuteration had significant effects on crystallizability. In summary, our protein yield was ~2-fold lower than in non-deuterated conditions, but still producing tens of milligrams of pure protein, yields sufficient for NPX. We managed to get ~65-80% D incorporation, also sufficient for NPX applications, at a cost that was reduced by ~4-fold compared to previously reported protocols.

X-ray structures of crystals from different deuteration stages presented in this project should be next in line for neutron crystallography studies and are also the base for another early stage manuscript that will compare effect of deuteration on the molecular level. Above-mentioned manuscript is not included in this thesis.

In manuscript III, with the working title "*With A Little Help from Algae: Deuteration Of Human Carbonic Anhydrase for Neutron Crystallography*" or "*Development of new methodologies to express deuterated proteins in algal-derived media*" we explored the possibilities for deuteration of proteins by supplementing the minimal growth media with a fully perdeuterated extract from algae. This experiment complements the manuscript II on deuteration, characterization and crystallization of CA IX SV and remains at the time of writing uncompleted. The research question was how much can this boost protein productivity and deuteration grade and at what

cost. Deuterated proteins and other cell components are expensive to produce due to the high cost of deuterated reagents. The price of commercially available single compounds to make deuterated media is high (e.g. 1 L of D_2O costs ≈ 1250 €, https://www.sigmaaldrich.com – accessed on 2019-03-20), therefore alternatives have been continuously explored.

Researchers in deuteration laboratories have different approaches to deuteration, mostly based on local experience and often tradition. One of them is to progressively adapt cells to gradually higher concentrations of D_2O in minimal media with D-labeled carbon source and salts, and once the cell growth is stabile in 100% deuterated media the cells are used for expression (Haertlein et al., 2016; Duff et al., 2015; Paliy et al., 2003). We wanted to avoid this step since cell growth in D_2O is affected by isotope effects and it results in slower growth and low yields. This needs to be compensated with larger culture volumes, which increases expenses and time required to prepare, grow and process larger cultures. An alternative approach is to use rich media that is based on lysed deuterated algal cells as a source of isotopically labeled metabolites and is also available commercially. The algal cells lysate is used as the main component of the media prepared in D_2O and only supplemented with low concentrations of salts. The basic idea of the experiments was to test and compare lysates from two photosynthetic microalgae *Botryococcus braunii* (*B. braunii*) and *Scenedesmus obliquus* (*S. obliquus*) and to determine its performance as a component of growth media in expression trials of CA IX SV. This exploratory study was conducted as part of Japan Society for the Promotion of Science (JSPS) Short Term Fellowship for Research in Japan on a research visit to the laboratories of Comprehensive Research Organization for Science and Society (CROSS) and National Institutes for Quantum and Radiological Science and Technology (QST). Thereby practices from two laboratories that support deuteration at Japan Proton Accelerator Research Complex (J-PARC) in Tokai, Japan, and Lund Protein Production Platform (LP3) and Deuteration and Macromolecular Crystallisation (DEMAX) that support deuteration at European Spallation Source (ESS) were taken into account.

Lysates were prepared from algae grown "in-house" in minimal deuterated media by using CO_2 as inorganic carbon and light energy. Another idea behind this experiment is of interest for neutron experiments like neutron reflectometry and SANS. The deuterated algal biomass can be valuable source of deuterated fatty acids and lipids, prior to preparation of the extracts for protein expression. Lipids do not have significant effect on the growth of *E. coli* so in this case we actually get two products out of algal biomass for the price of one.

Algal extracts of *B. braunii* and *S. obliquus* were prepared by using autolysis and hydrolysis. They were chemically analysed for protein and carbohydrate content and used for expression trials. Yields of extracts itself and protein were compared and

used for cost estimation. In order to bring the cost of the cultivation of algae in D_2O to as low as possible, the D_2O was recycled via rotary evaporation. This process provides some savings but eventually the D_2O becomes protiated via exchange with air and addition of protiated salt solutions. The final step of these experiments will be determination of incorporated deuterium by mass spectroscopy. We expect to achieve high yields of protein at high deuteration levels when using both deuterated algal lysate and D_2O, but unfortunately the results of MS analysis are not available yet.

In the paper IV "*From initial hit to crystal optimization with microseeding of human carbonic anhydrase IX—A case study for neutron protein crystallography*" we discuss crystallisation optimisation for large crystal growth. The construct used in this manuscript was hCA IX SV that proved to be a challenging protein to work with. In four years of working with this protein only a handful of large crystals was grown. Upon observation that hCA IX SV crystalizes in several different conditions, at various pH and salt concentration and different crystallization set-ups, when using microseed matrix screening (MMS), we decided to present the crystallization protocols and strategies that resulted in consistent crystallization.

Continuation of work presented in this paper resulted in a large single crystal (1.36 mm^3) that was tested at the Japan Proton Accelerator Research Complex (J-PARC) in Tokai, Japan and diffracted to 2.7 Å, but so far, the data was not collected. The neutron experiment at the Materials and Life Science Experimental Facility of the J-PARC was performed under a user program (Proposal No. 2019PX0020).

In the paper V "*Using neutron crystallography to elucidate the basis of selective inhibition of carbonic anhydrase by saccharin and a derivative*" we describe jointly refined X-ray and neutron crystal structures of hCA IX_{mimic} alone and in complex with two specific inhibitors – SAC and SGC with the aim of explaining isoform specificity of these two compounds.

The H-bonded network and a water molecule named "deep water" in the active site of unbound CA $IX_{mimic,}$ was corresponding to previously described CA II structures (Genis et al., 2009; Fisher et al., 2011; Domsic et al., 2008). The highly ordered H-bonded water network in the active site was mapped in details based on the positions of D atoms on the H-bonded water molecules. Four possibilities for a proton to leave the active site from Zn -bound water to the "proton shuttling" residue His64 on the edge of the active site were mapped in details.

SAC binds to the Zn similar to a sulfonamide-based CAI. Upon binding, it displaces four water molecules and results in net loss of six H-bonds between water and protein, while seven H-bonds are retained "as is" or reorganised to accommodate

SAC. In comparison to unbound CA IX_{mimic}, the Gln67 side chain moves but maintains the H-bond.

The SAC moiety of SGC binds in slightly different orientation than SAC alone, but it is still coordinated to the Zn. The linker is not involved in any H-bonds, while SGC makes three indirect interaction for a total of ten H-bonds to amino acid residues. Gln67 maintains the same configuration as in the structure with SAC and appears to be involved in water-mediated H-bond with glucose.

Modest 2-fold difference in K_i of 100 nM against 50 nM for SAC and SGC towards hCA IX could be explained by the presence of the linker and interactions that glucose may make to the protein. Additionally, there is a possibility for Arg130 in native hCA IX to make a direct H-bond to SGC, but in hCA IX_{mimic} the Arg130 is Asp, therefore a H-bond is not observed in our model. However, the isoform selectivity of SAC for CA IX over CA II is difficult to interpret as only Gln67 and some water positions are different between the two isoforms. As it may be, this difference is strong enough for the 60-fold difference in K_i or there are some other effects not observable in static crystal structures. The information about H-bonds provided insights to specific interactions of binding and selectivity and gives clues on how to improve SGC. Slight modification of SGC could displace one of the waters and directly H-bond to Gln67. The linker that is situated close to the hydrophobic region of the active site could be methylated or replaced with a hydrophobic moiety to reinforce hydrophobic interactions. And finally, the disordered glucose indicates that there is enough space for bulkier groups that could fill this area.

Section III: Conclusions and future perspectives

The work in this thesis provides examples for protein deuteration and supports the use of neutrons to characterise numerous features of ligand binding to protein that X-rays alone cannot reveal.

1) Deuteration

In the study case of hCA IX we demonstrate possibilities of deuteration optimization by culture media, and further explore the protein thermostability and crystallization behaviour for NPX.

We used different constructs, introduced in paper I, but they are all CAs. The next step would be to apply the methods developed by us to other proteins of interest and test cases. This work could, in particular for the algae extracts, also include protein deuteration with the help of yeast (*P. pastoris*) and higher eukaryotic expression systems.

In addition, further understanding of the background and molecular changes happening in organisms when used for *in vivo* deuteration is still pending. This would be another valuable piece of information since all neutron studies require deuterated organic compounds in addition to proteins, such as membrane phospholipids, sterols, short peptides and biological detergents. Genetic engineering to optimise organisms for use as deuteration hosts would be another challenge for the future.

2) Crystallisation and structure determination

So far, we determined the structure of hCA IX_{mimic} and solved all the steps of joint X-ray and neutron structure refinement. Only recently we achieved crystallisation of hCA IX SV (Figure 13a) and "full-length" hCA IX FL (Figure 13 b), and after data collection we are now in the process of final X-ray structure refinement. The next step would be to develop these systems for neutron crystallography, meaning we need to increase the yield of perdeuterated protein and grow bigger crystals.

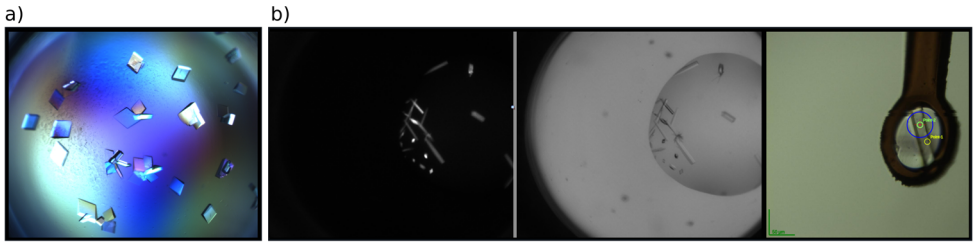


Figure 13: Crystals of hCA IX SV (a) and crystals of hCA IX FL: UV signal from the crystals, crystals in the crystalization drop and harvested crystal at the synchrotron beamline BioMAX (b).

3) Protein-inhibitor binding

To date we have solved the neutron structures of hCA IX_{mimic} alone and with two inhibitors. This showed interesting features regarding H-bonding and revealed certain amino acids as important for specific inhibitor binding. It would be important to verify these binding properties with the hCA IX FL construct. In addition, we would like to study the charged state of another inhibitor (Figure 14), that has several orders of magnitude higher selectivity for CA IX (K_i = 11 nM) over CA II (K_i= > 20000 nM) and poor cell membrane passive diffusion characteristics, a property that eliminates interactions with cytosolic CAs.

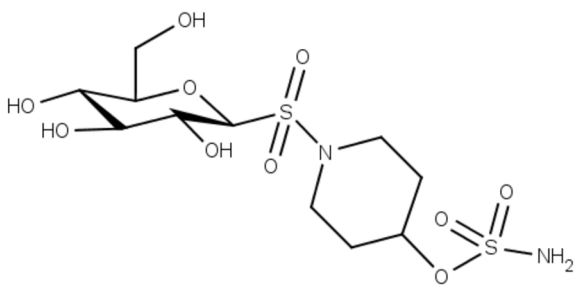


Figure 14: Highly specific inhibitor “5d” with sulfamate, linker molecule and a carbohydrate tail moiety. Adapted from (Moeker et al., 2014a).

Together with all other results this should give enough input into generation of new inhibitors to test. Additionally, we recently even obtained well diffracting crystals with the best inhibitor of paper I (polypeptide conjugate to AZM) and the structure analysis and refinement are ongoing. Besides completing the structure and for the first time describe the binding of such a hybrid-peptide inhibitor to a target by X-ray crystallography, the testing of this inhibitor in cell culture models would be a desired next step.

In the long run, I expect my project to have impact in two main areas:

a) Better methods to produce (per)deuterated bioreagents, which will open up the field for more researchers to use neutron scattering for life sciences.

b) Contributing to the process of ultimately finding new isoform specific inhibitors of hCA IX that might be tested in the clinic and that have a positive effect on the treatment of solid tumours.

Popular science summary

Proteins are one of foundations of a living cell's structure and function. They are for example building the cell skeleton, transporting essential molecules over the membranes, and transmitting signals to coordinate biological processes. One of the most prominent properties of proteins is that they act as enzymes and catalyse nearly all the chemical reactions in biological systems.

X-ray crystallography remains the primary tool in structural biology that provides information about structure and function of proteins involved in basic cellular functions or in pathology of disease, and it is also a critical tool used in drug discovery. In a typical X-ray experiment a well-ordered crystal of proteins is exposed to highly coherent and intense X-rays, and the diffraction of the beam is used to determine the protein structure. If the protein used in this method is an enzyme in complex with a small molecule, for example an inhibitor, it is possible to identify the structure of the ligand within the protein and also how they interact. In the case of structure-based drug discovery, this information can then help a medicinal chemist with further chemical optimisation of the inhibitor, and modifying it in an iterative process to finally develop a starting molecule into a drug that can be tested in clinical trials.

However, hydrogen (H) atoms - that account for approximately half of the atoms in proteins - play a crucial role in protein folding, enzyme catalysis, and ligand (substrate, product, or inhibitor) binding interactions, will remain "invisible" with X-rays. Even at ultra high-resolution structures it is very challenging to determine H positions using X-ray crystallography alone.

The difficulty of locating H atom can be circumvented by neutron protein crystallography, that offers a unique and complementary approach for locating H atoms by using the intrinsic strong neutron scattering properties of H, and also that of its isotope, deuterium (D). Due to the strong neutron scattering of D, it is possible to explicitly observe features involved with ligand binding to proteins, such as: protonation and H-bonding of amino acid residues; how water molecules interact with each other and the protein; the charged state of the ligand itself and how water and H-bonds modulate binding; the possibility to discern solvent species (H^+ vs OH^- vs H_2O vs H_3O^+). Deuterated biomolecules - where H are exchanged by D - are essential for biological experiments in neutron scattering and we have worked on improving methods to produce them. Deuteration is expensive and when it needs to

be done in living cells – e.g. bacteria, it is laborious and results in low yields. In the methods we developed, we simplified methodology, worked on improvement of yields of protein and deuteration grade, and made them as cost-effective as possible.

As a model for this study we selected human Carbonic anhydrase IX (hCA IX). The chemical reaction that describes any Carbonic anhydrase (CA) activity is very basic: carbon dioxide and water are rapidly converted to bicarbonate and a proton ($CO_2 + H_2O \rightarrow HCO_3^- + H^+$). In our human bodies there are 15 different types (isoforms) of the enzyme in different forms, depending on the tissue or cellular compartment they are located in. They can be found in tissues like lungs, kidney, intestines, muscles, etc. The CA isoforms differ in their amino-acid sequence (30 – 80 % identity) but they all show a high sequence conservation in their active site in which catalysis takes place. Human CAs are involved with nearly every physiological process in our bodies: respiration, acid/base homeostasis, gluconeogenesis, ureagenesis, bone resorption, and others. Most of the hCA isoforms (except hCA IV, IX and XII) are monomers and three hCA isoforms (hCA IX, XII, and XIV) have the transmembrane spanning domain.

hCA IX is usually not present in large amounts in healthy tissues, but it has been identified in more than 30 types of cancer to date. Its presence in solid tumours is associated with metastasis and indicates a poor prognosis of survival for the cancer patient. As such, hCA IX is an attractive target in cancer research as it could help with diagnosis of cancer progression, and possibly cancer therapy. There is strong evidence that identifying patients with high hCA IX expression, and subsequently treating them with hCA IX inhibitors, could be beneficial. However, treatment with currently available drugs that have been developed for other CAs e.g., the drug acetazolamide, inhibits many of the other hCAs, causing off-target binding and side-effects. To specifically target only cancers that are hCA IX-positive with isoform-specific inhibitors, there is a need for new approaches, including optimization of structure-based drug design.

We have solved joint X-ray and neutron structures of hCA IX_{mimic} (hCA II with 7 mutations that "mimic" the active site of hCA IX) alone and in complex with two specific inhibitors – Saccharin and a Saccharin-glucose conjugate. The structures showed interesting features regarding H-bonding and revealed certain amino acids as important for specific inhibitor binding interactions.

We also explored a new concept of enzyme inhibitors in which a small-molecule enzyme inhibitor is conjugated to a larger polypeptide molecule to increase affinity and specificity of the inhibitor. By testing several of these hybrid-peptides against hCA IX we found one with significantly increased affinity and specificity for hCA IX when compared to the small-molecule inhibitor alone.

Populärvetenskaplig sammanfattning

Proteiner utgör grunden för en levande cells struktur och funktion. De till exempel bygger upp cellskelettet, transporterar väsentliga molekyler över membranen och sänder signaler för att samordna biologiska processer. En av de mest framträdande egenskaperna hos proteiner är att de fungerar som enzymer och katalyserar nästan alla kemiska reaktioner i biologiska system.

Röntgenkristallografi är fortfarande det primära verktyget inom strukturbiologi som ger information om struktur och därmed också funktion av proteiner som är involverade i sjukdomspatologi, och det är också ett kritiskt verktyg för utveckling av läkemedel. I ett typiskt röntgenexperiment exponeras en kristall av proteiner för röntgenstrålar, och reflektionsmönstret från kristallen ger i slutändan en bild av hur proteinet ser ut. Om proteinet i kristallen har bundit en liten molekyl, exempelvis en inhibitor, är det möjligt att identifiera hur denna bindning ser ut och också hur de interagerar. Denna information kan sedan hjälpa en läkemedelskemist att ytterligare kemiskt optimera inhibitorn så att den binder starkare och på så vis, i en iterativ process, utveckla en molekyl som kan testas i kliniska prövningar och senare bli ett läkemedel.

Väteatomer (H) - som utgör ungefär hälften av atomerna i proteiner - spelar en avgörande roll vid veckning av proteiner, enzymkatalyser och interaktioner med ligander (substrat, produkt eller inhibitor). Dessa atomer förblir "osynliga" när man använder röntgenstrålar för att få strukturen av ett protein. Även vid väldigt hög upplösning, då man ser många olika atomer i detalj i proteinstrukturen, är det mycket utmanande att bestämma just positionerna för väteatomerna med endast röntgenkristallografi som analysmetod.

Neutronproteinkristallografi erbjuder ett unikt och komplementärt sätt att lokalisera H-atomer i proteinstrukturer. I denna teknik används H-atomernas, och även dess isotop deuteriums (tungt väte, D), egenskap att sprida neutroner istället för röntgenstrålning. Eftersom dessa atomer sprider neutroner så bra är det möjligt att observera i atomär detalj företeelser under ligandbindning till proteiner, såsom protonering och vätebinding till aminosyror, hur vattenmolekyler interagerar med varandra och proteinet, ligandens egen laddning och hur vatten och vätebindningar påverkar bindningen, samt olika former av väte (H^+, OH^-, H_2O och H_3O^+) och särskilja dem. Deutererade biomolekyler - där H har bytts ut med D - är nödvändiga

för biologiska experiment inom neutronkristallografi och vårt mål med detta arbete är att förbättra metoderna för deuterering.

Deuterering är en kostsam process och görs det i levande celler, till exempel i bakterier, är det påfrestande för cellerna och resulterar i liten mängd av proteinet som ska studeras. Vårt mål med detta arbete var att förenkla metoderna för deuterering, förbättra utbytet av protein, öka deutereringsgraden (hur många procent av väteatomerna som bytts ut till tungt väte) och göra processen så kostnadseffektiv som möjligt.

Vi valde det humana proteinet karbanhydras IX (hKA IX) som modell för denna studie. hKA IX omvandlar koldioxid och vatten till bikarbonat och en proton ($CO_2 + H_2O \rightarrow HCO_3^- + H^+$). I människokroppen finns det 15 olika varianter (isoformer), och dessa är specifika för olika vävnader eller del av cellen de befinner sig i. Man kan finna denna grupp av enzymer i tex lunga, njure, tarm och muskelvävnad, och de är involverade i ett antal fysiologiska processer som andning, syra- och bashomeostas, glukoneogenes och utsöndring av urea. De olika isoformerna skiljer sig åt i aminosyrasekvensen (de har 30-80% identitet sinsemellan) men i det aktiva centrumet, i vilket katalysen äger rum, är sekvenserna för isoformerna mycket konserverad. De flesta hKA (förutom hKA IV, IX och XII) är monomerer och tre hKA (hKA IX, XII och XIV) har en transmembran domän.

hKA IX finns vanligtvis inte i stora mängder i frisk vävnad, och har hittills identifierats i mer än 30 typer av cancer. Dess närvaro i tumörer är förknippad med metastas och indikerar en dålig prognos för cancerpatienten. Detta gör att hKA IX skulle kunna användas vid prognos av cancerförlopp och eventuellt också som markör vid cancerbehandling. Det finns forskning som visar att patienter med högt hKA IX-uttryck, som behandlas med hKA IX-inhibitor, har chans till en bättre prognos. Det finns också tillgängliga läkemedel riktade mot KA isoformer, t.ex. acetazolamid, men de är inte så specifika gentemot KA IX, vilket ger svag inbindning och bieffekter. För att utveckla isoformspecifika inhibitorer bör man inrikta sig mot cancerformer som är hKA-IX-positiva och med fördel använda strukturbaserad läkemedelsdesign.

Vi har löst röntgen- och neutronstrukturer av hKA IX_{mimic} (hKA II med 7 mutationer som ger ett liknande aktivt centrum som hKA IX) dels ensamt, utan någon tillsats, och dels med två specifika inhibitorer - Saccharin och ett Saccharin-glukoskonjugat. Strukturerna avslöjade vilka aminosyror som är grundläggande för den specifika interaktionen samt visade intressanta detaljer angående vätebindningarna vid inbinding av de två inhibitorerna. Vi undersökte också ett nytt koncept av enzyminhibitor: en liten molekyl sammansatt med en större polypeptid för att öka affiniteten och specificitet. Genom att testa flera av dessa hybridpeptider mot hKA IX fann vi en med signifikant ökad affinitet och specificitet jämfört med enbart en liten molekylinhibitor.

Acknowledgements

To begin with I would like to thank my supervisor Wolfgang Knecht and my co-supervisor Zoë Fisher for accepting me on this extraordinary journey! My PhD would not be possible without your guidance and support.

I also wish to thank my examiner Lars Hederstedt, co-supervisor Claes von Wachenfeldt and Torbjörn Säll & Emma Kritzberg for valuable advices and for keeping an eye on my progress as members of my supervisory committee. I thank to Lund University for funding my PhD project. I am grateful for financial support from Royal Physiographic Society of Lund, Jörgen Lindström's foundation, Crafoord Foundation and BioCARE (A strategic research area at Lund University).

I would like to thank European Spallation Source for supporting my PhD project. This fruitful collaboration between Lund University and ESS was additionally supported by Interreg/MAX4ESSFUN, Integrated Infrastructure Initiative No. 262348 European Soft Matter Infrastructure and SINE2020 grant the receives funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 654000. Special thanks to Esko Oksanen for supporting feedback during my midterm defence.

I would like to express my deepest gratitude to Japan Society for the Promotion of Science (JSPS) for funding my "Short Term Fellowship for Research in Japan" and Swedish Foundation for International Cooperation in Research and Higher Education (STINT) for rewarding me the Capstone Award.

I would like to thank our collaborators for technical support and materials. Special thanks to Professor Robert McKenna and Dr. Brian Mahon, for constructs and for contribution to our manuscripts. Thanks also to Dr. Lars Baltzer and Dr. Jie Yang from the Department of Chemistry-BMC, Uppsala University, Sweden, for providing conjugated polypeptide. Thanks to Dr. Ákos Végvári for MS analysis of problematic samples. I also wish to thank to Bénédicte Lafumat, Maria Gourdon and Saima Rehman from Lund Protein Production platform (LP3) for contributing to the manuscripts. Thanks also to Anna Leung and Robin Delhom (ESS) for lipid analysis of all the colourful samples. I am extremely grateful to Olena Ishchuk, Chalmers University of Technology, who provided me with encouragement and patience throughout the duration of my two internships and my PhD project.

Parts of my thesis would not have been possible without neutrons and X-rays. I would like to thank the ILL and ESRF for access to the beamlines LADI and FIP-BM30A; the ESMI for providing beam time at BIODIFF at the Heinz Maier-Leibnitz Zentrum (MLZ); the MAX IV laboratory – BioMAX, & I911-3 and I911-2 Cassiopeia beamline at MAX lab (Lund, Sweden); and finally, to Materials and Life Science Experimental Facility of the Japan Proton Accelerator Research Complex (J-PARC) in Tokai, Japan for access to the bemline iBIX, (user program proposal No. 2019PX0020). Thanks to all beamline scientists and LP3 staff for expert assistance and data collection.

I would also like to extend my deepest gratitude to Dr. Kazuhiro Akutsu from the Neutron Science and Technology Center, Comprehensive Research Organization for Science and Society (CROSS), Tokai, Japan for hosting me in Tokai for three months as part of my JSPS Short Term fellowship. Research in Japan would not have been possible without the support of Dr. Motoyasu Adachi, Dr. Satoru Fujiwara, Chie Shibazaki and Rumi Shimizu from the National Institutes for Quantum and Radiological Science and Technology (QST) (Tokai, Japan). My whole stay in Japan was definitely a wonderful experience and I am very happy to see that our collaboration is continuing.

I have met wonderful people at the Molecular Cell Biology! Thanks for creating great working environment, good luck and many thanks to: LP3 “gäng”, Manuel, Swati, Robin, Ryan, Beer, Humberto, Markus, Elisabeth, Chatarina, Mats, Olivier, Alex, Oscar, Nick, Klas, Mengshu, Marita, Allan, Lars, Colm, people from Jure’s group, Katsuya, Sofia, Fanny and many others! Thanks to Vinardas, Shakira, David and Anna R. for sharing the teaching duties & many thanks to all the participants of pizza workshops, BBQs & running!

Lastly, I would like to thank to my friends from Sweden and Slovenia, my family & Andrea for endless support and encouragement. Thanks to Erika and Paul for being awesome & big thanks to Falcon brotherhood and its admirers. Thank you Ksenia for introducing me and Andrea to the best badminton team ever! You guys made these five years in Sweden so much warmer!

References

Adachi M, Ohhara T, Kurihara K, et al. (2009) Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proceedings of the National Academy of Sciences of the United States of America* 106: 4641-4646.

Adams PD, Afonine PV, Bunkoczi G, et al. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55: 94-106.

Aggarwal M, Boone CD, Kondeti B, et al. (2013) Structural annotation of human carbonic anhydrases. *Journal of Enzyme Inhibition and Medicinal Chemistry* 28: 267-277.

Aggarwal M, Kovalevsky AY, Velazquez H, et al. (2016) Neutron structure of human carbonic anhydrase II in complex with methazolamide: mapping the solvent and hydrogen-bonding patterns of an effective clinical drug. *International Union of Crystallography Journal* 3: 319-325.

Ahlskog JKJ, Schliemann C, Mårlind J, et al. (2009) Human monoclonal antibodies targeting carbonic anhydrase IX for the molecular imaging of hypoxic regions in solid tumours. *British Journal Of Cancer* 101: 645.

Alterio V, Hilvo M, Di Fiore A, et al. (2009) Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX. *Proceedings of the National Academy of Sciences of the United States of America* 106: 16233-16238.

Alterio V, Langella E, Viparelli F, et al. (2012) Structural and inhibition insights into carbonic anhydrase CDCA1 from the marine diatom Thalassiosira weissflogii. *Biochimie* 94: 1232-1241.

Andriukonis E and Gorokhova E. (2017) Kinetic 15N-isotope effects on algal growth. *Scientific Reports* 7: 44181.

Anselmo AC, Gokarn Y and Mitragotri S. (2019) Non-invasive delivery strategies for biologics. *Nature Reviews Drug Discovery* 18: 19-40.

Asherie N. (2004) Protein crystallization and phase diagrams. *Methods* 34: 266-272.

Azadmanesh J, Trickel SR, Weiss KL, et al. (2017) Preliminary neutron diffraction analysis of challenging human manganese superoxide dismutase crystals. *Acta Crystallographica Section F* 73: 235-240.

Banerjee D and Slack F. (2002) Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *BioEssays* 24: 119-129.

Bennett B, Langan P, Coates L, et al. (2006) Neutron diffraction studies of *Escherichia coli* dihydrofolate reductase complexed with methotrexate. *Proceedings of the National Academy of Sciences of the United States of America* 103: 18493-18498.

Bennett BC, Gardberg AS, Blair MD, et al. (2008) On the determinants of amide backbone exchange in proteins: a neutron crystallographic comparative study. *Acta Crystallographica Section D* 64: 764-783.

Berman HM, Bhat TN, Bourne PE, et al. (2000) The Protein Data Bank and the challenge of structural genomics. *Nature Structural Biology* 7: 957.

Berman HM and Burley SK. (2016) *The Protein Data Bank and the Importance of Sustaining Primary Data Archives*. Available at: https://ncip.nci.nih.gov/blog/protein-data-bank-importance-sustaining-primary-data-archives/.

Bernal JD and Crowfoot D. (1934) X-ray photographs of crystalline pepsin. *Nature* 133: 794-795.

Bhatia CR and Smith HH. (1968) Adaptation and Growth Response of Arabidopsis Thaliana to Deuterium. *Planta* 80: 176-184.

Blakeley MP. (2016) Neutron crystallography aids in drug design. *International Union of Crystallography Journal* 3: 296-297.

Blakeley MP, Hasnain SS and Antonyuk SV. (2015) Sub-atomic resolution X-ray crystallography and neutron crystallography: promise, challenges and potential. *International Union of Crystallography Journal* 2: 464-474.

Blakeley MP, Kalb AJ, Helliwell JR, et al. (2004) The 15-K neutron structure of saccharide-free concanavalin A. *Proceedings of the National Academy of Sciences of the United States of America* 101: 16405-16410.

Blundell TL. (2017) Protein crystallography and drug discovery: recollections of knowledge exchange between academia and industry. *International Union of Crystallography Journal* 4: 308-321.

Brockwell D, Yu L, Cooper S, et al. (2001) Physicochemical consequences of the perdeuteriation of glutathione S-transferase from S. japonicum. *Protein Science* 10: 572-580.

Budayova-Spano M, Fisher SZ, Dauvergne M-T, et al. (2006) Production and X-ray crystallographic analysis of fully deuterated human carbonic anhydrase II. *Acta Crystallographica Section F* 62: 6-9.

Cai M, Huang Y, Yang R, et al. (2016) A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *Journal of Biomolecular NMR* 66: 85-91.

Chadwick J. (1932) Possible Existence of a Neutron. *Nature* 129: 312.

Chapman HN, Barty A, Bogan MJ, et al. (2006) Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics* 2: 839.

Chen JC and Unkefer CJ. (2017) Fifteen years of the Protein Crystallography Station: the coming of age of macromolecular neutron crystallography. *International Union of Crystallography Journal* 4: 72-86.

Chrastina A, Závada J, Parkkila S, et al. (2003) Biodistribution and pharmacokinetics of 125I-labeled monoclonal antibody M75 specific for carbonic anhydrase IX, an intrinsic marker of hypoxia, in nude mice xenografted with human colorectal carcinoma. *International Journal of Cancer* 105: 873-881.

Clabbers MTB, van Genderen E, Wan W, et al. (2017) Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Crystallographica Section D* 73: 738-748.

Cooper G. (2000) *The Cell: A Molecular Approach.,* Sunderland (MA): Sinauer Associates.

Covington AK, Paabo M, Robinson RA, et al. (1968) Use of the glass electrode in deuterium oxide and the relation between the standardized pD (paD) scale and the operational pH in heavy water. *Analytical Chemistry* 40: 700-706.

Crespi HL, Archer SM and Katz JJ. (1959) Culture of Algae and other Micro-organisms in Deuterium oxide. *Nature* 184: 729-730.

Crespi HL, Conrad SM, Uphaus RA, et al. (1960) Cultivation of Microorganisms in Heavy Water. *Annals New York Academy of Sciences* 84: 648-666.

Crespo I. (2012) *From Bragg's law to neutron diffraction*. Available at: http://nmi3.eu/news-and-media/from-braggs-law-to-neutron-diffraction.html.

D'Arcy A, Mac Sweeney A and Haber A. (2003) Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Crystallographica Section D* 59: 1343-1346.

Daboll HF, Crespi HL and Katz JJ. (1962) Mass cultivation of algae in pure heavy water. *Biotechnology and Bioengineering* 4: 281-297.

De Giovanni R. (1960) The effects of deuterium oxide on certain microorganisms. *Annals of the New York Academy of Sciences* 84: 644-647.

Del Prete S, Vullo D, Fisher GM, et al. (2014) Discovery of a new family of carbonic anhydrases in the malaria pathogen Plasmodium falciparum - the eta-carbonic anhydrases. *Bioorganic & Medicinal Chemistry Letters* 24: 4389-4396.

Domsic JF, Avvaru BS, Kim CU, et al. (2008) Entrapment of Carbon Dioxide in the Active Site of Carbonic Anhydrase II. *Journal of Biological Chemistry* 283: 30766-30771.

Dubochet J and Mcdowall AW. (1981) Vitrification of Pure Water for Electron-Microscopy. *Journal of Microscopy* 124: 3 - 4.

Duff AP, Wilde KL, Rekas A, et al. (2015) Robust High-Yield Methodologies for 2H and 2H/15N/13C Labeling of Proteins for Structural Investigations Using Neutron Scattering and NMR. *Methods in Enzymology* 565: 3-25.

Dunne O, Weidenhaupt M, Callow P, et al. (2017) Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic expression systems and high cell-density cultures. *European Biophysics Journal* 46: 425-432.

Duyvesteyn HME, Kotecha A, Ginn HM, et al. (2018) Machining protein microcrystals for structure determination by electron diffraction. *Proceedings of the National Academy of Sciences of the United States of America* 115: 9569-9573.

Ferraroni M, Del Prete S, Vullo D, et al. (2015) Crystal structure and kinetic studies of a tetrameric type II beta-carbonic anhydrase from the pathogenic bacterium Vibrio cholerae. *Acta Crystallographica section D* 71: 2449-2456.

Fiaux J, Bertelsen EB, Horwich AL, et al. (2004) Uniform and residue-specific 15N-labeling of proteins on a highly deuterated background. *Journal of Biomolecular NMR* 29: 289-297.

Fisher SZ, Aggarwal M, Kovalevsky AY, et al. (2012) Neutron diffraction of acetazolamide-bound human carbonic anhydrase II reveals atomic details of drug binding. *Journal of the American Chemical Society* 134: 14726-14729.

Fisher SZ, Kovalevsky AY, Domsic JF, et al. (2010) Neutron structure of human carbonic anhydrase II: implications for proton transfer. *Biochemistry* 49: 415-421.

Fisher SZ, Kovalevsky AY, Domsic JF, et al. (2009) Preliminary joint neutron and X-ray crystallographic study of human carbonic anhydrase II. *Acta Crystallographica Section F* 65: 495-498.

Fisher Z, Jackson A, Kovalevsky A, et al. (2017) Biological Structures. In: Felix F-A and David LP (eds) *Experimental Methods in the Physical Sciences.* Academic Press, 1-75.

Fisher Z, Kovalevsky AY, Mustyakimov M, et al. (2011) Neutron Structure of Human Carbonic Anhydrase II: A Hydrogen-Bonded Water Network "Switch" Is Observed between pH 7.8 and 10.0. *Biochemistry* 50: 9421-9423.

Franke B, Opitz C, Isogai S, et al. (2018) Production of isotope-labeled proteins in insect cells for NMR. *Journal of Biomolecular NMR* 71: 173-184.

Gardberg AS, Del Castillo AR, Weiss KL, et al. (2010) Unambiguous determination of H-atom positions: comparing results from neutron and high-resolution X-ray crystallography. *Acta Crystallographica Section D* 66: 558-567.

Garman EF and Weik M. (2017) X-ray radiation damage to biological macromolecules: further insights. *Journal of Synchrotron Radiation* 24: 1-6.

Gavira JA. (2016) Current trends in protein crystallization. *Archives of Biochemistry and Biophysics* 602: 3-11.

Gelin M, Delfosse V, Allemand F, et al. (2015) Combining 'dry' co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. *Acta Crystallographica section D* 71: 1777-1787.

Genis C, Sippel KH, Case N, et al. (2009) Design of a Carbonic Anhydrase IX Active-Site Mimic To Screen Inhibitors for Possible Anticancer Properties. *Biochemistry* 48: 1322-1331.

Gerlits O, Keen DA, Blakeley MP, et al. (2017) Room Temperature Neutron Crystallography of Drug Resistant HIV-1 Protease Uncovers Limitations of X-ray Structural Analysis at 100 K. *Journal of Medicinal Chemistry* 60: 2018-2025.

Golden E, Attwood PV, Duff AP, et al. (2015) Production and characterization of recombinant perdeuterated cholesterol oxidase. *Analytical Biochemistry* 485: 102-108.

Haertlein M, Moulin M, Devos JM, et al. (2016) Biomolecular Deuteration for Neutron Structural Biology and Dynamics. *Methods in Enzymology* 566: 113-157.

Hamman JH, Enslin GM and Kotzé AF. (2005) Oral Delivery of Peptide Drugs. *BioDrugs* 19: 165-177.

Hattori A, Crespi HL and Katz JJ. (1965) Effect of Side-Chain Deuteration on Protein Stability. *Biochemistry* 4: 1213-1225.

Haupt M, Blakeley MP, Fisher SJ, et al. (2014) Binding site asymmetry in human transthyretin: insights from a joint neutron and X-ray crystallographic analysis using perdeuterated protein. *International Union of Crystallography Journal* 1: 429-438.

Hazemann I, Dauvergne MT, Blakeley MP, et al. (2005) High-resolution neutron protein crystallography with radically small crystal volumes: application of perdeuteration to human aldose reductase. *Acta Crystallographica Section D* 61: 1413-1417.

Henderson R. (2015) Overview and future of single particle electron cryomicroscopy. *Archives of Biochemistry and Biophysics* 581: 19-24.

Hilvo M, Baranauskiene L, Salzano AM, et al. (2008) Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes. *Journal of Biological Chemistry* 283: 27799-27809.

Huang GY, Gerlits OO, Blakeley MP, et al. (2014) Neutron diffraction reveals hydrogen bonds critical for cGMP-selective activation: insights for cGMP-dependent protein kinase agonist design. *Biochemistry* 53: 6725-6727.

Ibrahim Hashim A, Cornnell HH, Coelho Ribeiro ML, et al. (2011) Reduction of metastasis using a non-volatile buffer. *Clinical & Experimental Metastasis* 28: 841-849.

Iverson TM, Alber BE, Kisker C, et al. (2000) A Closer Look at the Active Site of γ-Class Carbonic Anhydrases: High-Resolution Crystallographic Studies of the Carbonic Anhydrase from Methanosarcina thermophila. *Biochemistry* 39: 9222-9231.

Kats-Ugurlu G, Oosterwijk E, Muselaers S, et al. (2014) Neoadjuvant sorafenib treatment of clear cell renal cell carcinoma and release of circulating tumor fragments. *Neoplasia* 16: 221-228.

Katz JJ and Crespi HL. (1966) Deuterated Organisms: Cultivation and Uses. *Science* 151: 1187-1194.

Kendrew JC, Bodo G, Dintzis HM, et al. (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* 181: 662-666.

Kholodar SA, Ghosh AK and Kohen A. (2017) Chapter Three - Measurement of Enzyme Isotope Effects. In: Harris ME and Anderson VE (eds) *Methods in Enzymology.* Academic Press, 43-83.

Kikutani S, Nakajima K, Nagasato C, et al. (2016) Thylakoid luminal theta-carbonic anhydrase critical for growth and photosynthesis in the marine diatom Phaeodactylum tricornutum. *Proceedings of the National Academy of Sciences of the United States of America* 113: 9828-9833.

Kim CU, Song H, Avvaru BS, et al. (2016) Tracking solvent and protein movement during CO2 release in carbonic anhydrase II crystals. *Proceedings of the National Academy of Sciences of the United States of America* 113: 5257-5262.

Klopp J, Winterhalter A, Gébleux R, et al. (2018) Cost-effective large-scale expression of proteins for NMR studies. *Journal of Biomolecular NMR* 71: 247-262.

Köhler K, Hillebrecht A, Schulze Wischeler J, et al. (2007) Saccharin Inhibits Carbonic Anhydrases: Possible Explanation for its Unpleasant Metallic Aftertaste. *Angewandte Chemie International Edition* 46: 7697-7699.

Konermann L, Pan J and Liu Y-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews* 40: 1224-1234.

Koruza K, Lafumat B, Nyblom M, et al. (2018a) From Initial Hit to Crystal Optimization with Microseeding of Human Carbonic Anhydrase IX—A Case Study for Neutron Protein Crystallography. *Crystals* 8: 434.

Koruza K, Lafumat B, Végvári Á, et al. (2018b) Deuteration of human carbonic anhydrase for neutron crystallography: Cell culture media, protein thermostability, and crystallization behavior. *Archives of Biochemistry and Biophysics* 645: 26-33.

Koruza K, Mahon BP, Blakeley MP, et al. (2019) Using neutron crystallography to elucidate the basis of selective inhibition of carbonic anhydrase by saccharin and a derivative. *Journal of Structural Biology* 205: 147-154.

Kossiakoff AA and Spencer SA. (1981) Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin. *Biochemistry* 20: 6462-6474.

Kovalevsky A, Aggarwal M, Velazquez H, et al. (2018) "To Be or Not to Be" Protonated: Atomic Details of Human Carbonic Anhydrase-Clinical Drug Complexes by Neutron Crystallography and Simulation. *Structure* 26: 383 - 390.e383.

Kuchitsu K and Bartell LS. (1962) Effect of Anharmonic Vibrations on the Bond Lengths of Polyatomic Molecules. II. Cubic Constants and Equilibrium Bond Lengths of Methane. *The Journal of Chemical Physics* 36: 2470-2481.

LaGuerre A, Löhr F, Bernhard F, et al. (2015) Chapter Fifteen - Labeling of Membrane Proteins by Cell-Free Expression. In: Kelman Z (ed) *Methods in Enzymology.* Academic Press, 367-388.

Lakey JH. (2009) Neutrons for biologists: a beginner's guide, or why you should consider using neutrons. *Journal of the Royal Society Interface* 6 Suppl 5: S567-573.

Langan P and Chen JC. (2013) Seeing the chemistry in biology with neutron crystallography. *Physical Chemistry Chemical Physics* 15: 13705-13712.

Leitans J, Kazaks A, Balode A, et al. (2015) Efficient Expression and Crystallization System of Cancer-Associated Carbonic Anhydrase Isoform IX. *Journal of Medicinal Chemistry* 58: 9004-9009.

Liljas A, Kannan KK, Bergsten PC, et al. (1972) Crystal structure of human carbonic anhydrase C. *Nature: New biology* 235: 131-137.

Liu X, Hanson BL, Langan P, et al. (2007) The effect of deuteration on protein structure: a high-resolution comparison of hydrogenous and perdeuterated haloalkane dehalogenase. *Acta Crystallographica Section D* 63: 1000-1008.

Lomelino CL, Andring JT and McKenna R. (2018) Crystallography and Its Impact on Carbonic Anhydrase Research *International Journal of Medicinal Chemistry* 2018: 21.

Lomelino CL, Supuran CT and McKenna R. (2016) Non-Classical Inhibition of Carbonic Anhydrase. *International Journal of Molecular Sciences* 17: 1150.

Lou Y, McDonald PC, Oloumi A, et al. (2011) Targeting Tumor Hypoxia: Suppression of Breast Tumor Growth and Metastasis by Novel Carbonic Anhydrase IX Inhibitors. *Cancer Research* 71: 3364.

Luft JR and DeTitta GT. (1999) A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallographica Section D* 55: 988-993.

Mahon B, Pinard M and McKenna R. (2015a) Targeting Carbonic Anhydrase IX Activity and Expression. *Molecules* 20: 2323.

Mahon BP, Bhatt A, Socorro L, et al. (2016) The Structure of Carbonic Anhydrase IX Is Adapted for Low-pH Catalysis. *Biochemistry* 55: 4642-4653.

Mahon BP, Hendon AM, Driscoll JM, et al. (2015b) Saccharin: A lead compound for structure-based drug design of carbonic anhydrase IX inhibitors. *Bioorganic & Medicinal Chemistry* 23: 849-854.

Mahon BP, Lomelino CL, Ladwig J, et al. (2015c) Mapping Selective Inhibition of the Cancer-Related Carbonic Anhydrase IX Using Structure–Activity Relationships of Glucosyl-Based Sulfamates. *Journal of Medicinal Chemistry* 58: 6630-6638.

Manzoni F, Wallerstein J, Schrader TE, et al. (2018) Elucidation of Hydrogen Bonding Patterns in Ligand-Free, Lactose- and Glycerol-Bound Galectin-3C by Neutron Crystallography to Guide Drug Design. *Journal of Medicinal Chemistry* 61: 4412-4420.

Marcsisin SR and Engen JR. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Analytical and bioanalytical chemistry* 397: 967-972.

Mboge MY, Mahon BP, Lamas N, et al. (2017) Structure activity study of carbonic anhydrase IX: Selective inhibition with ureido-substituted benzenesulfonamides. *European Journal of Medicinal Chemistry* 132: 184-191.

Mboge MY, McKenna R and Frost SC. (2015) Advances in Anti-Cancer Drug Development Targeting Carbonic Anhydrase IX and XII. *Topics in anti-cancer research* 5: 3-42.

McDonald PC, Winum J-Y, Supuran CT, et al. (2012) Recent Developments in Targeting Carbonic Anhydrase IX for Cancer Therapeutics. *Oncotarget* 3: 84-97.

McKenna R and Supuran CT. (2014) Carbonic Anhydrase Inhibitors Drug Design. *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications.* Dordrecht: Springer Netherlands, 291-323.

McPherson A and Gavira JA. (2014) Introduction to protein crystallization. *Acta Crystallographica. Section F* 70: 2-20.

Meilleur F, Coates L, Cuneo M, et al. (2018) The Neutron Macromolecular Crystallography Instruments at Oak Ridge National Laboratory: Advances, Challenges, and Opportunities. *Crystals* 8: 388.

Meilleur F, Dauvergne M-T, Schlichting I, et al. (2005) Production and X-ray crystallographic analysis of fully deuterated cytochrome P450cam. *Acta Crystallographica Section D* 61: 539-544.

Meilleur F, Weiss KL and Myles DAA. (2009) Deuterium Labeling for Neutron Structure-Function-Dynamics Analysis. In: Foote RS and Lee JW (eds) *Micro and Nano Technologies in Bioanalysis: Methods and Protocols.* Totowa, NJ: Humana Press, 281-292.

Meldrum NU and Roughton FJ. (1933) Carbonic anhydrase. Its preparation and properties. *The Journal of Physiology* 80: 113-142.

Michalczyk R, Unkefer CJ, Bacik JP, et al. (2015) Joint neutron crystallographic and NMR solution studies of Tyr residue ionization and hydrogen bonding: Implications for enzyme-mediated proton transfer. *Proceedings of the National Academy of Sciences of the United States of America* 112: 5673-5678.

Mikulski R, West D, Sippel KH, et al. (2013) Water Networks in Fast Proton Transfer during Catalysis by Human Carbonic Anhydrase II. *Biochemistry* 52: 125-131.

Mittermaier AK and Kay LE. (2009) Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* 34: 601-611.

Moeker J, Mahon BP, Bornaghi LF, et al. (2014a) Structural insights into carbonic anhydrase IX isoform specificity of carbohydrate-based sulfamates. *Journal of Medicinal Chemistry* 57: 8635-8645.

Moeker J, Peat TS, Bornaghi LF, et al. (2014b) Cyclic secondary sulfonamides: unusually good inhibitors of cancer-related carbonic anhydrase enzymes. *Journal of Medicinal Chemistry* 57: 3522-3531.

Murata K and Wolf M. (2018) Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta* 1862: 324-334.

Muselaers CH, Boers-Sonderen MJ, van Oostenbrugge TJ, et al. (2016) Phase 2 Study of Lutetium 177-Labeled Anti-Carbonic Anhydrase IX Monoclonal Antibody Girentuximab in Patients with Advanced Renal Cell Carcinoma. *European Urology* 69: 767-770.

Nango E, Royant A, Kubo M, et al. (2016) A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* 354: 1552-1557.

Neutze R, Wouts R, van der Spoel D, et al. (2000) Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* 406: 752-757.

O'Dell WB, Bodenheimer AM and Meilleur F. (2016) Neutron protein crystallography: A complementary tool for locating hydrogens in proteins. *Archives of Biochemistry and Biophysics* 602: 48-60.

O'Brien ES, Lin DW, Fuglestad B, et al. (2018) Improving yields of deuterated, methyl labeled protein by growing in H2O. *Journal of Biomolecular NMR* 71: 263-273.

Oksanen E, Chen JC and Fisher SZ. (2017) Neutron Crystallography for the Study of Hydrogen Bonds in Macromolecules. *Molecules* 22.

Oosterwdk E, Ruiter DJ, Hoedemaeker PJ, et al. (1986) Monoclonal antibody G 250 recognizes a determinant present in renal-cell carcinoma and absent from normal kidney. *International Journal of Cancer* 38: 489-494.

Oosterwijk-Wakka JC, de Weijert MC, Franssen GM, et al. (2015) Successful combination of sunitinib and girentuximab in two renal cell carcinoma animal models: a rationale for combination treatment of patients with advanced RCC. *Neoplasia* 17: 215-224.

Opitz C, Isogai S and Grzesiek S. (2015) An economic approach to efficient isotope labeling in insect cells using homemade 15N-, 13C- and 2H-labeled yeast extracts. *Journal of Biomolecular NMR* 62: 373-385.

Ozawa K, Jergic S, Crowther JA, et al. (2005) Cell-free Protein Synthesis in an Autoinduction System for NMR Studies of Protein–Protein Interactions. *Journal of Biomolecular NMR* 32: 235-241.

Palatinus L, Brazda P, Boullay P, et al. (2017) Hydrogen positions in single nanocrystals revealed by electron diffraction. *Science* 355: 166-169.

Paliy O, Bloor D, Brockwell D, et al. (2003) Improved methods of cultivation and production of deuteriated proteins from E. coli strains grown on fully deuteriated minimal medium. *Journal of Applied Microbiology* 94: 580-586.

Paliy O and Gunasekera TS. (2007) Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Applied Microbiology and Biotechnology* 73: 1169-1172.

Pastorek J and Pastorekova S. (2015) Hypoxia-induced carbonic anhydrase IX as a target for cancer therapy: from biology to clinical use. *Seminars in Cancer Biology* 31: 52-64.

Perutz MF, Rossmann MG, Cullis AF, et al. (1960) Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* 185: 416-422.

Petit-Haertlein I, Blakeley MP, Howard E, et al. (2009) Perdeuteration, purification, crystallization and preliminary neutron diffraction of an ocean pout type III antifreeze protein. *Acta Crystallographica Section F* 65: 406-409.

Pinard MA, Boone CD, Rife BD, et al. (2013) Structural study of interaction between brinzolamide and dorzolamide inhibition of human carbonic anhydrases. *Bioorganic & Medicinal Chemistry* 21: 7210-7215.

Pinard MA, Mahon B and McKenna R. (2015) Probing the Surface of Human Carbonic Anhydrase for Clues towards the Design of Isoform Specific Inhibitors. *BioMed Research International* 2015: 15.

Privalov PL and Crane-Robinson C. (2017) Role of water in the formation of macromolecular structures. *European biophysics journal* 46: 203-224.

Ravelli RBG and McSweeney SM. (2000) The 'fingerprint' that X-rays can leave on structures. *Structure* 8: 315-328.

Richards OW. (1934) The Effect of Deuterium on the Growth of Yeast. *Journal of Bacteriology* 28: 289-294.

Rubinson KA. (2017) Practical corrections for p(H,D) measurements in mixed H2O/D2O biological buffers. *Analytical Methods* 9: 2744-2750.

Santos R, Ursu O, Gaulton A, et al. (2016) A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* 16: 19-34.

Schiebel J, Gaspari R, Wulsdorf T, et al. (2018) Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nature Communications* 9: 3559.

Schmidt A, Teeter M, Weckert E, et al. (2011) Crystal structure of small protein crambin at 0.48 Å resolution. *Acta crystallographica. Section F* 67: 424-428.

Schoenborn BP. (1969) Neutron Diffraction Analysis of Myoglobin. *Nature* 224: 143-146.

Schrödinger L. (2015) *The PyMOL Molecular Graphics System, Version 1.8*.

Shu F, Ramakrishnan V and Schoenborn BP. (2000) Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin. *Proceedings of the National Academy of Sciences of the United States of America* 97: 3872.

Silverman DN and McKenna R. (2007) Solvent-Mediated Proton Transfer in Catalysis by Carbonic Anhydrase. *Accounts of Chemical Research* 40: 669-675.

Sivashanmugam A, Murray V, Cui C, et al. (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Science* 18: 936-948.

Skora L, Shrestha B and Gossert AD. (2015) Chapter Eleven - Isotope Labeling of Proteins in Insect Cells. In: Kelman Z (ed) *Methods in Enzymology.* Academic Press, 245-288.

Studier FW. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 41: 207-234.

Sumner JB. (1926) The isolation and crystallization of the enzyme urease. Preliminary paper. *Journal of Biological Chemistry* 69: 435-441.

Supuran CT. (2008) Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nature Reviews Drug Discovery* 7: 168.

Supuran CT. (2011) Bacterial carbonic anhydrases as drug targets: toward novel antibiotics? *Frontiers in pharmacology* 2: 34-34.

Supuran CT. (2012) Structure-based drug discovery of carbonic anhydrase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry* 27: 759-772.

Supuran CT. (2016) Structure and function of carbonic anhydrases. *Biochemical Journal* 473: 2023-2032.

Supuran CT. (2017) Advances in structure-based drug discovery of carbonic anhydrase inhibitors. *Expert Opinion on Drug Discovery* 12: 61-88.

Supuran CT, Alterio V, Di Fiore A, et al. (2018) Inhibition of carbonic anhydrase IX targets primary tumors, metastases, and cancer stem cells: Three for the price of one. *Medicinal Research Reviews* 38: 1799-1836.

Švastová E, Žilka N, Zat'ovičová M, et al. (2003) Carbonic anhydrase IX reduces E-cadherin-mediated adhesion of MDCK cells via interaction with β-catenin. *Experimental Cell Research* 290: 332-345.

Tegler LT, Fromell K, Jonsson B-H, et al. (2011) Polypeptide Conjugate Binders that Discriminate between Two Isoforms of Human Carbonic Anhydrase in Human Blood. *ChemBioChem* 12: 559-566.

Teicher BA, Liu SD, Liu JT, et al. (1993) A carbonic anhydrase inhibitor as a potential modulator of cancer therapies. *Anticancer research* 13: 1549-1556.

Thomas SE, Mendes V, Kim SY, et al. (2017) Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections: A Paper Dedicated to John Kendrew. *Journal of Molecular Biology* 429: 2677-2693.

Urey HC, Brickwedde, F. G., Murphy, G. M. (1932) A Hydrogen Isotope of Mass 2. *Physical Review* 39: 164-165.

van Kuijk SJA, Yaromina A, Houben R, et al. (2016) Prognostic Significance of Carbonic Anhydrase IX Expression in Cancer Patients: A Meta-Analysis. *Frontiers in Oncology* 6: 69.

Verardi R, Traaseth NJ, Masterson LR, et al. (2012) Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins. In: Atreya HS (ed) *Isotope labeling in Biomolecular NMR.* Dordrecht: Springer Netherlands, 35-62.

Villanueva J, Canals F, Villegas V, et al. (2000) Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. *FEBS Letters* 472: 27-33.

Vukovic V and Tannock IF. (1997) Influence of low pH on cytotoxicity of paclitaxel, mitoxantrone and topotecan. *British Journal of Cancer* 75: 1167-1172.

Weber IT, Waltman MJ, Mustyakimov M, et al. (2013) Joint X-ray/Neutron Crystallographic Study of HIV-1 Protease with Clinical Inhibitor Amprenavir: Insights for Drug Design. *Journal of Medicinal Chemistry* 56: 5631-5635.

Whittington DA, Waheed A, Ulmasov B, et al. (2001) Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells. *Proceedings of the National Academy of Sciences of the United States of America* 98: 9545-9550.

Wieser ME and Brand WA. (1999) Isotope Ratio Studies Using Mass Spectrometry. In: Lindon JC (ed) *Encyclopedia of Spectroscopy and Spectrometry.* Oxford: Elsevier, 1072-1086.

Yang J, Koruza K, Fisher Z, et al. (2017) Improved molecular recognition of Carbonic Anhydrase IX by polypeptide conjugation to acetazolamide. *Bioorganic & Medicinal Chemistry* 25: 5838-5848.

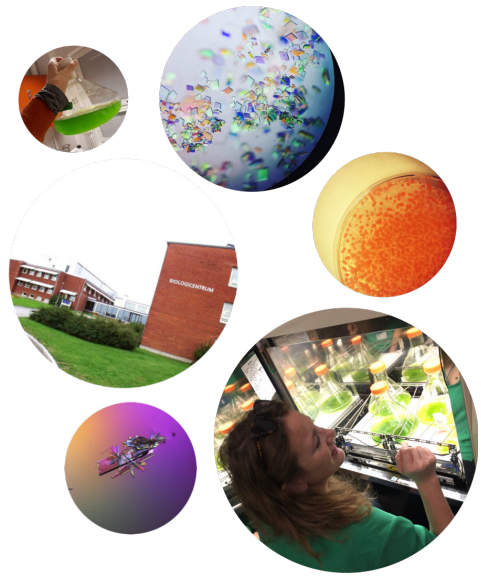
Yee AW, Blakeley MP, Moulin M, et al. (2017) Back-exchange of deuterium in neutron crystallography: characterization by IR spectroscopy. *Journal of Applied Crystallography* 50: 660-664.

Yee AW, Moulin M, Breteau N, et al. (2016) Impact of Deuteration on the Assembly Kinetics of Transthyretin Monitored by Native Mass Spectrometry and Implications for Amyloidoses. *Angewandte Chemie (International Ed. in English)* 55: 9292-9296.

Zatovicova M, Jelenska L, Hulikova A, et al. (2010) Carbonic anhydrase IX as an anticancer therapy target: preclinical evaluation of internalizing monoclonal antibody directed to catalytic domain. *Current Pharmaceutical Design* 16: 3255-3263.

Zatovicova M, Jelenska L, Hulikova A, et al. (2014) Monoclonal antibody G250 targeting CA : Binding specificity, internalization and therapeutic effects in a non-renal cancer model. *International Journal of Oncology* 45: 2455-2467.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903



Perdeuteration of Biological Macromolecules: A Case Study of Human Carbonic Anhydrases

Focus of my thesis is production of deuterated proteins for neutron protein crystallography. The long-term ambition of my work is to show that joint X-ray and neutron protein crystallography can provide insight into protein–ligand interactions and can give valuable input into structure-guided drug design and computational chemistry. As an example for this, human carbonic anhydrase IX was chosen, an emerging cancer target in need of development of isoform-specific inhibitors.

The outline of the thesis follows the workflow necessary for many protein structural biology projects: from optimization of protein expression and deuteration, to growing large crystals for neutron diffraction experiments, and finally preparing complexes of enzyme with promising new inhibitors and exploring their co-structures.

LUND UNIVERSITY

Department of Biology
Faculty of Science
Lund University
ISBN 978-91-7895-068-3

9 789178 950683