

Thesis for the degree of Doctor of  
Philosophy in Natural Science

Microcrystallization and  
Structural Studies of Proteins  
by Serial Crystallography  
and X-ray Scattering

Owens Uwangu



UNIVERSITY OF GOTHENBURG  
Department of Chemistry and Molecular Biology  
Gothenburg 2023

Thesis for the degree of Doctor of Philosophy in Natural Science

# Microcrystallization and Structural Studies of Proteins by Serial Crystallography and X-ray Scattering

Owens Uwangué

**Cover:** CYP3A4 microcrystals on a sitting drop well

Copyright ©2023 by Owens Uwangué

ISBN 978-91-8069-457-5 (Print)

ISBN 978-91-8069-458-2 (PDF)

Available online at <http://hdl.handle.net/2077/78474>

Department of Chemistry and Molecular

Biology Division of Biochemistry and Structural Biology

University of Gothenburg

SE-405 30, Göteborg, Sweden

Printed by Stema Specialtryck AB

Borås, Sweden, 2023

*To Maria, my love, my best friend. If laughter truly does prolong life, then may we live a thousand years of childlike play.*



# Abstract

Proteins are the most complex and versatile biomolecules known, with essential roles in all cellular and physiological processes. They constitute the main structural components of cells, catalyze most biochemical reactions, regulate gene expression, modulate immune responses, and form tissues. Proteins are composed of amino acids, which also have diverse functions as neurotransmitters, hormones, and regulators of various physiological pathways. Proteins have evolved and adapted their structures and functions over billions of years of evolutionary history. This thesis describes methods for expression, purification, and crystallization of proteins, as well as the techniques for analysis and characterization of protein structures at room temperature. This thesis also reports the results obtained from these methods and techniques. We applied various methods to obtain high-quality microcrystals for serial crystallography. We used serial crystallography to determine the structures of human Cytochrome P450 3A4 (CYP3A4) and soluble Epoxide Hydrolase bound to seven ligands. We detected a temperature-dependent difference in the binding mode of one ligand and improved resolution of some flexible loops at ambient temperature compared to cryo-temperature structures. (Paper I & II) We used time-resolved serial femtosecond crystallography to study the structural dynamics of *ba*<sub>3</sub>-type Cytochrome *c* oxidase (CcO) after CO photodissociation from reduced heme *a*<sub>3</sub>. We found that CO is stabilized at Cu<sub>B</sub> by a transient water molecule, unlike in *aa*<sub>3</sub>-type CcO. This explained the long-lived Cu<sub>B</sub> – CO complex and the high oxygen affinity of *ba*<sub>3</sub>-type CcO (Paper III). We confirmed CYP3A4 ligand binding by UV-vis absorbance spectroscopy, showed that CYP3A4 forms tetramers in solution and undergoes a distinct conformational change upon ketoconazole binding by X-ray scattering (Paper IV).

# Populärvetenskaplig sammanfattning

Proteiner är de mest komplexa och mångsidiga biomolekylerna som är kända, med viktiga roller i alla cellulära och fysiologiska processer. De utgör de huvudsakliga strukturella komponenterna i celler, katalyserar de flesta biokemiska reaktioner, reglerar genuttryck, modulerar immunreaktioner och bildar vävnader. Proteiner består av aminosyror, som också har olika funktioner som neurotransmittorer, hormoner och regulatorer av olika fysiologiska vägar. Proteiner har utvecklats och anpassat sina strukturer och funktioner under miljarder år av evolutionär historia. Avhandlingen beskriver metoder för uttryck, rening och kristallisering av proteiner, samt tekniker för analys och karakterisering av proteinstrukturer vid rumstemperatur. Denna avhandling rapporterar också resultat som erhållits från dessa metoder och tekniker. Vi använde olika metoder för att producera högkvalitativa mikrokristaller för seriell kristallografi. Vi använde seriell kristallografi för att bestämma strukturerna av humant Cytochrome P450 3A4 (CYP3A4) och soluble Epoxide Hydrolase bundet till sju ligander. Vi upptäckte en temperaturberoende skillnad i bindningsläget för en ligand och förbättrad upplösning av några flexibla loopar vid rumstemperatur jämfört med kryo-temperaturstrukturer. (Artikel I & II) Vi använde tidsupplöst seriell femtosekundskristallografi för att studera dynamiken hos *ba<sub>3</sub>*-typ Cytochrome *c* oxidase (CcO) efter CO-fotodissociation från reducerat heme *a<sub>3</sub>*. Vi fann att CO stabiliseras bundet till Cu<sub>B</sub> av en transient vattenmolekyl, till skillnad från *aa<sub>3</sub>*-typ CcO. Detta förklarar det långlivade Cu<sub>B</sub> – CO-komplexet och den höga syreaffiniteten hos *ba<sub>3</sub>*-typ CcO (Artikel III). Vi bekräftade CYP3A4-ligandbindning med UV-vis absorptionspektroskopi, samt använde röntgenspridning för att visa att CYP3A4 bildar tetramerer i lösning och genomgår en distinkt konformationsförändring vid ketoconazolbindning (Artikel IV).

# Publications

Here is the list of research papers that are included in this thesis.

- PAPER I      **Owens Uwangue**, Andreas Dunge, Johan Glerup, Monika Bjelcic, Gabrielle Wehlander and Gisela Brändén. *Room-temperature serial crystallography structure of human cytochrome P450 3A4*. Manuscript (2023).
- PAPER II      Andreas Dunge, **Owens Uwangue**, Cindy Phan, Jenny Gunnarsson, Gabrielle Wehlander, Helena Käck, and Gisela Brändén. *Exploring serial crystallography for drug discovery*. Manuscript (2023).
- PAPER III      C. Safari†, S. Ghosh†, R.Andersson, J. Johannesson, P. Båth, **O. Uwangue**, P. Dahl, D. Zoric, E. Sandelin, A. Vallejos, E. Nango, R. Tanaka, R. Bosman, P. Börjesson, E. Dunevall, G. Hammarin, G. Ortolani, M. Panman, T. Tanaka, A. Yamashita, T. Arima, M. Sugahara, M. Suzuki, T.Masuda, H. Takeda, R. Yamagiwa, K. Oda, M. Fukuda, T. Tosha, H. Naitow, S. Owada, K. Tono, O. Nureki, S. Iwata, R. Neutze and G. Brändén. *Time-resolved serial crystallography to track the dynamics of carbon monoxide in the active site of cytochrome c oxidase*. Manuscript under review (2023). † These authors contributed equally.
- PAPER IV      **Owens Uwangue**, Johan Glerup, Greger Hammarin, Per Börjesson and Gisela Brändén. *Investigation of the cytochrome P450 3A4 oligomeric state and ligand-induced conformational changes using X-ray solution scattering*. Manuscript (2023).

Here is a list of unrelated papers that I co-authored before and during my doctoral studies but are not included in this thesis.

- PAPER V Antonio Carlesso, Raquel Delgado, Oriol Ruiz Isant, **Owens Uwangue**, Dylan Valli, Roslyn M Bill, and Kristina Hedfalk. *Yeast as a tool for membrane protein production and structure determination*. FEMS yeast research (2022). <https://doi.org/10.1093/femsyr/foac047>
- PAPER VI Martin Palm, Matteo De Chiara, Ida Rahmqvist, Daniel Jaén-Luchoro, Sakshi Khaiwal, Simon Stenberg, Jonas Boström, Sandra Pueyo Hurtado, Marta Tous Mohedano, **Owens Uwangue**, Silvana Smits, Edward R. B. Moore, Leopold Parts, Gianni Liti, Jonas Warringer and Anne Farewell. *Mapping the Antibiotic and Heavy Metal Phenotype-Genetic Landscape of Escherichia coli*. Manuscript (2022)
- PAPER VII Hanna Thomsen, Marco Agnes, **Owens Uwangue**, Linnéa Persson, Matilda Mattsson, Fabrice E. Graf, Eleni-Marina Kasimati, Konstantina Yannakopoulou, Marica B. Ericson, Anne Farewell. *Increased antibiotic efficacy and noninvasive monitoring of Staphylococcus epidermidis biofilms using per-cysteamine-substituted  $\gamma$ -cyclodextrin – A delivery effect validated by fluorescence microscopy*. International Journal of Pharmaceutics (2020). <https://doi.org/10.1016/j.ijpharm.2020.119646>
- PAPER VIII Ibrahim Mamedov, Farid Naghiyev, Abel Maharramov, **Owens Uwangue**, Anne Farewell, Per Sunnerhagen, Mate Erdelyi. *Antibacterial activity of 2-amino-3-cyanopyridine derivatives*. Mendeleev Communications (2020). <https://doi.org/10.1016/j.mencom.2020.07.031>

# Contribution report

## **PAPER I**

I was responsible for protein production, purification, and crystallization. I contributed to the data collection, data processing, refinement, prepared figures, and I co-wrote the manuscript.

## **PAPER II**

I participated in multiple experiments and contributed data collection.

## **PAPER III**

I contributed to protein production, purification, and crystallization. I participated in the experiment and contributed to the collection of X-ray diffraction data.

## **PAPER IV**

I was responsible for protein production and purification. I performed the absorbance spectroscopy measurements. I contributed to the data collection and analysis of the X-ray scattering experiments, and I co-wrote the manuscript.

# Abbreviations

<b>ATP</b>	Adenosine triphosphate
<b>CcO</b>	Cytochrome <i>c</i> oxidase
<b>CO</b>	Carbon monoxide
<b>CYP(s)</b>	Cytochromes P450(s)
<b>CYP3A4</b>	Cytochrome P450 3A4
<b>DMSO</b>	Dimethylsulfoxid
<b>EET(s)</b>	Epoxyeicosatrienoic acid(s)
<b>ETC</b>	Electron transport chain
<b>FT</b>	Fixed target
<b>IDA</b>	Iminodiacetic acid
<b>IEC</b>	Ion exchange chromatography
<b>IMAC</b>	Immobilized metal affinity chromatography
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>LC</b>	Liquid chromatography
<b>LCP</b>	Lipidic cubic phase
<b>MR</b>	Molecular Replacement
<b>MW</b>	Molecular weight
<b>NMR</b>	Nuclear magnetic resonance
<b>NTA</b>	Nitrilotriacetic acid
<b>PCR</b>	Polymerase chain reaction
<b>PDB</b>	Protein Data Bank
<b>PEG</b>	Polyethylene glycol
<b>pI</b>	Isoelectric point
<b>RH</b>	Organic substrate(s)
<b>RT</b>	Room temperature
<b>SAXS</b>	Small-angle x-ray scattering
<b>SBDD</b>	Structure-based drug design
<b>SEC</b>	Size exclusion chromatography
<b>sEH</b>	soluble Epoxide Hydrolase
<b>SFX</b>	Serial femtosecond crystallography
<b>SRS(s)</b>	Substrate recognition site(s)
<b>SSX</b>	Serial synchrotron crystallography
<b>SX</b>	Serial crystallography
<b>TR</b>	Time-resolved
<b>UV</b>	Ultraviolet
<b>WAXS</b>	Wide-angle x-ray scattering
<b>XFEL(s)</b>	X-ray free-electron laser(s)
<b><math>\delta</math>-ALA</b>	$\delta$ -Aminolevulinic acid

# Table of Contents

<b>INTRODUCTION</b>	<b>1</b>
1.1 Protein	1
1.1.1 Protein structure and conformation	1
1.1.2 Hierarchical structure of proteins	2
1.2 Serial X-ray crystallography	4
1.3 Cytochrome P450	6
1.3.1 Cytochrome P450 oxidation mechanism	7
1.3.2 Cytochrome P450 3A4	9
1.3.3 CYP3A4 architecture and flexibility	10
1.4 Soluble epoxide hydrolase	13
1.5 Cytochrome <i>c</i> oxidase	14
1.5.1 Cytochrome <i>ba</i> <sub>3</sub> oxidase	15
1.6 Scope of this thesis	16
<b>METHODOLOGY</b>	<b>17</b>
2.1 Protein production	17
2.1.1 CYP3A4 expression in Escherichia coli XL1-Blue	17
2.2 Protein purification	18
2.2.1 Cell disruption and centrifugation	19
2.2.3 Liquid chromatography	20
2.3 Protein crystallization	22
2.3.1 Supersaturation, nucleation, and crystal growth	23
2.3.2 Membrane protein crystallization	24
2.4 Crystallography	25
2.4.1 X-ray crystallography	26
2.4.2 Serial crystallography	27
2.4.2 X-ray diffraction	29
2.4.3 Data collection, processing, and refinement	32
2.5 Absorbance spectroscopy	36
2.6 X-ray scattering	38
<b>RESULTS AND DISCUSSION</b>	<b>41</b>

3.1 Microcrystallization of CYP3A4 for RT FT-SSX (Paper I)	41
3.1.1 Production of CYP3A4 macrocrystals	41
3.1.2 From macro to microcrystals	43
3.1.3 Batch crystallization	44
3.1.4 Ligand co-crystallization	45
3.1.5 Conclusions from CYP3A4 crystallization	46
3.1.6 Data collection, processing, and structure refinement	47
3.1.7 RT structure of CYP3A4	47
3.1.8 Summary	49
3.2 Exploring sEH as model for drug discovery by RT FT-SSX (Paper II)	50
3.2.1 Production of sEH microcrystals	50
3.2.2 Data collection, processing, and structure refinement	50
3.2.3 RT structures of sEH-ligand complexes and temperature related observations	51
3.2.4 Summary	53
3.3 Structural dynamics of reduced CO-bound <i>ba</i> <sub>3</sub> -type CcO by TR SFX (Paper III)	54
3.3.1 Production of <i>ba</i> <sub>3</sub> -type CcO microcrystals	54
3.3.2 Data collection, processing, and structure refinement	54
3.3.3 RT structure of reduced CO-bound <i>ba</i> <sub>3</sub> -type CcO	55
3.3.4 Photodissociation of CO from reduced CO-bound <i>ba</i> <sub>3</sub> -type CcO	55
3.3.5 Summary	57
3.4 CYP3A4 ligand interaction and structural studies by X-ray scattering (Paper IV)	58
3.4.1 CYP3A4 ligand binding titration studies by absorbance spectroscopy	58
3.4.2 Structural studies by X-ray scattering	60
3.4.3 Summary	62
<b>CONCLUDING SUMMARY</b>	<b>65</b>
<b>ACKNOWLEDGEMENTS</b>	<b>67</b>
<b>BIBLIOGRAPHY</b>	<b>71</b>

# Chapter 1

## Introduction

### 1.1 Protein

The word protein was coined in 1838 by the Swedish chemist Jöns Jacob Berzelius in a letter to Dutch chemist Gerardus Johannes Mulder.<sup>7</sup> Berzelius adapted the term from the Greek word *proteios*, meaning “primary” to emphasize the fundamental significance of this class of molecule.<sup>8</sup> Proteins are involved in every aspect of physiology and comprises the major structural elements of all cells. They play crucial, life-sustaining biological roles, e.g., gene expression regulation, catalysis of most chemical reactions, immune system regulation, and are major constituents in tissue formation.<sup>8, 9</sup> Protein building blocks (amino acids) also have diverse biological functions as neurotransmitters, hormones, and regulators of various physiological processes.<sup>9</sup>

#### 1.1.1 Protein structure and conformation

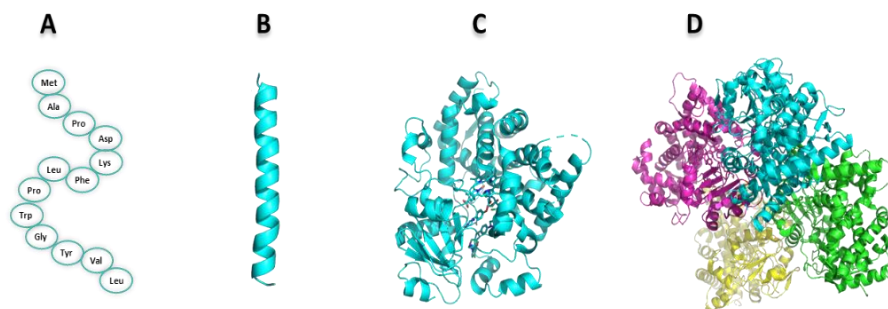
Proteins are the most structurally complex and functionally sophisticated biomolecules known from a chemical perspective.<sup>10</sup> The structure and function of each protein has been developed and fine-tuned over billions of years of evolutionary history, but not until crucial progress in sequencing that occurred in the 1950s was our current understanding of their chemical properties and biological function established.<sup>11</sup> Proteins are composed of 20 amino acids, which are small organic molecules with an alpha (central) carbon atom covalently bonded to an amino group, a carboxyl group, a hydrogen atom, and a variable side chain. Multiple amino acids are linked within a protein by peptide bonds into a polypeptide chain.<sup>10, 12, 13</sup> In 1958, the first three-dimensional structure of protein was established through the study of the myoglobin molecule at 6 Å resolution by X-ray analysis.<sup>14, 15</sup>

### 1.1.2 Hierarchical structure of proteins

Protein structure was hypothesized to have at least four levels of hierarchical organization (Figure 1) by Danish researcher Kaj Ulrik Linderstrøm-Lang in 1952. The Linderstrøm-Lang proposal defined a protein's primary structure as the linear amino acid sequence of its polypeptide chain (Figure 1A).<sup>16-18</sup> In 1973, Chris Anfinsen showed that the higher levels of protein structure are solely determined by the sequence of amino acids in the primary structure.<sup>19</sup> However, this is only applicable when optimal environmental conditions (temperature, pH, solvent, ionic strength) are met. Interactions with ligands, substrates, or other proteins can have substantial effect on the three-dimensional structure of a protein.<sup>20</sup> Secondary structure refers to the local spatial arrangements ( $\alpha$ -helices,  $\beta$ -sheets, and turns) of polypeptide backbone excluding the side chains (Figure 1B). The secondary structure results from the hydrogen bonds between partially negatively charged oxygen atoms and partially positively charged nitrogen atoms of the polypeptide backbone.<sup>21</sup> Tertiary structure describes the three-dimensional shape of a protein molecule formed by all its atoms (Figure 1C), and is mainly driven by non-specific hydrophobic interactions.<sup>22,23</sup> Quaternary structure refers to the precise spatial arrangement of several protein subunits in a multi-domain structure (Figure 1D). Protein folding has been studied for more than 60 years, since the first demonstration that proteins can spontaneously refold to their native state after denaturation<sup>24</sup>.

Development in X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy techniques have influenced the general understanding of protein folding by providing structural information on their final folded states. This has resulted in the further understanding of the folding mechanism, identification of key interactions that stabilize protein structures, enabled the design and synthesis of novel protein with desired folds and functions, and drug design with improved efficiency. X-ray crystallography studies have helped with the exploration of the relationship between sequence, structure, and function.<sup>25-30</sup>

The biological function of a protein depends on the spatial arrangement of its atoms, which determines its three-dimensional structure. The structure of a protein enables its catalytic activity, its interactions with other proteins, and its regulation by various factors. By elucidating the structure of a protein, we can gain a deeper understanding of its molecular mechanism and generate hypotheses about how to manipulate, modulate, or modify it. For instance, we can use structural information to design site-specific mutations that alter the function of a protein. Alternatively, we can predict potential ligands that bind to a protein and affect its activity or stability.<sup>10, 31</sup> Structural information of proteins provides insight for structure-based drug design (SBDD). SBDD is an approach that exploits the structural information of a protein to identify or design novel chemical compounds that can interact with the protein and inhibit its function.<sup>32</sup> Protein structure may also reveal the evolutionary relationships between different proteins and organisms.<sup>33</sup>



**Figure 1. Representation of the hierarchical structure of proteins.** (A) Linear amino acid sequence linked by peptide bond defines the primary structure. (B) Spatial arrangement of polypeptide backbone atoms determines the secondary structure ( $\alpha$ -helices,  $\beta$ -sheets, and turns). (C) Tertiary structure describes the three-dimensional shape of a protein molecule formed by all its atoms. (D) Quaternary structure is the arrangement of several protein subunits in a multi-domain structure. (PDB 2V0M)

## 1.2 Serial X-ray crystallography

Serial crystallography enables atomic-resolution structural determination of proteins without the need for large protein crystals. Serial X-ray crystallography is a technique that uses X-ray free-electron lasers (XFELs) or synchrotron X-rays to collect diffraction data from many small crystals (microcrystals) of macromolecules, such as proteins or nucleic acids.<sup>34-36</sup> Unlike conventional crystallography, which requires large crystals and cryogenic cooling, serial crystallography allows the study of room-temperature structures and time-resolved dynamics of biomolecules. Serial crystallography can also overcome the problem of radiation damage, which limits the resolution and accuracy of conventional methods. Radiation damage is minimized because microcrystals are exposed to the X-ray beam for a very short duration and only once, and by using ultra-short and ultra-intense X-ray pulses (at XFELs).<sup>35, 37-40</sup>

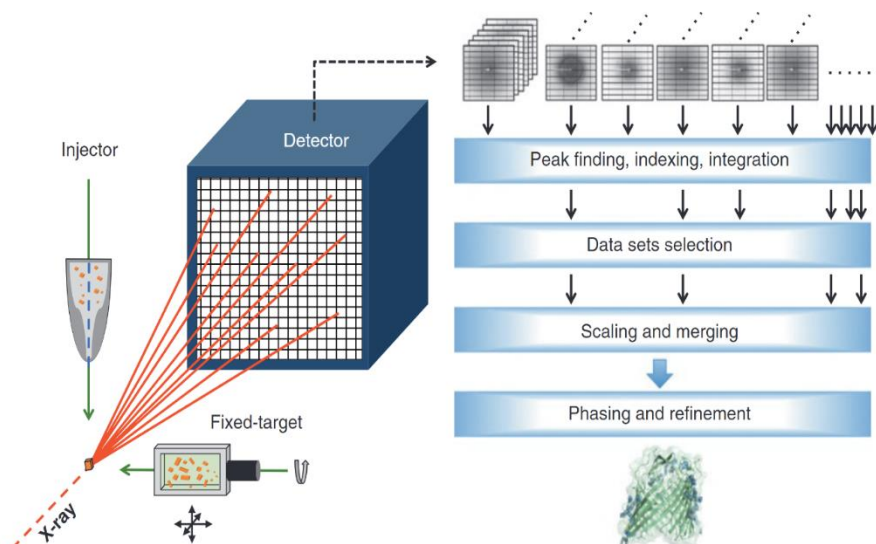
Serial crystallography can be performed using different methods of sample delivery and data analysis (Figure 2). Some of the common sample delivery methods are:

*Fixed-target* method utilizes standard crystal loops or silicon chips to mount crystals on a moving stage i.e., a goniometer or a more complex motor driven scanning stage like that of the roadrunner. These methods have low sample consumption and background scattering depends on the system.<sup>41-44</sup>

*High-viscosity sample-injection* method delivers crystals embedded in a viscous lipid matrix. This method has low sample consumption, relatively high background (to liquid jets), and is particularly suited for membrane proteins.<sup>45, 46</sup> This injection-method can also be used for crystals in solution that have been mixed with a viscous delivery medium i.e., grease or agarose gel.<sup>47, 48</sup>

*Gas Dynamic Virtual Nozzle (GDVN)* method generates a liquid jet containing crystals in vacuum. This method has low background scattering, but high sample consumption. It is the only method available for high repetition rate sources.<sup>49-52</sup>

*Tape drive* method uses a Kapton tape to transport crystals to the X-ray focus. This method is like fixed target systems, but with fewer moving parts.<sup>53</sup>



**Figure 2. Serial X-ray crystallography.** From microcrystals to structure. *Adapted figure from Diederichs et al.*<sup>1</sup>

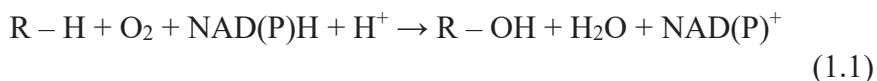
## 1.3 Cytochrome P450

Cytochromes P450 (CYPs) are a superfamily of heme-containing enzymes found across all kingdoms of life. Over 270 different CYP families containing more than 12,000 individual genes have been identified, humans have 57 CYP genes and 33 pseudogenes. CYPs found in prokaryotes are soluble, whereas eukaryotic counterparts are found attached via a membrane anchoring helix to the inner mitochondrial membrane or to the endoplasmic reticulum.<sup>54</sup> The first evidence of CYPs was found in rabbit liver microsomes in the 1950's with an observation that isolated microsomes were able to metabolize a variety of drugs when exposed to chemically prepared TPNH (NADPH).<sup>55</sup> The novel protein was later isolated and designated as P450, reflecting the pigment-450 characteristic of its heme group, which exhibited a peak absorbance at 450 nm when reduced and complexed with carbon monoxide (CO).<sup>56-58</sup> Drug metabolism refers to the biochemical modification of xenobiotics (foreign compounds) upon exposure to endogenous enzyme. Typically, drug metabolism reduces their pharmacological activity and increases water solubility to facilitate elimination in urine or bile. Cytochromes P450 enzymes predominantly catalyze a monooxygenase reaction (1.1), whereby an organic substrate (RH) undergoes aliphatic hydroxylation by one oxygen atom, while the other oxygen atom is reduced to water.

CYPs pathways are classified by similar gene sequences. Genes with similar sequences are grouped into cytochrome P450 pathways. Each pathway has a number for the family (e.g., CYP1, CYP2) and a letter for the subfamily (e.g., CYP1A, CYP2D). Within each subfamily, the isoforms or enzymes are distinguished by another number (e.g., CYP1A1, CYP2D6). CYPs determine the potential for drug interactions, and drugs that share common pathways have increased potential for drug-drug interaction. Not all drugs affect CYP. However, some drugs can inhibit, induce, or be substrates for a specific CYP pathway, changing the metabolism of co-administered drugs (drug – drug interaction). Inhibitors increase concentrations of drugs that are metabolized using the same pathway, causing toxicity. Inducers decrease concentrations of drugs using the same pathway, causing

treatment failure.<sup>59, 60</sup> CYPs metabolic effects are not limited to the liver but extends to the small intestine. Drugs absorbed from the small intestine often go through first-pass metabolism by CYP3A4. CYPs inhibitor i.e., Grapefruit juice blocks CYP3A4 in the small intestine and increases systemic levels of CYP3A active drugs. Some drugs are both substrate and inhibitor, and requires caution as co-prescription of other CYP3A inhibitors and inducers as they may be detrimental with cascading effects.<sup>59-63</sup>

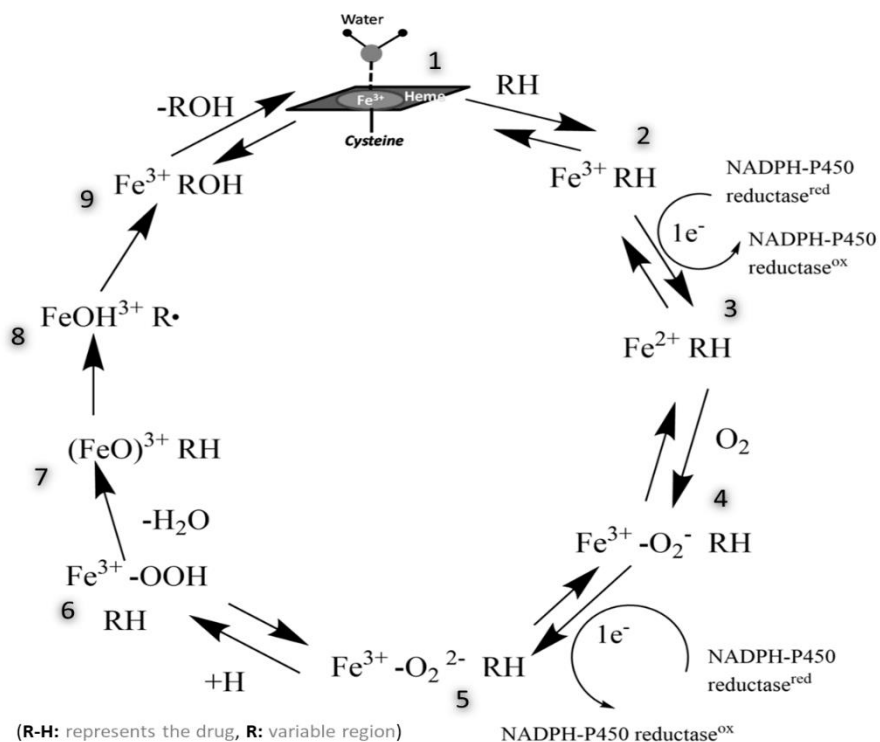
The complexity of CYPs systems is why its study is crucial, and with improved understanding of CYPs, extensive drug interaction studies are performed on new drugs before they reach the market. Drugs interaction studies are combined with prediction based on metabolic pathways without testing, as testing every new drug in combination with every other drug before on every pathway would be insurmountable.



### 1.3.1 Cytochrome P450 oxidation mechanism

The catalytic cycle that drives the oxidation mechanism of CYPs is widely conserved across the enzyme family (Figure 3). For it to occur, most CYPs require the procurement of two electrons delivered by a flavin- or [2Fe-2S]-cluster-containing redox partner, and one molecule of oxygen per catalytic cycle.<sup>54</sup> Prior to catalysis, the iron forms a ferric, six-coordinate low-spin complex with the porphyrin ring, the cysteine axial ligand and a water molecule (1). The binding of a substrate lowers the redox potential of the heme by  $\approx 100$  mV, which allows electron transfer from its redox partner (NAD(P)H).<sup>64</sup> The axial water ligand dissociates upon substrate binding, which results increase redox potential of active site iron atom, and a change in its spin state from low to high (2). Ferric iron is reduced to ferrous iron by an electron transferred from NAD(P)H in first electron transfer step (3). Molecular oxygen binds to ferrous iron to form the superoxol

intermediate (4). A second electron transfer from NAD(P)H (5). Oxygen is protonated twice from surrounding solvents, breaking the O – O bond, releasing water, and leaving (Fe-O)<sup>3+</sup> R-H complex (6 – 7). It's now thermodynamically and kinetically favorable for Fe-ligated oxygen atom to transfer to the substrate forming R-OH (Hydrophilic R-H is hydroxylated to soluble R-OH) (8 – 9). The substrate is released, and enzyme reverts to its resting state (1).

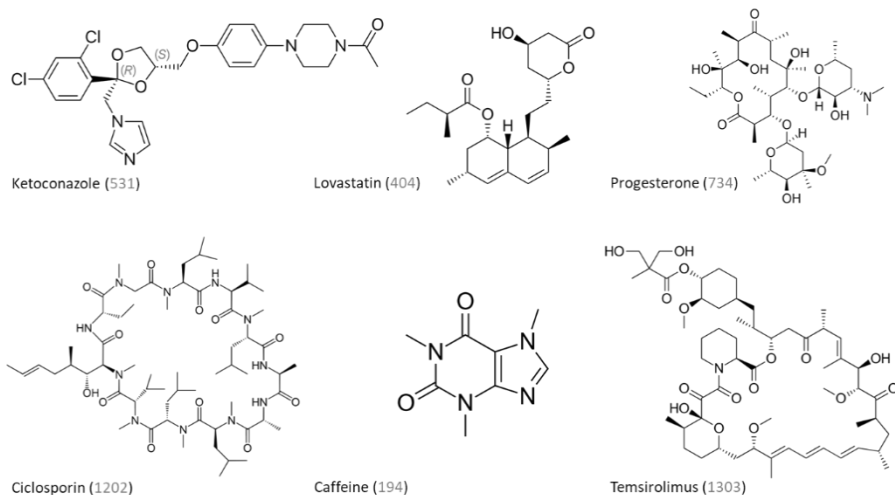


**Figure 3. General catalytic mechanism of Cytochrome P450 reactions.** (1) The enzyme resting state. Ferric heme is coordinated to a water molecule. (2) Drug (R-H) binds to the active site -> conformation change to Fe<sup>3+</sup> CYP450 – RH complex. (3) Complex is reduced by NADPH. (4) Oxygen molecule binds to the heme iron (5) Complex is further reduced by NADPH. (6 – 7) Oxygen is protonated twice, releasing water, and leaving an (Fe-O)<sup>3+</sup> R-H complex. (8 – 9) It's now thermodynamically and kinetically favorable for oxygen to transfer to the substrate forming R-OH. R-OH is released, and enzyme reverts to resting state.

### 1.3.2 Cytochrome P450 3A4

Cytochrome P450 3A4 (CYP3A4) is the most abundant CYP isoform in Humans and is mainly expressed in the liver and gastrointestinal tract.<sup>65</sup> The enzyme is responsible for about 45% of all phase I metabolism and up to 70% of gastrointestinal CYP activity.<sup>66</sup> CYP3A4 metabolizes approximately 50% of drugs marketed today and is a primary interest in drug design.<sup>65</sup> CYP3A4 is a member of the Cytochrome P450 family of oxidizing enzymes, and they represent a large class of heme-containing enzymes that catalyze the metabolism of multitudes of substrates both endogenous and exogenous compounds (e.g., drugs, chemicals, pollutants, steroids, bile and fatty acids, Vitamin D and other natural products and xenobiotics).<sup>67</sup>

The general understanding of enzyme-substrate interaction involves the binding of substrate to static active site of an enzyme to allow catalysis (lock & key mechanism).<sup>68</sup> While some CYP-substrate interactions have been observed to follow this lock & key model,<sup>69-71</sup> more recent studies demonstrates that other CYP-substrate interactions can be more complex than the stoichiometric one-step, two-state substrate enzyme-substrate interaction,<sup>72-77</sup> and of particular interest CYP3A4.<sup>72-74, 77, 78</sup> Viewed through the lock & key mechanism, this complex interaction involves the lock morphing to fit a vast range of key shapes and sizes. Common structural features and variations (e.g., topology, the size of active sites, degree of active site flexibility, preferred access/egress paths to active sites, and the water network within the active site) have been identified through Cytochrome P450 X-ray crystallography.<sup>79-85</sup> CYP3A4 interacts a large range of structurally diverse molecules (Figure 4),<sup>68</sup> its active site undergoes drastic conformational changes (>80% increase in active site volume) upon binding to erythromycin or ketoconazole.<sup>65</sup> CYP3A4 active site malleability provides a structural basis for ligand promiscuity.

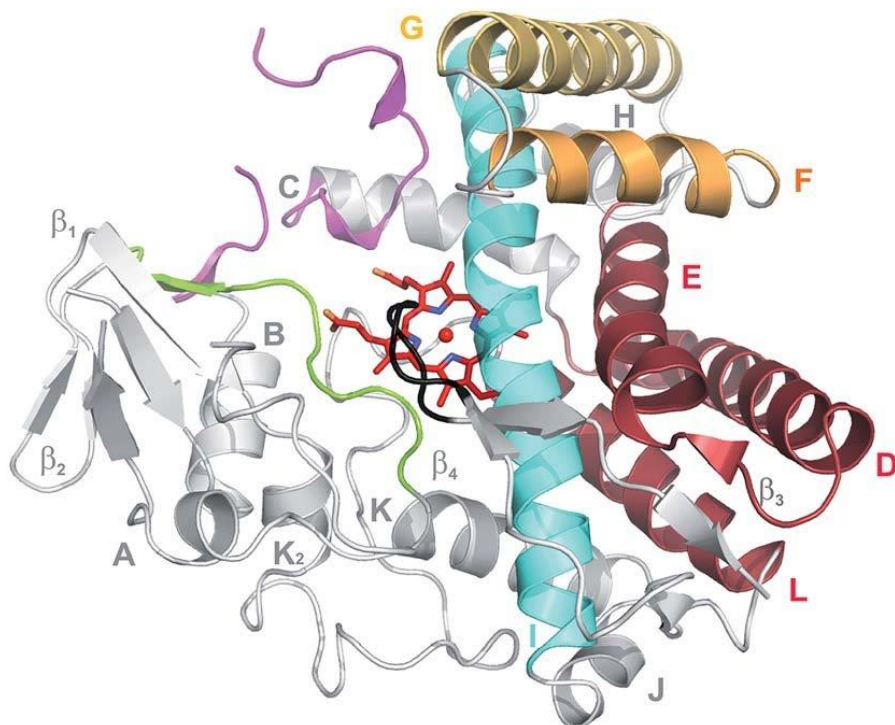


**Figure 4. CYP3A4 versatility represented through its interaction with a large range of structurally diverse molecules.** Six drugs (molar mass) listed from the >250 drugs that act as substrates, inhibitors, or inducers to CYP3A4.

### 1.3.3 CYP3A4 architecture and flexibility

97 crystal structures of CYP3A4 have been deposited in the Protein Data Bank (PDB), providing insights into its molecular mechanism and substrate specificity. CYP3A4 exhibits remarkable conformational flexibility, which enables it to accommodate and bind various ligands in its large active site, estimated to have a maximum volume of  $520 \text{ \AA}^3$ .<sup>86-88</sup> The enzyme can adopt multiple conformations depending on the ligand bound, reflecting its promiscuous nature.<sup>89</sup> CYP3A4 is considered the most versatile human CYPs in terms of their catalytic activity and substrate range.<sup>90</sup> The structural diversity of the compounds that interact with CYP3A4 is evident in the different crystal structures. Remarkably, the cavity can adjust to fit multiple ligands simultaneously, and may also assume a conformation that may expose a noncanonical binding site (Figure 6).<sup>86</sup> The general topology and structural fold of CYP enzymes are highly conserved across different species. The structural folds are described by an alphanumeric nomenclature that assigns a letter to each  $\alpha$ -helix and a number to each  $\beta$ -sheet, following the order from the N-terminus (Figure 5).<sup>91-93</sup> The

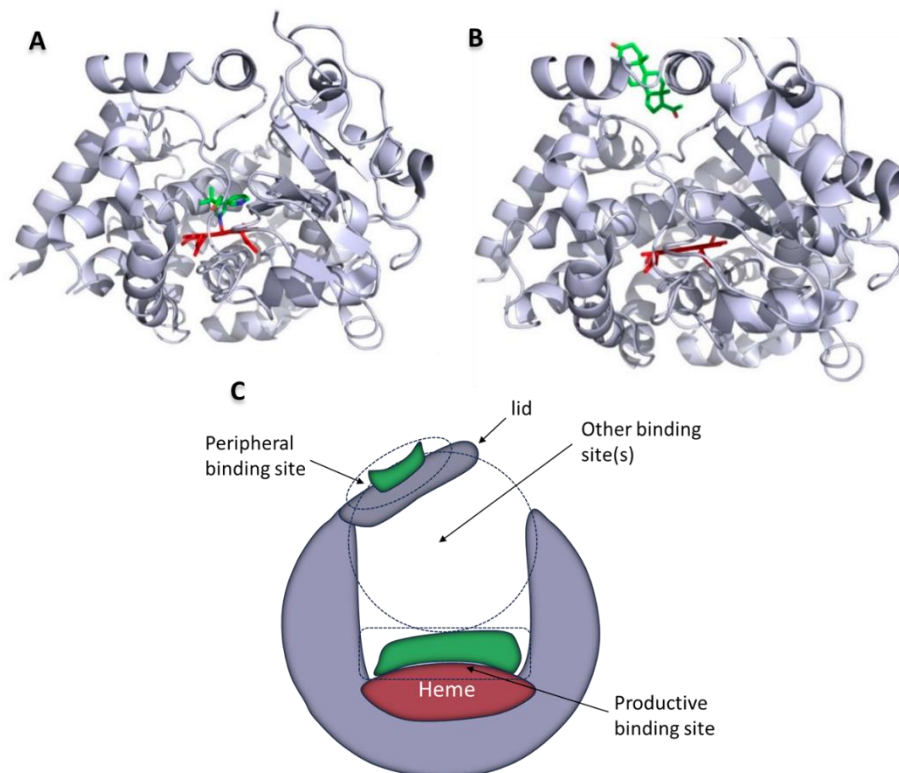
protein center surrounding the heme prosthetic group exhibits the highest structural conservation among different CYPs, indicating a common oxygen activation mechanism. This core structure consists of four helices (D, E, I and L), helices J and K, two  $\beta$ -sheet regions and a loop that binds heme.<sup>94</sup>



**Figure 5. Typical structure of a cytochrome P450**, exemplified by CYP165B7 showing the preponderance of  $\alpha$ -helices in the structure (labelled A–L). The core 4 helix bundle of the P450 is shown in deep red, except for the I-helix that is shown in turquoise. The regions comprising the active site are shown in colors (BC loop region in pink, F-helix shown in orange, the G-helix shown in gold, the I-helix shown in turquoise, the  $\beta$ -strand following the K-helix shown in green and the C-terminal loop shown in black). The central heme moiety is shown as sticks. *Adapted figure from Greule et al.*<sup>6</sup>

The membrane anchor and the substrate recognition regions are the variable regions of CYPs. The substrate recognition sites (SRSs, Figure 5), which have low sequence similarity, are located in conserved positions near the substrate-access and the catalytic site.<sup>95</sup> The F/G region, which forms the lid of the active site, are flexible regions that modulate substrate entry and exit and the opening and closing

movements of the active site (Figure 6C). The SRSs regions are vital for aligning the substrate in the active site and thus influence the regio and stereoselectivity of the oxidation process.<sup>80, 96, 97</sup> Unlike prokaryotic CYPs, eukaryotic CYPs are usually membrane-bound and typically anchored to the ER membrane by a short N-terminal trans-membrane  $\alpha$ -helix. The transmembrane helix is not essential for function and is often removed to simplify CYP expression and crystallization.<sup>98</sup>

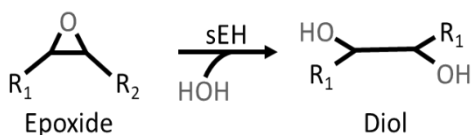


**Figure 6. Ligand binding regions of CYP3A4.** (A) Inhibitor metyrapone (green) bound to heme iron (red) by one of its pyridine nitrogen (PDB 1W0F). (B) endogenous hormone progesterone is bound on the CYP3A4 surface to a pocket made by the F/G region (PDB 1W0G). (C) Cartoon representation of CYP3A4 (purplish gray) with the different ligand binding regions. The binding site close to the heme (red) that leads to substrate oxidation is sometimes called the productive binding site. The other sites, including the peripheral binding site are part of the large active site pocket. *Crystal structures adapted Urban et al.*<sup>5</sup>

## 1.4 Soluble epoxide hydrolase

The epoxide hydrolases (EHs) are a family of enzymes present in all living organisms. They catalyze the hydrolysis of epoxide-containing lipids, which are important mediators of cellular signaling and inflammation.<sup>99, 100</sup> Mammalian soluble epoxide hydrolase (sEH) is a bifunctional enzyme that is encoded by the EPHX2 gene and is expressed in a vast number of tissues.<sup>101, 102</sup> The highest levels of sEH expression are found in the liver and kidney, followed by other organs such as the lungs, heart, brain, adrenals, spleen, intestines, urinary bladder, placenta, skin, mammary gland, testis, leukocytes, vascular endothelium, and smooth muscle.<sup>99, 103</sup> sEH was first identified in the 1970s in the cytosol of mouse hepatic cells by its ability to hydrolyze substrates with epoxide groups, such as the insect juvenile hormone and epoxy fatty acids derived from lipids, such as epoxystearate.<sup>104</sup> Since identified, sEH has accumulated more interest over time with over 100 unique structures solved – over 90% within the last decade.

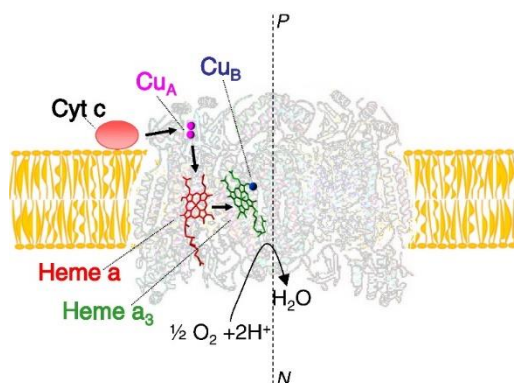
sEH catalyzes the hydrolysis of epoxides to diols by adding a water molecule (Figure 7). sEH efficiently cleaves lipid-derived signaling molecules, such as the Epoxyeicosatrienoic acids (EETs), which are epoxidized lipids formed from arachidonic acid by cytochrome p450.<sup>102</sup> sEH inhibition reduces EETs degradation, which can enhance the beneficial cardiovascular actions of EETs.<sup>105</sup> By dilating blood vessels, preventing the movement of vascular smooth muscle cells, and reducing inflammation, EETs have beneficial effects on the cardiovascular system. These natural lipids are degraded to less active diols by soluble epoxide hydrolase (sEH), so blocking this enzyme could increase the positive cardiovascular effects of EETs.<sup>105</sup> sEH inhibition has led to an increase in circulating EETs and a reduction in blood pressure in a number of animal models.<sup>106</sup> sEH role in diseases such as hypertension, pain, diabetes, and cardiac hypertrophy, makes its inhibition an ideal therapeutic target.<sup>102, 105, 107</sup>



**Figure 7. sEH epoxide hydrolysis.** sEH catalyzes the hydrolysis of epoxides to diols by adding a water molecule.

## 1.5 Cytochrome *c* oxidase

Cytochrome *c* oxidase (CcO) is a large protein complex located in the inner membrane of the mitochondria. It consists of 13 subunits, of which three are mitochondrial DNA-encoded and ten are nuclear DNA-encoded. The mitochondrial enzyme has two heme groups, cytochrome *a* and cytochrome *a*<sub>3</sub>, and two copper atoms, Cu<sub>A</sub> and Cu<sub>B</sub> (Figure 8). The cytochrome *a*<sub>3</sub> and Cu<sub>B</sub> together make a binuclear center where oxygen is reduced.<sup>108-111</sup> CcO is the final component of the mitochondrial electron transport chain (ETC) in eukaryotes and is regarded as one of the major regulation sites for oxidative phosphorylation. It catalyzes the transfer of electrons from cytochrome *c* to molecular oxygen, resulting in the formation of water and the pumping of protons across the inner mitochondrial membrane. This creates an electrochemical gradient that drives the synthesis of adenosine triphosphate (ATP) by ATP synthase, providing chemical energy for cellular functions.<sup>109, 112</sup> All living things on earth depend on energy generation, a vital process that uses adenosine triphosphate (ATP) as the main energy carrier. ATP stores energy from biochemical and metabolic reactions that involve direct or indirect transfer of phosphate groups.<sup>112</sup> Mitochondrial defects with clinical phenotypes such as stroke, hepatic failure, cardioencephalomyopathy, and Leigh's syndrome are often result from low CcO activity.<sup>109</sup>



**Figure 8. CcO dimer assembly with the redox binuclear center.** The enzyme structure is fitted into a bilayer (sketched). The electron transfer pathway from cytochrome *c* to the heme *a*<sub>3</sub>-Cu<sub>B</sub> via the Cu<sub>A</sub> site and heme *a* is indicated with small arrows. PDB ID: 2OCC. Adapted figure from Sarti *et al.*<sup>4</sup>

### 1.5.1 Cytochrome *ba*<sub>3</sub> oxidase

*ba*<sub>3</sub> cytochrome *c* oxidase (*ba*<sub>3</sub> CcO) is a heme-copper oxidase embedded in the plasma membrane of some bacteria that catalyzes the conversion of molecular oxygen to water (1.2), coupled with proton pumping.<sup>113</sup> Bacterial oxidases have a simpler structure than mitochondrial oxidases, consisting of only a few subunits that are encoded and assembled by one genetic system and a small number of essential biogenesis factors.<sup>114</sup> The *ba*<sub>3</sub> CcO from *Thermus thermophilus* consist of three subunits, and is expressed at high temperatures and under low oxygen supply.<sup>115-117</sup> *Thermus thermophilus* thrives under lower oxygen conditions and can grow at high temperatures ranging from 50 – 80 °C.<sup>118</sup> *ba*<sub>3</sub> CcO has four redox-active sites, of which three are in subunit I and include a low-spin heme *b* and *a* binuclear catalytic center composed of a high-spin heme *a*<sub>3</sub> and Cu<sub>B</sub>. The fourth site is the di-nuclear Cu<sub>A</sub> center, in subunit II, which acts as the main electron acceptor from soluble cytochrome *c*<sup>552</sup>.<sup>119</sup> In *Thermus thermophilus*, cytochrome *c*<sup>552</sup> delivers electrons one-by-one to cytochrome *ba*<sub>3</sub> oxidase at the positive side of the membrane, while the protons are taken up from the negative side of the membrane, creating a proton electrochemical gradient (1.3). The electrons are transferred from Cu<sub>A</sub> at the docking site through the low-spin heme *b* to the high-spin heme *a*<sub>3</sub>-Cu<sub>B</sub> binuclear center.<sup>113</sup> In addition to the enzymatic O<sub>2</sub> reduction, *ba*<sub>3</sub> CcO involved in proton pumping across the membrane i.e., the movement of protons that are not involved in the reduction of O<sub>2</sub> to H<sub>2</sub>O (1.4). The ratio of proton pumping by *ba*<sub>3</sub> CcO is lower than that of *aa*<sub>3</sub> CcO (~ 0.5 H<sup>+</sup>/*e* instead of 1 H<sup>+</sup>/*e*).<sup>113, 119-122</sup> The proton pumping mechanism of CcO is a long-standing question in bioenergetics. Despite extensive studies, the mechanism is still elusive. Structural biology methods may provide new insights into the structure-function correlation of proton pumping in CcO.



## 1.6 Scope of this thesis

This thesis focuses on the microcrystallization of proteins and structural studies involving three different proteins: Cytochrome P450 3A4, soluble Epoxide Hydrolase, and *ba*<sub>3</sub>-type Cytochrome *c* oxidase. The thesis is structure into three chapters:

**Chapter 1:** Introduces proteins with some historical context and why understanding protein structures is important, followed by a brief look into serial X-ray crystallography and some the methods used, and an introduction to the three proteins further explored in the thesis.

**Chapter 2:** Describes the methodology followed in this thesis for production, purification, and crystallization of proteins. This chapter also describes some crystallography, absorbance spectroscopy, and X-ray scattering techniques.

**Chapter 3:** Includes the results and discussion from the four research papers presented in this thesis.

In **paper I**, we describe the different methods explored for producing large quantities of well-diffracting CYP3A4 microcrystal, we solve the room-temperature structure of CYP3A4, and we discuss temperature-induced structural difference when compared to a cryogenic structure.

In **paper II**, we explore methods for producing well diffracting sEH microcrystals, we report seven RT sEH-ligand complexes obtained by serial crystallography, and we discuss temperature-induced structural difference when compared to cryogenic sEH-ligand complexes.

In **paper III**, we present the structural changes of reduced CO-bound *ba*<sub>3</sub>-type cytochrome *c* oxidase (CcO) from *Thermus thermophilus* and the time-resolved serial femtosecond crystallography results after photodissociation of carbon monoxide (CO) from reduced CO-bound *ba*<sub>3</sub>-type CcO.

In **Paper IV**, we confirmed the binding of Ketoconazole and Lovastatin to CYP3A4 in solution, and investigated CYP3A4 oligomeric state and ligand-induced conformational changes in solution by X-ray scattering.

# Chapter 2

## Methodology

### 2.1 Protein production

The theoretical steps for producing a recombinant protein are relatively simple. You take your gene of interest, insert it into an appropriate expression vector, transfer it into the host of choice, induce and then, the protein is ready for purification and characterization. However, in reality, many things can go wrong. Poor growth of the host, inclusion body formation, protein inactivity, and even no protein production at all are some of the challenges often encountered along the pipeline. The selection of the host cell whose protein synthesis machinery will generate the valuable protein determines the framework of the whole process.<sup>123</sup>

#### 2.1.1 CYP3A4 expression in *Escherichia coli* XL1-Blue

*E. coli* XL1-Blue is a strain of bacteria that is commonly used for cloning and protein production. *E. coli* XL1-Blue is derived from *E. coli* K-12 and has several mutations that make it suitable for these applications, such as *recA1*, which reduces recombination events; *endA1*, which improves the quality of plasmid DNA; *lacIqZΔM15*, which allows blue-white color screening for recombinant plasmids; and *Tn10*, which confers tetracycline resistance. *E. coli* XL1-Blue also contains the  $\lambda$ DE3 lysogen, which carries the T7 RNA polymerase gene under the control of the *lacUV5* promoter, allowing for strong expression of genes under the T7 promoter. *E. coli* XL1-Blue can be transformed with plasmid or lambda vectors to produce high-quality plasmid DNA or recombinant proteins. Some challenges may arise when using *E. coli* XL1-Blue for cloning or protein production, such as protein folding, protein stability, protein solubility, protein toxicity, and protein purification. To overcome these challenges, various strategies have been developed, such as optimizing gene sequence, modifying expression vector, engineering host strain, improving

culture conditions, and using alternative inducers.<sup>124</sup> A plasmid encoding human CYP3A4 was previously obtained from the LINK scheme (University of Dundee, Dundee, U.K.) and the coding sequence, modified according to Williams et al.<sup>125</sup> with a 6-polyhistidine-tag in the N-terminal, and was subcloned into the expression vector pCWori by Ekroos et al.<sup>78</sup> pCWori is a low copy number expression vector that confers ampicillin resistance to transformed cells. It contains the tac promoter, which allows for Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)-inducible expression of genes under the control of the lac operator. It also contains the lacIQ gene, which encodes a repressor protein that binds to the lac operator and prevents transcription in the absence of IPTG.<sup>126, 127</sup> CYP3A4 was expressed in *E. coli* XL1-Blue with IPTG and  $\delta$ -Aminolevulinic acid ( $\delta$ -ALA) were added to induce protein expression and heme biosynthesis respectively. IPTG is a molecular mimic of allolactose, a natural metabolite of lactose that activates the lac operon. Unlike allolactose, IPTG is not hydrolyzed by beta-galactosidase, so it does not get consumed or degraded by the cells. This means that IPTG can maintain a constant level of induction during an experiment, which makes it ideal for protein overexpression.<sup>127</sup>  $\delta$ -Aminolevulinic acid is a natural compound that is involved in the synthesis of porphyrins, which are essential for the formation of heme.<sup>128</sup>

## 2.2 Protein purification

Protein purification is a series of processes that aim to isolate a specific protein of interest from a complex mixture of proteins. The term isolation implies the attainment of a pure population of the target protein, which is often challenging and impractical. Therefore, the term enrichment is more appropriate, as it indicates the increase in the relative abundance of the target protein in the final product. Protein purification relies on the differential properties of the target protein that distinguish it from the other proteins in the complex mixture.<sup>129</sup> These properties include electrostatic interactions (based on the isoelectric point and the charge of the protein at a given pH), solubility in various

salt concentrations, molecular size, and specific binding affinity to a ligand. These properties are exploited in different stages of protein purification.

### 2.2.1 Cell disruption and centrifugation

The initial step of each purification procedure is the lysis of the cells that harbor the protein unless the protein is secreted by the organism into the extracellular medium, i.e., protein secreting insect cells.<sup>130</sup> The choice of the method for cell disruption depends on the stability of the cells and the sensitivity of the protein. Some of the common methods are:

*Freeze-thaw cycles* involves freezing the cell suspension in a mixture of dry ice and ethanol or a freezer, followed by thawing the sample at room temperature or 37°C. Multiple cycles are necessary for efficient lysis, and the process can be quite lengthy.

*Sonication* uses pulsed, high frequency acoustic waves to disrupt and break down cells, bacteria, spores, and minced tissue. The acoustic waves are generated by an instrument with a vibrating probe that is submerged in the cell suspension. The probe transfers mechanical energy to the sample, causing the formation and collapse of tiny vapor bubbles. This creates shock waves that propagate through the sample. To avoid excessive heating, the sample is exposed to ultrasonic treatment in multiple short pulses while immersed in an ice bath.

*Liquid homogenization* is the most common cell disruption technique for small volumes and cultured cells. This technique lyses cells by applying shear stress to the cell membranes, which occurs when the cell or tissue suspension is forced through a narrow gap. The French press is often the liquid homogenizer of choice for breaking bacterial cells mechanically, as it allows for higher volumes ranging from 40 – 250 mL.<sup>131</sup>

*Bead mill homogenization* uses beads to disrupt samples in a tube. The homogenizer unit shakes the tubes rapidly, creating collisions between the samples, the beads, and the tube walls. These collisions result in the homogenization of the samples.

*Manual grinding* uses pestle and mortar to pulverize tissues frozen in liquid nitrogen. This is the most prevalent technique for disrupting plant cells, as it is quickest and most effective way to access plant proteins and DNA, due to the high tensile strength of the cell wall polysaccharides, such as cellulose.<sup>131</sup>

*Solution-based cell lysis* i.e., detergent permeabilization is a gentler and simpler option than physical disruption of cell membranes. Solution-based cell lysis dissolves the lipid barrier of cells by solubilizing proteins and disrupting the interactions among lipids, proteins, and lipid-proteins. Proteins in the solution are digested by proteases released during cell lysis. To prevent this, protease inhibitor is added to the sample before lysis.

### 2.2.3 Liquid chromatography

Liquid chromatography (LC) is a technique that separates the components of a mixture based on their differential interactions with a liquid solvent and a solid material. The technique is classified into different types according to the physical states of the solvent and the material, which are called the mobile phase and the stationary phase, respectively. The choice of the mobile and stationary phases depends on the properties of the mixture and the desired separation.<sup>132</sup> In protein production, the mobile phase is the protein solution, and the stationary state depends on the protein properties being exploited. Liquid-solid column chromatography, the most common type of chromatography, uses a liquid mobile phase that flows through the solid stationary phase, carrying the separated components along with it. The liquid mobile phase is pressurized by pumps and passes through a column filled with a solid adsorbent material. The components of the mixture have different affinities for the adsorbent material, which results in different retention times and separation efficiencies as they elute from the column. Three different LC methods were used in the production of proteins and are listed in order:

*Immobilized metal affinity chromatography (IMAC)* is a type of affinity chromatography that uses metal ions to selectively bind to proteins or peptides that have an affinity for them. IMAC is often used to purify

proteins that have a polyhistidine-tag, which is a short sequence of histidine residues that can bind strongly to metals such as nickel, cobalt, zinc, or copper.<sup>133</sup> IMAC works by immobilizing metal ions on a solid support, such as agarose or polyacrylamide beads, using a chelating agent, such as iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA).<sup>134</sup> The metal ions form coordinate covalent bonds with the nitrogen atoms of the chelating agent, creating a stable complex. The beads are then packed into a column and equilibrated with a buffer that contains a low concentration of imidazole, which competes with histidine for binding to the metal ions. The sample containing the target protein is then loaded onto the column and allowed to bind to the metal ions. The non-specifically bound proteins are then washed away with the same buffer, while the target protein remains attached to the column. The target protein can then be eluted by increasing the concentration of imidazole in the buffer, which displaces the histidine from the metal ions. IMAC is a powerful technique for protein purification because it is highly specific, efficient, and versatile. The next LC method used was in exploit of the isoelectric properties and protein charge. The high salt content of the elution buffer used in previous affinity chromatography step resulted in some interference and required a buffer exchange to proceed. Buffer was exchanged by dialysis.

*Dialysis buffer exchange* is a process of changing the buffer solution of a sample by using a semi-permeable membrane. It is often used to remove small molecules or contaminants from proteins or nucleic acids, or to prepare them for different downstream applications. The sample is placed in a dialysis device that is immersed in a large volume of buffer solution (dialysate). The small molecules in the sample can pass through the membrane pores, while the large molecules are retained. The dialysis process resulted in some protein precipitate formation that was separated by centrifugation before the next LC step.

*Ion exchange chromatography (IEC)* is a method that separates molecules based on their charge. The technique exploits the different affinities of the molecules for a resin that has a fixed charge. The resin can be either positively or negatively charged, depending on the type

of ion exchange chromatography. Amino acids are zwitterionic compounds that have both positive and negative charges on their functional groups.<sup>135</sup> The net charge of a protein depends on the pH of the solution, which affects the ionization state of the amino acids. The protein has a net zero charge at a specific pH value, which is called the isoelectric point (pI).<sup>136</sup> When the pH is higher than the pI, the protein has a negative net charge and binds to an anion exchange resin, which has a positive charge. When the pH is lower than the pI, the protein has a positive net charge and binds to a cation exchange resin, which has a negative charge. This allows the separation of molecules that have similar properties but different charge states. The molecular size of the protein is exploited in the final LC purification step.

*Size exclusion chromatography (SEC)* is a method of separating molecules based on their size and shape. It uses a column filled with porous beads that act as a sieve. The pore sizes of the beads are used to estimate the size of the molecules. Molecules that are larger than the pores of the beads cannot enter them and are eluted first. Molecules that are smaller than the pores of the beads can enter them and are eluted later.<sup>137, 138</sup>

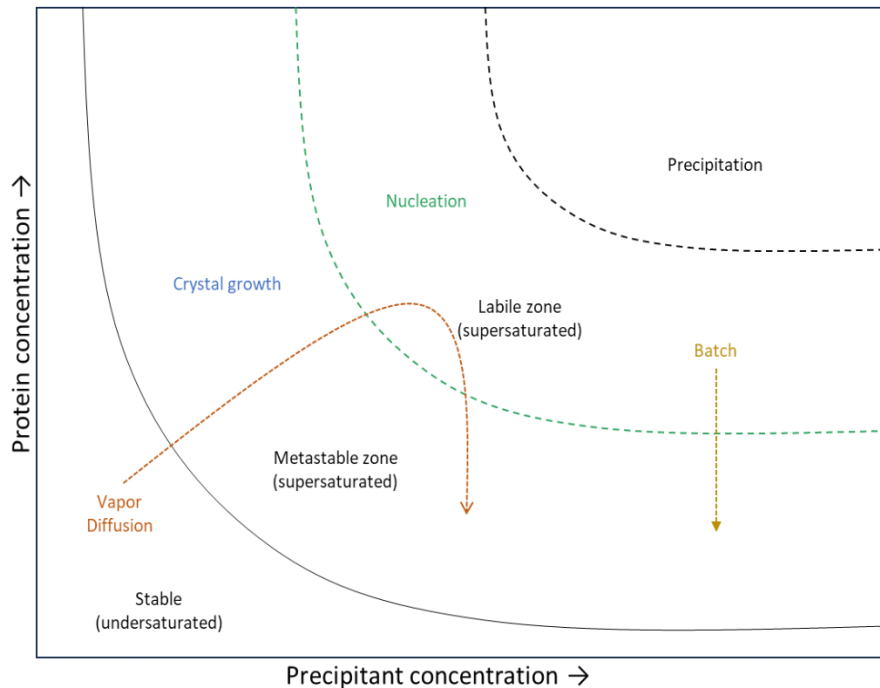
## 2.3 Protein crystallization

Protein crystallization is the process of formation of a regular array of individual protein molecules stabilized by crystal contacts. Proteins are usually amorphous and unstable in the solid state and tend to lose their native structure and function. However, proteins in crystalline form can maintain their conformation and activity with enhanced stability. The aim of crystallization is to obtain a well-ordered crystal that is free of contaminants and large enough to produce a diffraction pattern when exposed to x-ray. This diffraction pattern can be used to determine the three-dimensional structure of the protein. Proteins can be crystallized under suitable conditions that favor the balance between solubility and supersaturation. Protein crystallization is a complex and challenging task that depends on many factors that influence the protein behavior and crystal quality. It is a process of exploring, in a systematic way, the

ranges of the parameters that affect crystal formation, finding one or more sets of conditions that produce some crystals, and then optimizing the conditions to obtain the best crystals. This is usually done by performing a large number of crystallization experiments, analyzing the results and using the information to improve the conditions in subsequent experiments.<sup>139</sup>

### 2.3.1 Supersaturation, nucleation, and crystal growth

Protein crystallization requires the creation of a supersaturated state. Supersaturation is a non-equilibrium state of a solution when it contains more solute than the solvent can hold at a given temperature and pressure. This means that the solution is unstable, and the excess solute can precipitate out of the solution if disturbed or seeded with a crystal. Supersaturation can occur with solids, liquids or gases dissolved in a solvent. For example, water can become supersaturated with sugar<sup>140</sup>, carbon dioxide<sup>141</sup>, salt<sup>142</sup>, etc. In protein crystallization the goal is to change the conditions so that the protein solubility in solution is significantly decreased, resulting in a supersaturated solution (Figure 9). This is achieved by adding mild precipitants such as neutral salts or polymers, and by adjusting parameters that affect the solubility, such as temperature, ionic strength, and pH. When a stable nucleus forms in a supersaturated solution, it will continue to grow until the system reaches equilibrium. The crystal growth or precipitate formation depends on the non-equilibrium forces and the degree of supersaturation in the solution.<sup>139</sup> Microcrystals were obtained by vapor diffusion and batch mode crystallization. In vapor diffusion, a drop containing a mixture of sample and reagent is equilibrated with a liquid reservoir of reagent in a closed system. The drop has a lower reagent concentration than the reservoir initially. To reach equilibrium, water vapor evaporates from the drop and condenses in the reservoir. This causes the sample to increase in relative supersaturation. The concentration of both the sample and the reagent rises as water leaves the drop for the reservoir. Equilibrium is achieved when the reagent concentration in the drop is approximately equal to that in the reservoir.<sup>143</sup> The batch method involves mixing protein with the required amount of precipitant to directly induce nucleation.<sup>144</sup>



**Figure 9. Phase diagram schematic.** The solubility curve separates undersaturated and supersaturated regions. Supersaturated region is further divided into three zones: metastable, labile and precipitation. In the metastable region, nuclei grow into crystals without nucleation. In the labile region, both nucleation and growth can occur. The last region, at very high supersaturation, is called the precipitation region, where precipitation is most likely. Crystal growth requires a supersaturated solution of the protein of interest. Creating such a solution is the first goal in protein crystallization. Schematic also shows the kinetic trajectories of vapor diffusion and batch crystallization.

### 2.3.2 Membrane protein crystallization

Membrane proteins are essential for many cellular functions and diseases. To design drugs that target them, their molecular mechanisms need to be understood. This requires obtaining their three-dimensional structures by crystallography, which is challenging because of the difficulty of growing high-quality crystals.<sup>145</sup> Membrane proteins can be crystallized several methods that falls within two main approaches.

*In surfo* crystallization is a method for growing crystals of membrane proteins in detergent micelles.<sup>146</sup> It is based on the vapor diffusion technique, which involves equilibrating a drop of protein solution with

a reservoir of precipitant solution. The drop and the reservoir are separated by a vapor phase, which allows water to evaporate from the drop and increase its concentration until supersaturation is reached. At this point, crystals may start to form and grow in the drop. *In surfo* crystallization requires finding the optimal detergent and buffer conditions for solubilizing and stabilizing the membrane protein, as well as screening many crystallization conditions to identify the ones that promote crystal formation and quality. Some advantages of *in surfo* crystallization are that it is easy to perform, compatible with most crystallization robots and screens, and similar to the method used for soluble proteins. Some disadvantages are that it may disrupt the native conformation and function of membrane proteins, cause aggregation and degradation, and result in fragile crystals with high solvent content.<sup>147</sup>

*In meso* crystallization is a method for growing crystals of membrane proteins in a lipidic environment that mimics the natural membrane.<sup>148</sup> It involves mixing the protein with a lipid, such as monoolein, to form a bicontinuous cubic phase (also called lipidic cubic phase or LCP) that has a three-dimensional network of water channels surrounded by lipid bilayers. The protein molecules can diffuse and align in the water channels and form crystals when the conditions are favorable. *In meso* crystallization has several advantages over conventional detergent-based crystallization, such as preserving the native conformation and function of membrane proteins, reducing aggregation and degradation, and allowing for in situ data collection.<sup>147, 149-151</sup>

## 2.4 Crystallography

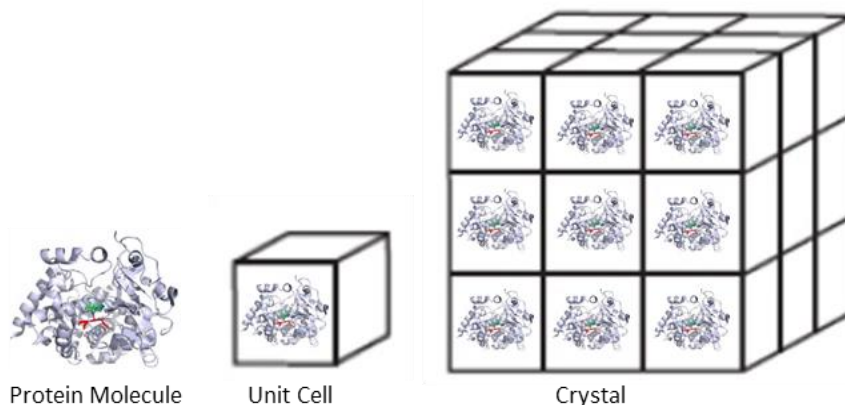
Crystallography methods use radiation beams to probe the atomic and molecular structure of a substance by analyzing its diffraction patterns. The most common type of radiation is X-rays, but neutrons and electrons are also used. Crystallography is suitable for studying atomic structures because optical microscopy cannot resolve them due to the large wavelength of visible light compared to the size of atoms and bonds. The shorter wavelength of radiation allows for better resolution

of structures at this scale. A major difference between crystallography and optical methods is that the beam cannot be focused to form images directly. Instead, the structure is reconstructed from the diffraction pattern using various techniques.<sup>152</sup>

### 2.4.1 X-ray crystallography

X-ray crystallography is a method that uses X-rays, which are a type of electromagnetic radiation with very short wavelengths (usually between 0.05 and 0.3 nanometers), to study the structure of molecules that form crystals. Crystals are solids that have a regular and repeating pattern of atoms or ions. X-rays are very weakly scattered by individual molecules, and the scattered signal would be overwhelmed by the background noise, which includes X-rays bouncing off air and water molecules. A crystal is a solid that has many molecules aligned in the same way, so that their scattered X-rays can combine constructively and increase the signal strength. The repeating portion of the crystal lattice that shows the three-dimensional pattern of the entire crystal is called a unit cell (Figure 10). There are seven basic types of unit cells, which differ in their shape and symmetry. The shape of a unit cell can be described by a parallelepiped with three edge lengths  $a$ ,  $b$  and  $c$ , and three angles  $\alpha$ ,  $\beta$  and  $\gamma$  between them. A crystal can be seen as a device that boosts the X-ray scattering signal from a single molecule. When X-rays pass through a crystal, they scatter in different directions and form a specific pattern on a detector. This pattern, called the diffraction pattern, reveals information about the shape and arrangement of the atoms or ions in the crystal. X-rays were discovered by Wilhelm Röntgen in 1895, and he received the first Nobel Prize in Physics for his work in 1901. In 1912, Max von Laue, Walter Friedrich, and Paul Knipping showed that X-rays could be diffracted by inorganic crystals, and Von Laue won the Nobel Prize in Physics in 1914 for this discovery. Diffraction is the phenomenon of waves, such as X-rays or light, bending and interfering with each other when they encounter an obstacle, such as a crystal. William H. Bragg and William L. Bragg, a father and son team, used X-ray diffraction to determine the structure of sodium chloride, a common salt, in 1913. They were awarded the

Nobel Prize in Physics in 1915 for their contribution to the field of X-ray crystallography.<sup>153</sup>

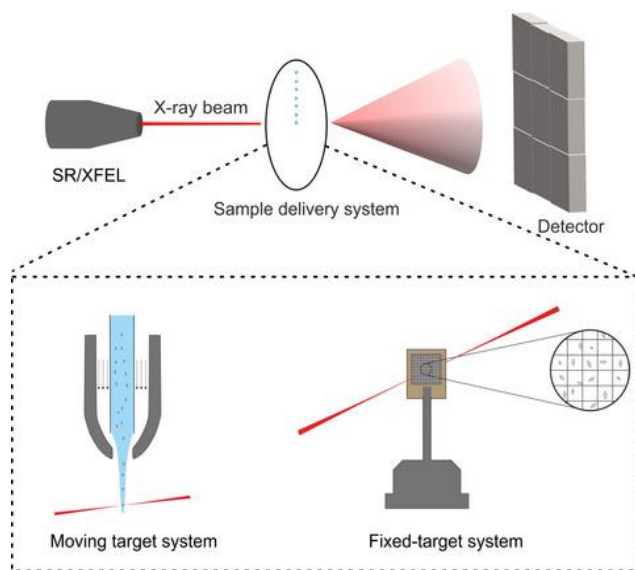


**Figure 10. A simplified representation of a unit cell.** Protein molecule represents the crystal's asymmetric unit (unique part of the crystal structure). A crystal's asymmetric unit may contain a portion of a biological assembly, a whole biological assembly, or multiple biological assemblies.

## 2.4.2 Serial crystallography

X-ray crystallography using single crystals is a powerful technique for structural biology. This technique uses a goniometer to rotate a single crystal in a cryo-loop while exposing it to X-rays. This way, the crystal covers the entire symmetry and produces complete diffraction data. This has some limitations, and one of them is radiation damage, which occurs when X-rays interact with the atoms in the crystals and cause ionization and structural changes. Radiation damage reduces the quality of the diffraction data and affects the accuracy of the structure determination. To minimize radiation damage, data collection is usually performed at cryogenic temperatures i.e., the crystal is flash-frozen in liquid nitrogen before mounting and cooled by a cryo-stream during data collection. However, cryo-freezing may damage the crystal lattice and require the use of cryo-protectants, which may interfere with the protein structure. Cryo-freezing may not completely prevent radiation damage and can limit the information on protein molecular dynamics.<sup>154</sup> Radiation damage can also affect the high-resolution

regions of the data if the atomic movements are too large.<sup>155</sup> To overcome these challenges, new serial crystallography (SX) methods have been developed. SX is a method that uses intense X-ray pulses from free-electron lasers (XFELs) or synchrotrons to collect data from thousands of microcrystals in random orientations. Each crystal is exposed to X-rays only once and for a very short time, which reduces radiation damage and allows data collection at room temperature or near-physiological temperature, which can reveal the molecular mechanisms of molecules and better understanding of function. To deliver the crystals, one method is the GDVN-based liquid jet injector, which has been used successfully at the XFEL facility with a high frequency of X-ray pulses. Another method is the fixed-target scanning method, which uses fewer crystals and causes less physical stress to them than injection methods (Figure 11).<sup>35, 156</sup>



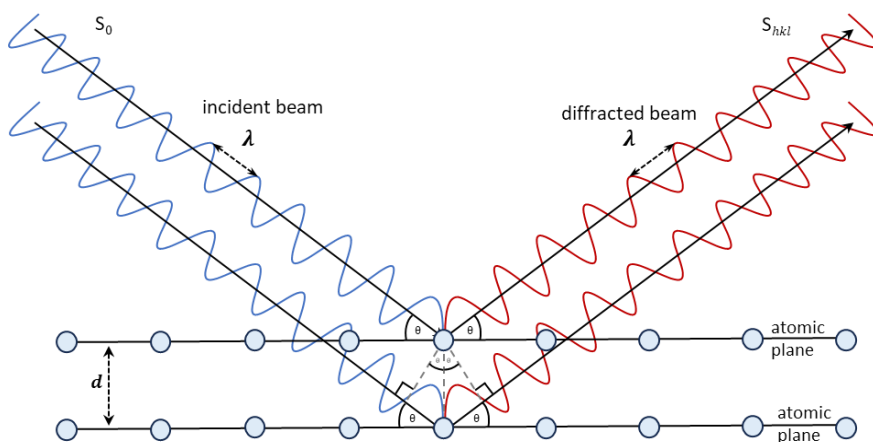
**Figure 11. Schematic representation of serial crystallography setup.** The X-ray beam hits the moving crystals (flowing crystals in GDVN-based liquid jet injectors/raster grid scanning in fixed-target systems) once and diffraction images are collected by the detector. *Adapted figure from Zhao et al.*<sup>3</sup>

## 2.4.2 X-ray diffraction

The electrons in a crystal can be seen as sources of small waves of X-rays. When these waves combine, they can interfere with each other: they can be in sync, out of sync, or somewhere in between. This depends on the direction of the incoming and outgoing waves and the relative positions of the electrons. The total distance from the X-ray source to the detector affects the interference. If the difference between the distances travelled by two waves via different electrons is a whole number of wavelengths, then the waves will scatter in sync and their intensities will add up. If it is a half-number of wavelengths, they will scatter out of sync and cancel out. In a crystal, there are flat surfaces that pass through the points where the atoms are arranged. These surfaces are called Miller planes and represented with indices  $(h,k,l)$ . When a beam of X-rays hits a crystal, some of the X-rays are reflected by these planes. The reflected X-rays form a pattern that can be used to study the structure of the crystal. However, not all Miller planes reflect X-rays. Only the ones that satisfy Bragg's law (Figure 12) can produce a visible pattern. Bragg's law is a rule that relates the wavelength of the X-rays, the angle of incidence, and the distance between the Miller planes. We can simplify the conditions for scattering in sync by imagining that the waves are reflected by flat surfaces that go through the atoms. The relation between the scattering angle and the distance between these surfaces is given by Bragg's law (2.1). Bragg's law is a formula that relates the wavelength of X-rays, the angle of incidence, and the distance between atomic layers in a crystal. It is written as:

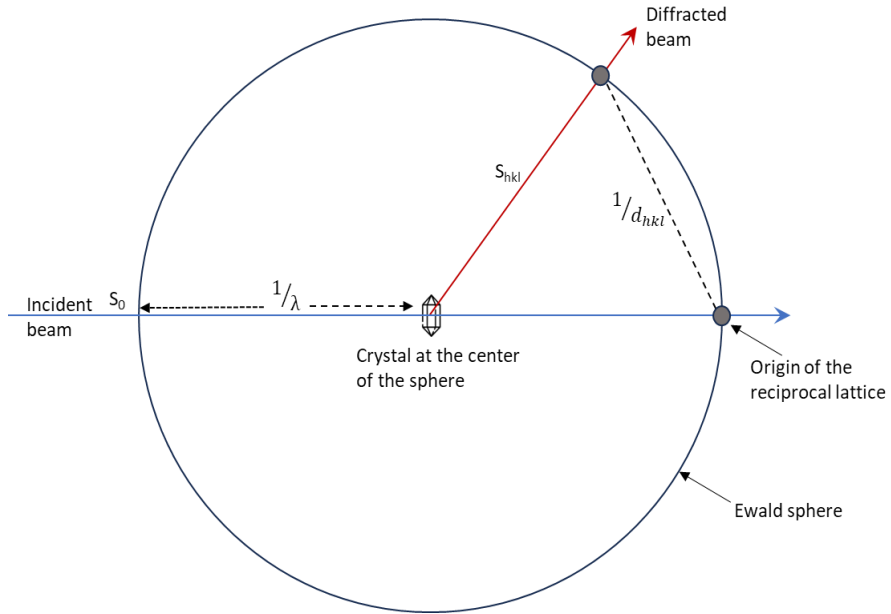
$$n\lambda = 2d_{hkl}\sin\theta \tag{2.1}$$

where  $n$  is an integer (diffraction order),  $\lambda$  is the wavelength,  $d$  is the distance between layers (path difference), and  $\theta$  is the angle of incidence.



**Figure 12. Representation of Bragg's law in a crystal.** Beam of parallel X-rays penetrating a set of parallel lattice planes of spacing  $d$  and at an angle of incidence  $\theta$ . The 'reflected' waves will combine to form a diffracted beam. Atoms are shown as blue spheres along atomic planes.

The relationship between the wavelength of the incident and diffracted X-ray beam, diffraction angle for a given reflection, and the reciprocal lattice of a crystal is best represented by the Ewald sphere (Figure 13). The Ewald sphere is a concept that helps to understand how X-rays are diffracted by crystals. It is a sphere with a radius of  $1/\lambda$ , where  $\lambda$  is the wavelength of the X-rays. The center of the sphere is the origin of the reciprocal lattice, which is a mathematical representation of the crystal structure. The direction of the incoming X-rays is along a radius of the sphere. When a reciprocal lattice point lies on the surface of the sphere, it means that the X-rays are reflected by a plane of atoms in the crystal at a certain angle. This angle satisfies Bragg's law, which relates the wavelength, the interplanar spacing, and the angle of incidence. The reflected X-rays form a diffraction pattern that can be used to determine the structure of the crystal. The Ewald sphere was introduced by Paul Peter Ewald, a German physicist and crystallographer, in 1913. He used it to explain the theory of X-ray interference in crystals.<sup>153</sup>



**Figure 13. Two-dimensional representation of the Ewald sphere.** The Ewald circle represents in reciprocal space all the possible points where planes (reflections) could satisfy the Bragg equation. In 3-dimensions an Ewald Sphere represents all the possibilities.

The amplitude ( $|F_{hkl}|$ ) and phase ( $\phi_{hkl}$ ) that represents the overall scattering from a particular set of Bragg planes ( $hkl$ ) is termed the structure factor and is the key parameter for solving and refining crystal structures. It reflects the dependence of the reflection intensity on the atomic arrangement of the crystal, regardless of the experimental method or conditions. The structure factor ( $F_{hkl}$ ) represented in the Fourier sum:

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (2.2)$$

where  $F_{hkl}$  is the structure factor,  $hkl$  is the scattering vector,  $n$  is the number of atoms in the unit cell,  $f_j$  is the scattering factor of the  $j$ -atom at coordinates  $x$ ,  $y$ ,  $z$ . The diffraction intensities reflect the electron density distribution ( $\rho$ ) in the unit cell, and the diffraction phenomenon can be expressed by the equation:

$$F_{hkl} = \iiint V \rho_{xyz} e^{2\pi i(hx+ky+lz)} \partial x \partial y \partial z \quad (2.3)$$

where  $\rho_{xyz}$  is the electron density at coordinate (x, y, z) and V represents the volume of the unit cell. The collection of structure factors for all the reflections ( $hkl$ ) allows the reconstruction of the three-dimensional electron density ( $\rho$ ) map of the crystal, which represents its structure.

This is obtained by an inverse of the Fourier sum:

$$\rho_{xyz} = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz+\varphi_{hkl})} \quad (2.4)$$

where modulus sign ( $||$ ) around F means F is a phaseless quantity and  $\varphi$  signifies phases.<sup>157</sup>

### 2.4.3 Data collection, processing, and refinement

The crystal scatters the X-ray beam into discrete spots, also known as reflections. The experimental data consist of the amplitudes of these reflections, while the phases are obtained by Molecular Replacement (MR), a method that uses a known structure as a model. The resolution of the data is determined by the distance of the reflections from the center of the detector, with higher resolution corresponding to larger distances. However, the signal-to-noise ratio ( $I/\sigma$ ) decreases with increasing resolution, and there is a limit beyond which the reflections are too weak to be measured. To improve the signal-to-noise ratio in X-ray diffraction experiments with synchrotron radiation, it is important to match the X-ray beam size and the crystal size, and to minimize the extra scattering materials in the X-ray path. This way, the diffraction signal is enhanced by illuminating the whole crystal volume with less X-ray dose, and the background scattering is reduced by removing unnecessary materials around the crystal.

In SSX with injector methods, the optimal combination of X-ray beam size, crystal size, and stream velocity should be chosen to ensure that

the crystal is fully exposed to X-rays while it traverses the beam. The injector stream diameter should be close to the crystal size, but not too small to avoid clogging. In SSX with fixed-target methods, various systems have been developed to use low-scattering materials and less crystallization media around crystals. The goniometer or scanning stage, which holds the fixed-target samples, allows precise control of crystal characterization and data collection. Crystals can be located with a grid scan, and a still image can be collected. Fixed-target methods also provides the ability for obtaining rotation data over a small angle by slight oscillation of the goniometer.<sup>1, 158-160</sup>

The quality of the data can be assessed by various statistics, such as R-factors,  $I/\sigma$ , and completeness. These statistics are also used to select the reflections that are reliable enough for the subsequent map calculation, model building, and refinement.<sup>161</sup> The main objective of collecting diffraction data is to acquire a complete data set with high precision and accuracy. Precision refers to the consistency of the measured intensities, as indicated by the values of  $CC_{1/2}$  and  $I/\sigma$  for the merged data, especially in the highest resolution shell. Accuracy refers to the closeness of the measured intensities to their true values.<sup>160</sup>

$CC_{1/2}$  is the Pearson correlation and is a useful statistic for detecting weak signals in the high-resolution shells of crystallographic data sets. It is calculated by splitting the intensity measurements of each reflection into two random and independent subsets, and then computing the Pearson correlation coefficient between the intensities of the two subsets. The  $CC_{1/2}$  is reliable and independent of scaling issues, or invariant to scaling if scaling is applied.<sup>162</sup>

Data processing involves two main stages: integration and scaling and merging. Integration consists of four steps: spot finding, indexing, refinement and integration. Scaling and merging are usually treated as a single step. Integration provides useful information about the crystal orientation, unit cell parameters, and spot intensities. Scaling and merging provide the most useful statistics about the data quality, such as completeness, redundancy, resolution, R-factors,  $I/\sigma$ , and  $CC_{1/2}$ . Indexing is the process of determining the crystal's unit cell

parameters, orientation, and symmetry from the diffraction pattern. It requires accurate knowledge of the direct beam position, the crystal-detector distance, and the radiation wavelength. Errors in these parameters can cause indexing failures. Integration is the process of estimating the spot intensities from the image pixels. It requires subtracting the background noise, which is present both around and under the spots, from the pixel counts. This is done by interpolating a background plane from the pixels surrounding each spot. Scaling and merging are the process of adjusting and combining the measured intensities to account for experimental errors and crystal symmetry. For 2D integration, it also involves merging partial reflections into complete ones. This process evaluates the quality and resolution of the data collection and processing and determines the true symmetry of the crystal.<sup>163</sup> The integration, scaling and merging of a serial crystallography (SX) dataset may be performed using a variety of programs within CrystFEL, e.g., *indexamajig* for batch indexing, integration and data reduction by accessing indexing programs i.e. *Mosflm* and *XDS*, *ambigator* for resolving indexing ambiguities, *cell explorer* for examining the distribution of unit cell parameters, *partialator* for full scaling and post-refinement accurate merging of data and outlier rejection, etc.<sup>164</sup>

The end result of an X-ray structure determination is the electron density map. Electron density maps are graphical representations of the agreement between the structural model and the experimental data in X-ray crystallography. The maps depict the probability distribution of electrons in space, derived from the combination of the model parameters and the measured diffraction intensities. Two types of electron density maps are commonly used in X-ray crystallography: the 2Fo-Fc map and the Fo-Fc map. The Fo-Fc map, also called a difference or omit map, displays the regions of electron density that are not explained by the model or that are overrepresented by the model. The 2Fo-Fc map incorporates the Fo-Fc map and shows the electron density around the atoms in the model. These two maps help to refine and validate the model. Even in high-resolution structures, some parts of the model may have poor electron density, indicating that they adopt

multiple conformations in the crystal. This is often observed in flexible side chains or surface loops of the model.<sup>165</sup> Structure factor amplitudes and phases are required to calculate the electron density of an experimental dataset, but only the structure factor can be obtained from the experimental dataset. Theoretical phases may be estimated by molecular replacement using software (*Phaser*, *Molrep*, *MrBump*) assuming similar known structures exist.

Molecular replacement (MR) is a technique that uses a known structure of a molecule as a template to determine the structure of a similar molecule in a crystal. MR solves the phase problem in crystallography by using the phases of the template structure to estimate the phases of the target structure. The application of MR has increased with the growth of the database of known structures. The phase problem is a major challenge in x-ray crystallography. The x-ray detector only measures the intensities of the diffracted waves, but not their phases. Each diffracted wave or structure factor has an amplitude and a phase, which are both essential for the reconstruction of the electron density map of the crystal. The amplitude can be obtained by taking the square root of the intensity, but the phase is lost during the data collection. MR accounts for up to 70% of deposited macromolecular structures and has the benefits of being fast, cheap and highly automated.<sup>166</sup> The atomic model is then taken through several iterations of reciprocal-space refinement using programs (*REFMAC*, *SHELX*, *BUSTER*) and manual model corrections using *Coot* to improve estimated phases. Refinement is the process of adjusting the atomic coordinates to improve the agreement between the calculated and observed diffraction data and is validated by R-factor. The work R-factor ( $R_{\text{work}}$ ) measures how well the calculated model structure factor amplitudes ( $F_c$ ) agree with the observed diffraction data ( $F_o$ ).

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad (2.5)$$

R-factor is not a good measure of accuracy: it is not independent, as it is the target function that is minimized during refinement, and can be

manipulated by using too many refinement parameters, which leads to overfitting.<sup>167</sup> Therefore, the free R-factor ( $R_{\text{free}}$ ) is often used to validate the refinement process. The free R-factor is computed from 5% of the reflections that are not used in refinement, and thus provides an independent measure of the refinement progress. The decrease of the R-factors during refinement is more important than their absolute values. Protein structures that are well refined usually have R-factors between 15% and 25%, depending on the resolution and data quality. However, for small organic compounds, R-factors as low as 2–5% are common. To ensure the validity and reliability of R-free, it is essential to maintain the same set of test reflections throughout the entire structure determination and refinement process.<sup>167-169</sup> *CCP4* is a comprehensive software packages that guides the workflow for the structure determination and refinement, and contains most of the programs mentioned in this section.<sup>170, 171</sup>

## 2.5 Absorbance spectroscopy

Absorbance spectroscopy is a technique that uses a spectrometer to measure the amount of light absorbed by a sample at different wavelengths. It can be used to determine the electronic structure and the energy gap between the molecular orbitals of the sample. Absorbance spectroscopy can also reveal other properties of the sample, such as concentration, phase transitions, or composition changes. A spectrometer produces a beam of light with multiple wavelengths and shines it on a sample. The sample absorbs some wavelengths and transmits others. The spectrometer detects the intensity of the transmitted light and compares it with the incident light. The absorbance is calculated as the logarithm of the ratio of the incident and transmitted light intensities. The absorbance changes with wavelength, and this change is called the absorption spectrum. The absorption spectrum is characteristic of each sample and depends on its molecular structure and environment. The absorption spectrum of a solution can provide information about the molecular properties of the solute. The wavelengths where the absorbance is maximum indicate

the energy gap between the molecular orbitals of the solute. The magnitude of the absorbance at these wavelengths depends on the molar attenuation coefficient,  $\epsilon$ , which is also called the molar extinction coefficient or the molar absorptivity coefficient. The molar attenuation coefficient can be calculated using the Beer-Lambert law, which relates the absorbance of a solution to its concentration, path length, and wavelength. The Beer-Lambert law can be written as:

$$A = \epsilon cl \tag{2.6}$$

Where  $A$  is absorbance of the solution,  $\epsilon$  is the molar attenuation coefficient of the solute ( $\text{M}^{-1}\text{cm}^{-1}$ ),  $c$  is the molar concentration of the molecule in solution (M), and  $l$  is the path length through the sample i.e., the width of the cuvette (cm)

Absorbance spectroscopy can be used to study the structure, function, and interactions of heme proteins, which are proteins that contain a heme group as a prosthetic group. A heme group is a complex of an iron atom and a porphyrin ring, which has four pyrrole rings connected by methine bridges. The heme group can bind different ligands, such as oxygen, carbon monoxide, nitric oxide, or water, to the iron atom, and can also change its oxidation state from ferrous ( $\text{Fe}^{2+}$ ) to ferric ( $\text{Fe}^{3+}$ ). These changes affect the electronic structure of the heme group, and thus its absorbance spectrum. One of the characteristic features of the absorbance spectrum of heme proteins is the Soret band, which is a strong peak in the ultraviolet region, around 400 nm. The Soret band arises from the  $\pi \rightarrow \pi^*$  transitions of the porphyrin ring, and its position and intensity depend on the ligation and oxidation state of the iron atom.<sup>172</sup> For example, in CYP3A4, the Soret band shifts from 415 nm in the oxidized form to 450 nm in the deoxidized form. The Soret band can also be affected by the conformation of the protein and the environment of the heme group.

By measuring the absorbance spectrum of a heme protein under different conditions, such as varying temperature, pH, ligand concentration, or light exposure, one can obtain information about the

thermodynamics and kinetics of ligand binding, protein folding, and protein-ligand interactions. For example, one can use flash photolysis to induce rapid ligand dissociation from the heme group by a short pulse of light, and then monitor the absorbance changes over time to determine the rate constants and mechanisms of ligand rebinding.

## 2.6 X-ray scattering

X-ray scattering techniques are methods of obtaining structural, elemental, and atomic information about a sample by analyzing the patterns of X-rays that are diffracted and scattered by the sample. The intensity, angle, polarization, wavelength, and energy of the scattered X-rays depend on the properties of the sample, such as its crystal structure, composition, morphology, and orientation. X-ray scattering techniques can be applied to various types of samples, from periodic crystals to novel materials and complex biological molecules/polymers.<sup>173-175</sup> These techniques are non-destructive and can be combined with X-ray microscopy and X-ray spectroscopy for a comprehensive analysis. Any sample that can scatter X-rays can be studied with X-ray scattering. Some proteins may precipitate under prolonged exposure to X-ray, and the exposure duration required for precipitation to form is dependent on protein tolerance and experimental setup i.e., protein medium, temperature, rapidly flowing in a capillary or stationary sample, etc. The parameters that can be measured with X-ray scattering include sample size, shape, dispersity, porosity, morphology, orientation and more.<sup>176, 177</sup>

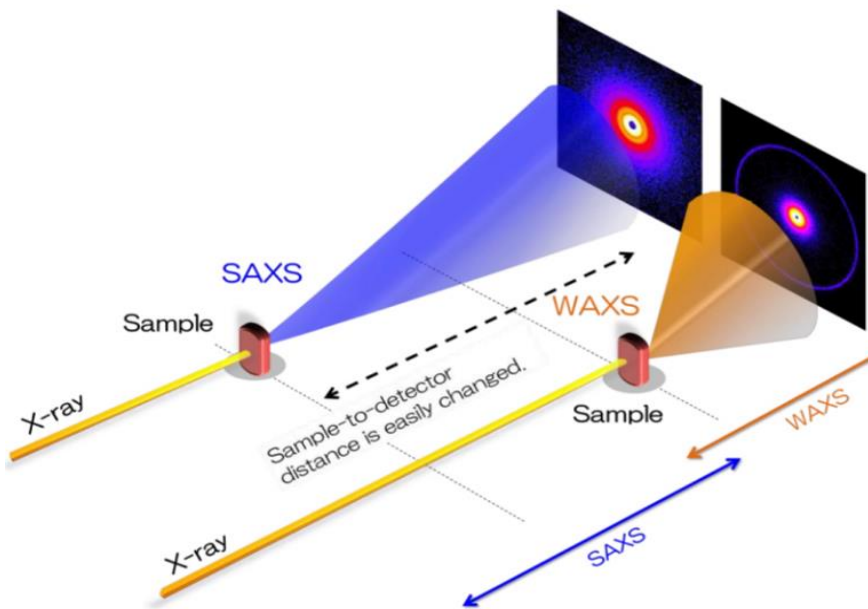
The scattering of X-rays is caused by the interaction of X-rays with electrons in the sample. The scattering intensity is proportional to the number of electrons in the sample, so heavier elements with more electrons tend to scatter more strongly than lighter elements.<sup>178</sup> Another factor that affects the scattering is the contrast between the sample and its surrounding medium. The contrast is determined by the difference in electron density between the sample and the medium. A high contrast is desirable for a clear identification of the sample, while a low contrast should be avoided. For example, silicon particles in a silica gel would

have a low contrast and would be difficult to analyze with X-ray scattering. To enhance the contrast, samples can be immersed in a solution with a different electron density than the sample.<sup>179</sup>

The scattered X-rays from different parts of the sample can interfere with each other, either constructively or destructively. Constructive interference occurs when the scattered X-rays are in phase, meaning that they have the same wavelength and direction. Destructive interference occurs when the scattered X-rays are out of phase, meaning that they have different wavelengths or directions. The detectors that are used to measure the scattered X-rays capture mainly the constructive interference patterns, which reveal information about the location and arrangement of electrons in the sample.<sup>180-182</sup>

The detectors produce a 2D scattering pattern, which shows the distribution of scattered X-rays as a function of angle and intensity. The beamstop is located at the center of the pattern, blocking the direct beam of X-rays that are not scattered by the sample. The 2D scattering pattern can be converted into a 1D scattering curve, which shows the intensity of scattered X-rays as a function of angle only. The 1D scattering curve can be fitted to theoretical models or empirical data to infer the size, shape and surface features of the sample.<sup>183</sup>

The resolution of X-ray scattering techniques depends on the scattering angle of the X-rays. The small-angle X-ray scattering (SAXS) technique can achieve nanoscale resolution (from 100 nm to 1 nm) by measuring the X-rays that are scattered at low angles (less than 5°). The SAXS technique can be used to study the large-scale microstructure of the sample, such as its shape, size, orientation, and aggregation. The wide-angle X-ray scattering (WAXS) technique can achieve atomic resolution (from 1 nm to 0.1 nm) by measuring the X-rays that are scattered at high angles (more than 5°). The WAXS technique can be used to study the small-scale structure of the sample, such as its crystallinity, molecular orientation, and chain packing. The SAXS and WAXS techniques can be performed simultaneously or interchangeably by adjusting the distance between the detector and the sample (Figure 14).<sup>176, 183-185</sup>



**Figure 14.** SAXS and WAXS can be done simultaneously by moving the detector closer or further away from the sample. *Adapted figure from Teledyne Princeton Instruments<sup>2</sup>*

# Chapter 3

## Results and discussion

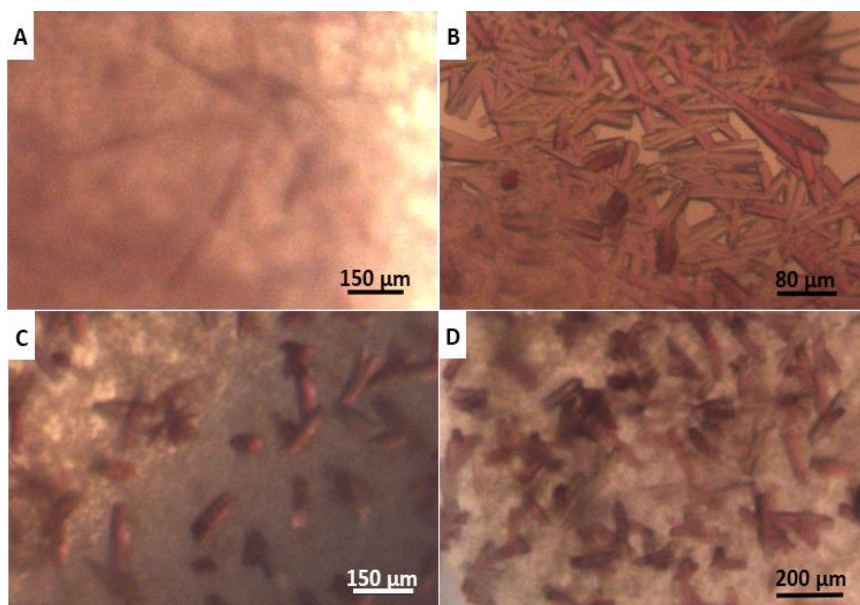
### 3.1 Microcrystallization of CYP3A4 for RT FT-SSX (Paper I)

In this section, I describe the different methods explored for producing large quantities of well-diffracting CYP3A4 microcrystal by vapor diffusion (sitting and hanging drop), batch mode, with/without seeding, with/without ligand co-crystallization. We also solve the RT structure of CYP3A4 and discuss temperature-induced structural difference when compared to a cryogenic structure. The expression and purification of CYP3A4 protein in *Escherichia coli* XL1 blue is presented in Paper I.

#### 3.1.1 Production of CYP3A4 macrocrystals

*Screening and production of CYP3A4 macrocrystals:* Initial screenings to obtain CYP3A4 crystals were performed using the hanging drop vapor diffusion method. This allowed us to screen multiple conditions with minor loss in protein consumption, as protein production was the most demanding and time-consuming aspect. Large needle-shaped crystals formed under these conditions tested with 25 mg/mL protein (26 – 29% PEG4000 + 0.2 M sodium tartrate + 20% glycerol, 17 – 20% PEG4000 + 0.3 M sodium formate + 20% glycerol, 15 – 17% PEG4000 + 10 mM calcium chloride + 20% glycerol), with sizes up to 700  $\mu\text{m}$  in length and averaged approximately  $20 \times 20 \times 300 \mu\text{m}^3$  in size (Figure 15A). The condition that yielded our first batch of crystals presented us with a few issues, i.e., the crystals were too large for our aims, and they were surrounded in high amounts of protein precipitate. Our next goal was to reduce the amount of protein precipitate formed around the crystals, and we hypothesized that reducing viscosity by screening conditions with reduced glycerol and PEG concentrations may reduce protein precipitation. While screening to reduce precipitation, we observed that the crystals produced using sitting drop vapor diffusion

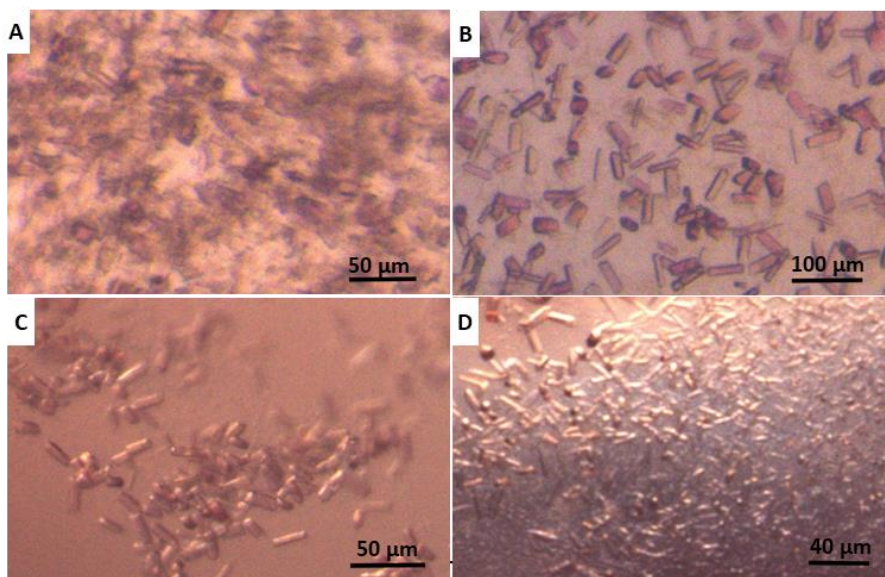
were more fully formed and uniform in size, and the sitting drop method was used for all following screenings. Rod-shaped crystals with greatly reduced protein precipitation formed under a condition tested with 60 mg/ml protein (14% PEG3350 + 0.1 M sodium malate), with sizes up to 200  $\mu\text{m}$  in length and averaging approximately  $30 \times 30 \times 160 \mu\text{m}^3$  in size (Figure 15B ). The crystals grew aggregated, fused together, and was impossible to separate without damaging them. We had managed to reduce the protein precipitation, but the crystals were still quite large for intended FT-SSX experiments, and very high amount of protein would be required to scale up at 60 mg/ml. Reducing the protein concentration to 50 mg/ml resulted in rod-shaped crystals of similar sizes but with less aggregation and without the crystals fusing (Figure 15C). A further reduction to 40 mg/ml resulted in a more uniform population of crystals of similar sizes with even less aggregation (Figure 15D). These crystals diffracted up to 6  $\text{\AA}$  when tested at BioMAX beamline, MAX IV.



**Figure 15. Production of Cytochrome P450 3A4 macrocrystals** (A) Crystals  $\leq 700 \mu\text{m}$  in length and averaging approximately  $20 \times 20 \times 300 \mu\text{m}^3$ . (B) Crystals fused in a lattice  $\leq 200 \mu\text{m}$  in length and  $\approx 30 \times 30 \times 160 \mu\text{m}^3$ . (C – D) Crystals are less aggregated, similar sizes to B, and diffracted up to 6  $\text{\AA}$ . *Figure from paper I.*

### 3.1.2 From macro to microcrystals

To obtain microcrystals, larger crystals were crushed to make seeds using a previously established<sup>186</sup> setup with some alterations. To achieve uniformity in seed sizes, after crushing for 20 – 30 minutes with two ceramic beads in one Eppendorf tube, the tube was allowed to sit for 10 minutes, which allowed the larger seeds to sediment. Larger seeds were further crushed until the seeds were uniform in size when observed under a microscope. Uniform sized seeds were serially diluted to produce a variety in seeds density. 5 – 10% (v/v) final seed concentration of varying seed density were tested. Microcrystals formed from 10% seeds with an average size of  $10 \times 10 \times 15 \mu\text{m}^3$  (Figure 16A). Microcrystals lacked uniformity in shape and stability (melted after 4 – 7 days). 11 – 20% (v/v) seeds were tested following these results. 11 – 13% seeds produced similar results to 10%, and higher seed concentration resulted in protein precipitation. The use of seeds was very resource demanding as it required larger crystals to be grown first, and because the microcrystals obtained lacked uniformity and stability, this method was abandoned for a more direct crystallization process without seeds. Further screenings were performed using sitting drop vapor diffusion without seeds with a goal of reducing the size of the crystals formed. This was achieved by reduction of protein and PEG concentration in small increments – we observed that the size of the crystals formed was sensitive to the precipitant concentration. The optimal crystallization condition of CYP3A4 for our RT FT-SSX experiment was observed to be 11 – 13% PEG 3350, 0.1 M sodium malate, 25 – 30 mg/ml protein concentration without seeds and performed using sitting drop method. 12% PEG + 25 mg/ml protein produced uniform rod-shaped microcrystals formed with minor precipitation. Microcrystals averaging approximately  $20 \times 20 \times 40 \mu\text{m}^3$  in size (Figure 16B) diffracted up to 2.95 Å when tested at BioMAX, MAX IV.



**Figure 16. CYP3A4 macro to microcrystals** (A) Microcrystals from seeding with an average size of  $\approx 10 \times 10 \times 15 \mu\text{m}^3$  – microcrystals were unstable and melted after 4 – 7 days. (B) Microcrystals without seeding  $\approx 20 \times 20 \times 40 \mu\text{m}^3$  in size diffracted up to 3 Å. (C) Microcrystals formed using batch method  $\approx 10 \times 10 \times 20 \mu\text{m}^3$  in size diffracted up to 4.5 Å. (D) Microcrystals formed using hanging drop method with similar protein/buffer/seed conditions to (A). *Figure from paper I.*

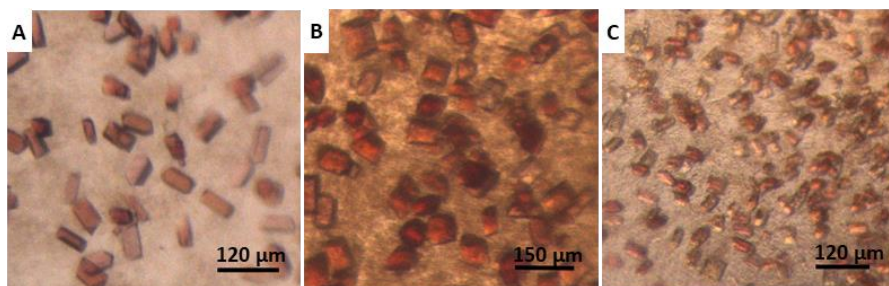
### 3.1.3 Batch crystallization

This was performed using similar conditions to produce larger volumes of crystals simultaneously and reduce the strain caused on the microcrystal during pipetting. Microcrystals may be easily damaged during transfer by from sitting drop wells to PCR tubes so they may be transported for experiments, and then again when pipetted from PCR tube to chip. Batch method aim to reduce the amount of pipetting needed and lessen the damage to microcrystals from pipetting. Uniform rod-shaped microcrystals formed with an average size of  $10 \times 10 \times 20 \mu\text{m}^3$  (Figure 16C). Microcrystals diffracted up to 4.5 Å when tested at BioMAX, MAX IV. Hanging drop vapor diffusion was used during the initial screenings of untested buffer conditions to reduce the amount of protein used. The crystals formed by hanging drop (Figure 16D) were not a representation of expected crystal shape/size when matching

buffer conditions are scaled up using sitting drop method (Figure 16A) but hanging drop crystallization screenings did provided information on if an untested condition would yield crystals.

### 3.1.4 Ligand co-crystallization

This was performed by mixing DMSO dissolved ligand of varying concentration with protein and incubating on ice for 4 h before using the protein for crystallization screenings. Co-crystallization was used as an alternate method to ligand soaking of introducing ligands into our microcrystals. The ligand soaking experiment involves incubating CYP3A4 microcrystals in DMSO dissolved ligand of varying concentration for 1 – 4 h before data collection. We observed crystal degradation after 10 – 20 min when soaked with ketoconazole concentration  $> 250 \mu\text{M}$ . To circumvent this problem, we opted for co-crystallization to test if we can get microcrystals forming with higher ketoconazole concentrations. Ketoconazole co-crystallization altered the shape and size of the crystals formed depending on concentration.  $50 \mu\text{M}$  ketoconazole formed rod-shaped microcrystals with an average size of  $20 \times 20 \times 60 \mu\text{m}^3$  (Figure 17A),  $1 \text{ mM}$  ketoconazole formed square microcrystals with an average size of  $60 \times 60 \times 60 \mu\text{m}^3$  (Figure 17B), and  $2 \text{ mM}$  ketoconazole rod-shaped microcrystals with an average size of  $15 \times 15 \times 35 \mu\text{m}^3$  (Figure 17C). The concentration of ketoconazole affected the stability of the microcrystals –  $50 \mu\text{M}$  remained stable for months,  $1 \text{ mM}$  melted after 5 – 7 days, and  $2 \text{ mM}$  melted after 3 – 7 days.  $1 \text{ mM}$  ketoconazole co-crystallized microcrystals diffracted up to  $5 \text{ \AA}$  when tested at BioMAX. Both the co-crystallization and soaking methods require more screening for optimal ligand concentration and crystallization condition to produce well diffracting microcrystals. This will potentially allow us to observe ligand binding at RT conditions and to perform RT time-resolved ligand binding experiments using CYP3A4 microcrystals.



**Figure 17. CYP3A4 co-crystallization with ketoconazole** (A) 50  $\mu\text{M}$  ketoconazole formed rod-shaped microcrystals with an average size of  $20 \times 20 \times 60 \mu\text{m}$ . Microcrystals remained stable for months. (B) 1 mM ketoconazole formed square microcrystals with an average size of  $60 \times 60 \times 60 \mu\text{m}^3$ . Microcrystals were less stable and degraded after 5 – 7 days. (C) 2 mM ketoconazole formed rod-shaped microcrystals with an average size of  $15 \times 15 \times 35 \mu\text{m}^3$ . Microcrystals were even less stable and degraded after 3 – 7 days. 1mM ketoconazole co-crystallized. Microcrystals diffracted up to 5  $\text{\AA}$  when tested at BioMAX. *Figure from paper I.*

### 3.1.5 Conclusions from CYP3A4 crystallization

Small-volume vapor diffusion hanging drop was the most protein efficient method to screen for new buffer conditions. However, the conditions from the hanging drop method were not fully compatible with sitting-drop, and we had to adjust them when scaling up using hanging-drop. Hanging-drop vapor diffusion produced the best diffracting microcrystals, but microcrystals would stick to the sides of the plates. Batch crystallization had the benefit of requiring less manual handling than sitting drop. It also preserved the crystals better as they did not need to be transferred from the plate to a tube for transportation. Batch crystallization avoided the issue of sample loss due to crystals adhering to the plate and the pipette tip when harvesting the crystals. However, batch crystallization did not produce as high-quality crystals of CYP3A4 as sitting-drop vapor diffusion did. We aim to further optimize the batch methods to achieve better quality crystals in the future, as we found it to be the most time and resource efficient method for obtaining large volumes of microcrystals needed for serial crystallography.

### 3.1.6 Data collection, processing, and structure refinement

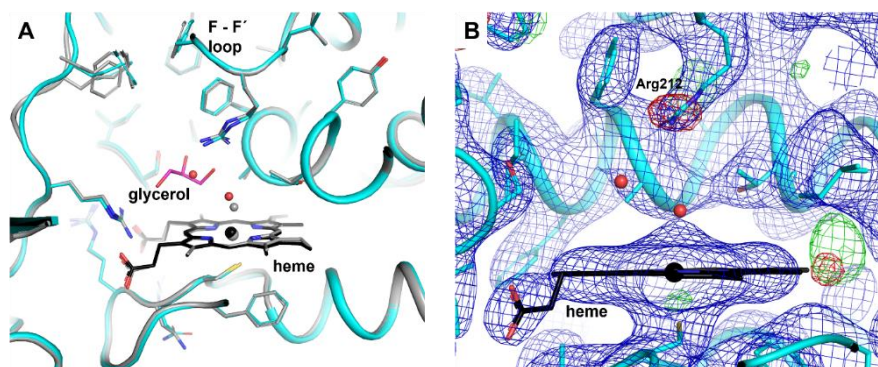
The methods for sample preparation, data collection, processing, and refinement of RT CYP3A4 structure are elaborated upon in Paper 1. FT-SSX diffraction data was collected at BioMAX beamline (Max IV Laboratory) with an X-ray energy of 12.7 keV,  $2.3 \times 10^{12}$  photons  $s^{-1}$  flux with 100% transmission, and 11 ms exposure time. CrystFEL was used for spot finding, indexing, integration, scaling and merging of images with diffracting spots. 169142 images were collected and 66748 were indexed and merged for an indexing rate of 39.5 %. RT structure was solved by molecular replacement with Phaser using cryogenic structure (PDB ID 5VCC) as a search model. Models were built in Coot and refined in REFMAC with an upper resolution limit of 2.95 Å.

### 3.1.7 RT structure of CYP3A4

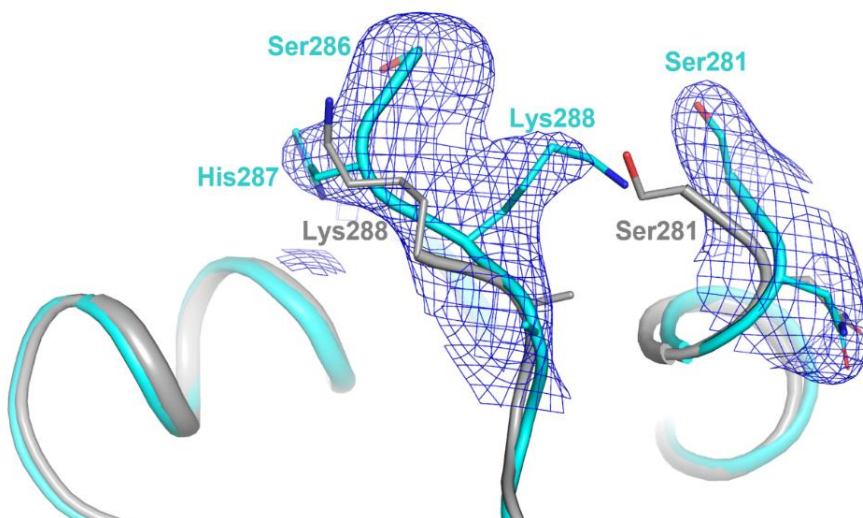
We solved the room-temperature SX structure of CYP3A4 (2.95 Å) and compared it with an apo cryo-structure (PDB ID 5VCC, 1.7 Å resolution). The structures are very similar, with a root-mean-square deviation of 1.14 Å for all atoms. Our structure has two water molecules in the active site (Figure 18A), but no other water molecules due to the limited resolution. Most cryo-structures have either a ligand or a cryo-protectant molecule such as glycerol in the active site. Many CYP3A4 inhibitors interact directly with the heme group or through a water molecule. In our structure, a water molecule contacts the heme iron at 2.8 Å, slightly longer than usual. This may reflect the uncertainty of our data. We also see electron density for another molecule near the water molecule and Arg212, where glycerol often binds in cryo-structures. However, the density does not match glycerol, even though our protein buffer contains it. We modelled this as a water molecule, but it could also be a sodium ion. The active site pocket is well ordered, including the F – F' loop lid region. Arg212 shows some side-chain flexibility and negative difference density (Figure 18B). The pocket has a collapsed conformation with the lid closed, as expected without a ligand. The collapsed active site is well ordered in our room-temperature structure, as in other structures. We did not observe

increased flexibility in the crystalline state at ambient temperature. However, temperature may influence the open-state structure.

The F – F' loop, which is highly flexible, may adopt more distinct conformations at room-temperature than in the frozen-state structures. We noticed that two loops, G – H and H – I, are better resolved in the room-temperature data than in the cryo-structure, despite the lower resolution (2.95 Å for SX and 1.7 Å for cryo). The room-temperature data reveals some extra residues in these loops (Figure 19).



**Figure 18. Room-temperature structure of CYP3A4.** (A) The room-temperature SX structure of CYP3A4 is shown in cyan with the heme group in black and water molecules in red overlaid on the cryo-structure shown in grey (PDB ID 5VCC). The active site glycerol of the cryo-structure is shown in pink. (B) Zoom-in of the active site of the room-temperature CYP3A4 structure. The 2fofc electron density map contoured at 1  $\sigma$  is displayed in blue and the fofc difference electron density map at +3.5  $\sigma$  and -3.5  $\sigma$  in green and red, respectively. *Figure from paper 1.*



**Figure 19. Increased ordering of loops at room-temperature.** The room-temperature SX structure of CYP3A4 is shown in cyan overlaid on the cryo-structure shown in grey (PDB ID 5VCC). The 2fofc electron density map contoured at  $1\sigma$  is displayed in blue. Additional residues of the H – I loop could be modelled in the room-temperature structure due to more well-resolved electron density. *Figure from paper 1.*

### 3.1.8 Summary

We tested different methods to obtain high-quality microcrystals of Cytochrome P450 3A4 (CYP3A4) for FT-SSX. We used sitting-drop vapor diffusion to produce microcrystals for solving the room-temperature structure of human CYP3A4, which was our main experimental goal. We compared the room-temperature structure with a high-resolution cryo-temperature structure to examine temperature-induced structural changes in the active site of CYP3A4, but we did not observe any significant differences. However, we noticed an increased order of loops in the room-temperature structure compared to the cryo-temperature structure. We also explored batch crystallization as a potential method for obtaining large volumes of microcrystals required for future TR SX experiments, and we found it promising with further optimization.

## 3.2 Exploring sEH as model for drug discovery by RT FT-SSX (Paper II)

In this section, we explore methods for producing well diffracting sEH microcrystals for RT SSX. We report seven sEH-ligand complexes obtained by SX at 2.1 - 2.5 Å resolution. We observed that the structures show high similarity to those obtained by cryo-crystallography, but with some loop regions better captured. This indicates that the RT structures may better reflect the physiological state.

### 3.2.1 Production of sEH microcrystals

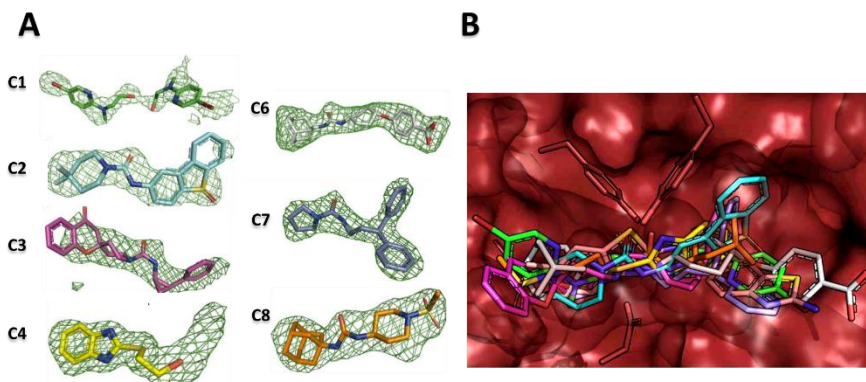
We describe the expression, purification, and crystallization of sEH in detail in Paper II. sEH microcrystals were obtained by seeding. We produced a concentrated seed solution by growing large sEH crystals (100-1000 µm) using vapor diffusion. We crushed the crystals with microseed beads and vortexing, adapting a method by Dods et al. (2017).<sup>186</sup> We optimized batch crystallization conditions using a method we call hybrid crystallization in sitting drop plates, which involved matching the reservoir well solution concentration to the crystallization drop concentration after the addition of protein, and varying the seed concentration, buffer concentrations and protein concentration until optimal conditions were met for the production of sEH microcrystals (20 – 40 µm).

### 3.2.2 Data collection, processing, and structure refinement

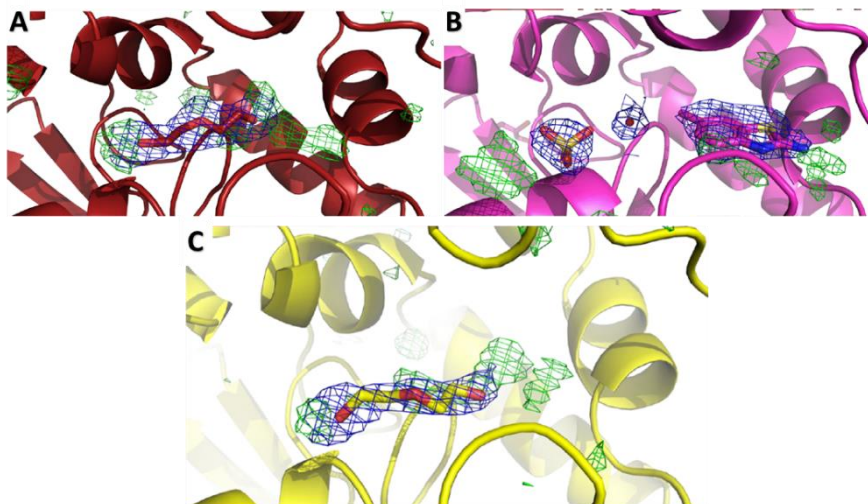
The methods for sample preparation, data collection, processing, and refinement of ligand-bound sEH structure are described in detail in Paper II. RT FT-SSX diffraction data was collected at the BioMAX beamline (MAX IV Laboratory) using 12.7 keV X-ray with 100% transmission and 11 ms exposure time. CrystFEL was used for spot finding, indexing, integration and merging of the images with diffraction spots. Structures were solved by molecular replacement with Phaser using previously solved cryogenic structure as a search model and refined using Buster. Real-space rebuilding and refinement were performed in COOT.

### 3.2.3 RT structures of sEH-ligand complexes and temperature related observations

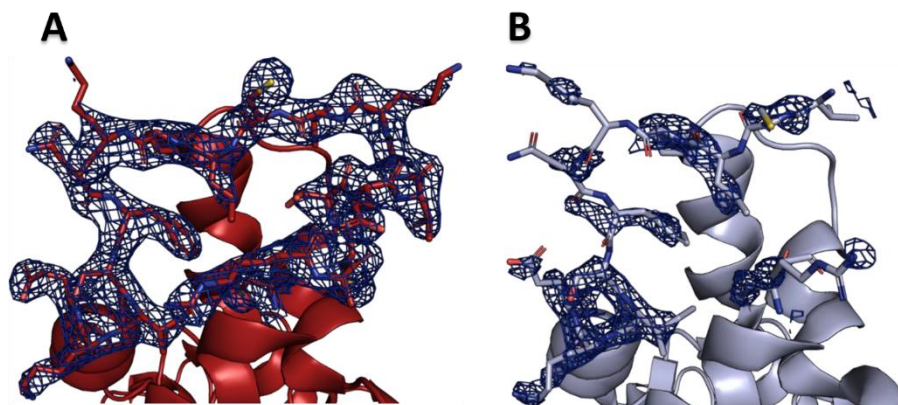
All but one ligand (compound 5) is shown bound in the active site binding pocket (Figure 20B), which shares similarities with sEH-ligand complexes previously solved by cryogenic methods.<sup>187-190</sup> Three of the ligands had a low molecular weight (<300 Da) and were classified as fragments. Two copies of compound 1 (C1, Figure 20A) were found in the active site (Figure 20B). This is a common occurrence in fragment screening caused by high compound concentrations. All but compound 5 shared ligand binding orientations to previous cryo structures. The structure from microcrystals with compound 5 had a large positive difference electron density in the active site. But the compound was not clear, and the density matched a PEG molecule (Figure 21A). An alternative interpretation of the density is compound 5 bound at two partly overlapping sites. The room-temperature structure was different from the cryo-structure, where compound 5 was at a different position and not linked to the catalytic triad. Also, in the cryo-structure (Figure 21B), a sulphate ion and a water molecule is found where the PEG molecule was at room-temperature (Figure 21C). Temperature affects the structures of sEH, especially the loops. The room-temperature models (Figure 22A) show the loops better than the cryo-structure (Figure 22B). This means that SX can give more details on less ordered regions. In our SX structures (Figure 22A), we model the loop with residues 65 to 95 in the N-terminal domain, but this loop is only partly seen in the cryo-structure (Figure 22B).



**Figure 20. Protein-compound complex structures of sEH.** (A) Shows the different compounds that form complexes in their respective conformations numbered according to Table S2 in paper II. The FoFc omit difference electron density map (green) is contoured at  $+3.5 \sigma$  in each case. (B) Displays an overlay of the seven room-temperature sEH complex-structures zoomed-in on the active site. *Adapted figure from paper II.*



**Figure 21. Temperature-dependent difference in fragment binding-mode.** The room-temperature structure from crystals soaked with compound 5 is compared with the cryo-structure in complex with compound 5 and the room-temperature apo structure. (A) In the room-temperature structure of sEH obtained from crystals soaked with compound 5 the active-site density is best fitted with a PEG molecule. (B) In the cryo-structure of compound 5 (PDB ID 5AI8), the fragment is not overlapping with the PEG position. (C) The room-temperature apo structure is displayed with a PEG molecule bound in the active site. For each structure, the 2FoFc electron density map is contoured at  $1 \sigma$  (blue) and the FoFc electron density map at  $+3 \sigma$  (green). *Adapted figure from paper II.*



**Figure 22. Ordering of loops.** An example of a loop (Pro65 to Ala95) with higher degree of order in the room-temperature structures compared to the cryo-structures is shown. (A) Shows the 2FoFc electron density map associated with crystals soaked with compound 5 at room-temperature. (B) Shows the corresponding cryo-structure (PDB ID 5AI8). The respective 2FoFc electron density maps are contoured at  $1\sigma$  (blue). *Adapted figure from paper II.*

### 3.2.4 Summary

We investigated methods to produce well-diffracting microcrystals of soluble Epoxide Hydrolase (sEH) for RT SSX. We obtained sEH microcrystals by seeding and optimized them using hybrid crystallization in sitting drop plates. We used serial crystallography to determine the structures of sEH in complex with seven different ligands, and we solved the ligand-bound structures at resolutions ranging from 2.1 to 2.5 Å. We compared the room-temperature structural data with previously published cryo-temperature structures, and we reported one example of a temperature-dependent difference in ligand-binding mode. We also observed that flexible loops were better resolved at ambient temperature.

### 3.3 Structural dynamics of reduced CO-bound *ba*<sub>3</sub>-type CcO by TR SFX (Paper III)

In this section, I present the structural changes of reduced CO-bound *ba*<sub>3</sub>-type cytochrome c oxidase (CcO) from *Thermus thermophilus* and the TR-SFX results after photodissociation of carbon monoxide (CO) from reduced CO-bound *ba*<sub>3</sub>-type CcO.

#### 3.3.1 Production of *ba*<sub>3</sub>-type CcO microcrystals

The procedure for the expression, purification, and crystallization of *ba*<sub>3</sub>-type CcO are expanded upon in Paper III. The crystallization procedure was performed under anaerobic conditions. All solutions used in the crystallization process were flushed with nitrogen gas before entering the anaerobic chamber. To ensure complete reduction, microcrystals formed were further treated with dithionite. Absorbance spectroscopy was used to verify the stability of reduced CO-bound *ba*<sub>3</sub>-type CcO in lipidic cubic phase (LCP) microcrystals. This confirmed that the microcrystals remained in the reduced CO-bound state for at least 5 days after loading them into gas-tight Hamilton syringes. This indicates that the syringes preserve an anaerobic environment and prevent oxygen contamination of the microcrystals during transportation and throughout our experiment at the Spring-8 Angstrom Compact free electron LASer (SACLA).

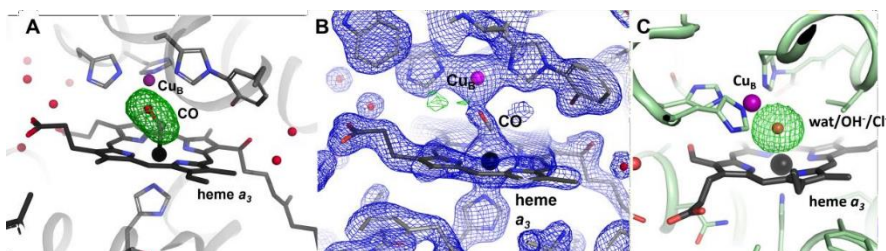
#### 3.3.2 Data collection, processing, and structure refinement

The methods for sample preparation, injection, data collection, processing, and refinement of CO-bound *ba*<sub>3</sub>-type CcO structure are described in detail in Paper III. SFX diffraction data was collected at the BL3 beamline at SACLA using 7.5 keV X-rays with a 30 Hz repetition rate and a <10 fs pulse duration. Cheetah software with a SACLA-adapted pipeline was used to process the data and CrystFEL to select, index and merge the images with diffraction spots. For the reduced CO-bound (dark-state) dataset, 201,274 images were collected, of which 41,059 had diffraction spots and 26,999 were indexed and merged. The crystal hit rate and the indexing rate were 20.4 % and 13.4 %, respectively. Molecular replacement with Phaser

was used to solve the structures (dark and laser-on) using the SFX structure of oxidized *ba*<sub>3</sub>-type CcO (PDB ID 5NDC) as a search model. Models were built in Coot and refined in REFMAC with an upper resolution limit at 2.0 Å.

### 3.3.3 RT structure of reduced CO-bound *ba*<sub>3</sub>-type CcO

The unbiased Fo-Fc map from reduced CO-bound *ba*<sub>3</sub>-type CcO without laser illumination (dark-state) showed an elongated electron density in the active site (Figure 23A), where we modelled a carbon monoxide molecule coordinating with the heme *a*<sub>3</sub> Fe in a bent conformation (Figure 23B). The oxidized SFX structure had a spherical Fo-Fc electron density map feature in the active site, corresponding to a single atom species, possibly a water molecule or hydroxide ion (Figure 23C), or a heavier atom such as chloride. The Fe-Cu distance was 4.9 Å with carbon monoxide bound, slightly longer than 4.8 Å in the oxidized SFX structure of *ba*<sub>3</sub>-type CcO.



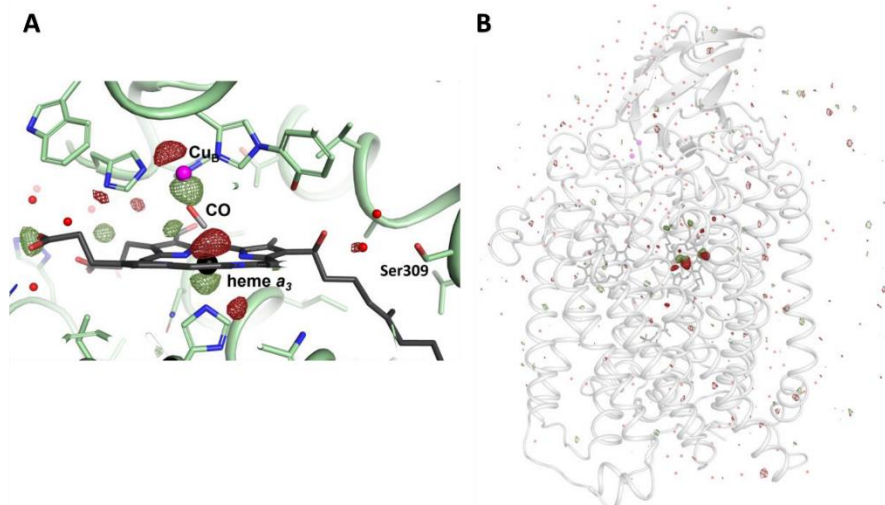
**Figure 23. Active site structure of *ba*<sub>3</sub>-type CcO in the reduced CO-bound and oxidized states from SFX data.** (A) The positive Fo-Fc omit map density (green), calculated without the active site ligand, contoured at 4.5  $\sigma$  for the reduced CO-bound structure. (B) The reduced CO-bound structure with a CO molecule modelled into the active site. The 2Fo-Fc electron density (blue) is contoured at 2  $\sigma$  and the Fo-Fc electron density (positive in green, negative in red) is contoured at 3.5  $\sigma$  around the active site ligand. (C) The positive Fo-Fc omit map density (green), calculated without the active site ligand, contoured at 4.5  $\sigma$  for the oxidized structure (PDB ID 5NDC). *Figure from paper II.*

### 3.3.4 Photodissociation of CO from reduced CO-bound *ba*<sub>3</sub>-type CcO

We used time-resolved X-ray diffraction to study the structural dynamics of CO photodissociation from heme *a*<sub>3</sub> in the active site of

*ba*<sub>3</sub>-type CcO. We applied a pump laser pulse to trigger CO dissociation and recorded the diffraction data 2 ms later. To compare the dark state (reduced CO-bound, no interleaved light images) structure and the photoactivated state (laser-on) structure, we calculated the  $F_o(2\text{ ms}) - F_o(\text{dark})$  isomorphous difference Fourier electron density map (Figure 24, A and B). The map shows strong negative and positive features above and below Cu<sub>B</sub> (Figure 24A), indicating a movement of Cu<sub>B</sub> towards heme *a*<sub>3</sub>. This is in agreement with previous time-resolved X-ray diffraction studies of bovine CcO upon CO flash-photolysis.<sup>191</sup> Similar features are seen on both sides of the heme *a*<sub>3</sub> Fe, suggesting an out-of-plane displacement of the Fe towards the proximal side of the heme. This is consistent with the reduced un-liganded form of heme *a*<sub>3</sub>.<sup>192</sup> Additionally, a weaker positive feature is observed near the CO ligand, implying a partial displacement of CO from the heme *a*<sub>3</sub> Fe. The  $F_o(2\text{ ms}) - F_o(\text{dark})$  difference map (Figure 24B) shows few other strong features, confirming that the light-induced structural changes are confined to the active site of the enzyme.

The refined active-state structure shows that at 2 ms, the heme *a*<sub>3</sub> - CO bond is cleaved, and the CO molecule is bound to Cu<sub>B</sub> instead. Remarkably, the CO switch is accompanied by the transient appearance of a water molecule in the active site, which may stabilize the Cu<sub>B</sub> - CO complex and account for its longer lifetime in *ba*<sub>3</sub>-type CcO. Our results reveal a different structural behavior of *ba*<sub>3</sub>-CcO compared to bovine *aa*<sub>3</sub>-CcO upon CO release from heme *a*<sub>3</sub>.



**Figure 24. Isomorphous difference Fourier electron density map.**  $F_0(2 \text{ ms}) - F_0(\text{dark})$  isomorphous difference Fourier electron density map displaying differences between the dark (reduced CO-bound) state and the activated (light) state, contoured at  $3.5 \sigma$ . Positive density is shown in dark green and negative density in dark red. (A) Zoomed-in view of the heme  $a_3$  – Cu<sub>B</sub> active site. (B) Zoomed-out view of the protein. *Figure from paper III.*

### 3.3.5 Summary

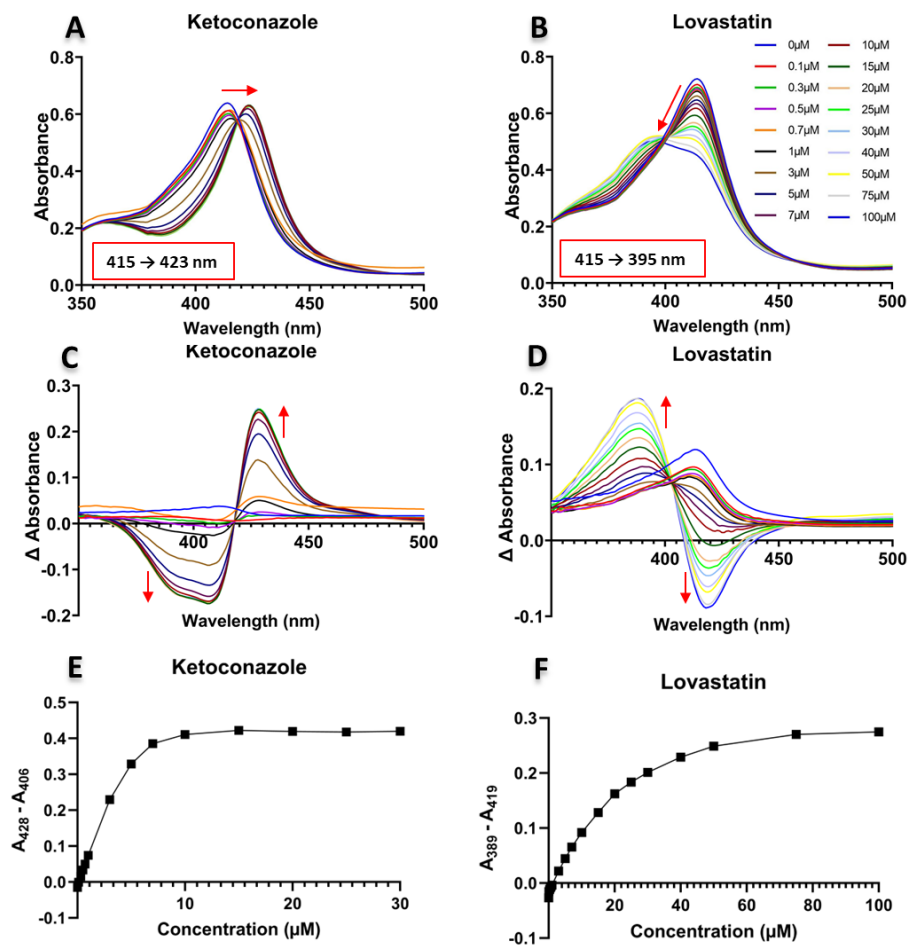
We used time-resolved serial femtosecond crystallography to investigate the structural changes of the active site after flash-photolysis of carbon monoxide (CO) from the reduced heme  $a_3$  of  $ba_3$ -type CcO. Our data reveal how unlike the  $aa_3$ -type enzyme, CO is stabilized on Cu<sub>B</sub> by interactions with a transiently ordered water molecule. Our results provide a structural rationale for the long lifetime of the Cu<sub>B</sub> – CO complex in  $ba_3$ -type CcO, and consequently, the very high oxygen affinity of the enzyme.

## 3.4 CYP3A4 ligand interaction and structural studies by X-ray scattering (Paper IV)

In this section, we confirmed the binding of Ketoconazole and Lovastatin to CYP3A4 in solution, and investigated the ligand interaction and oligomeric state of CYP3A4 by X-ray scattering.

### 3.4.1 CYP3A4 ligand binding titration studies by absorbance spectroscopy

Binding of endogenous substrates and xenobiotics to CYP3A4 results in two types of spectral changes in the UV-visible heme Soret spectrum, referred to as Type I and Type II [11, 23].<sup>68, 193</sup> Type I binders coordinate to the heme via a water molecule.<sup>194-196</sup> Lovastatin is observed in the experiment to function as a Type 1 drug, as it induces a spectral change in the Soret band from 415 to 395 nm (Figure 25B). Type 1 drugs are typically substrates of CYP3A4 for enzymatic oxygenation.<sup>194</sup> Type II drugs inhibit CYP3A4 by pushing away the water and binding directly to the 6th position of the heme iron without changing the spin state.<sup>194-196</sup> This results in a spectral change in the Soret band from 415 to 423 nm (Figure 25A). Apo CYP3A4 was titrated with 0 – 20 % DMSO without ligand as a control. Difference spectrum (Figure 25, C and D) was calculated by subtracting the absorbance values of apo CYP3A4 (DMSO) from ligand bound. Titration curve values (Figure 25, E and F) were obtained by subtracting valleys from peaks of the difference spectrum of 0 – 100  $\mu$ M ligand concentrations. The titration curve was used to estimate the dissociation constants ( $K_d$ ) of Ketoconazole and Lovastatin to be 3  $\mu$ M and 15  $\mu$ M respectively.



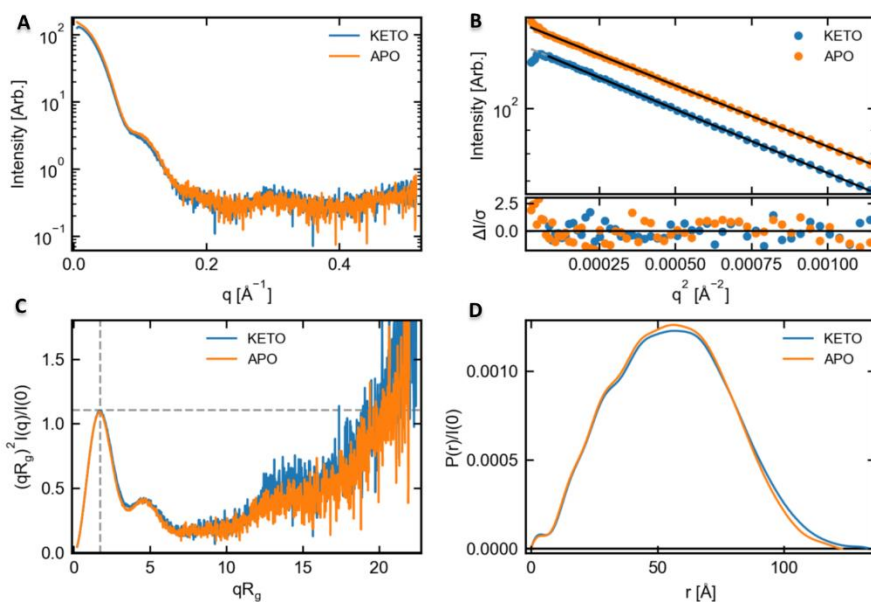
**Figure 25: CYP3A4 absorbance spectroscopy.** (A – B) Type I (Lovastatin) and Type II (Ketoconazole) spectral change from ligand titration experiment. (C – D) Difference spectrum obtained by subtracting the absorbance values of unbound CYP3A4 (DMSO) from bound. (E – F) The titration curve was used to estimate the dissociation constants ( $K_d$ ) of Ketoconazole and Lovastatin to be  $3 \mu\text{M}$  and  $15 \mu\text{M}$  respectively. *Figure from paper IV.*

### 3.4.2 Structural studies by X-ray scattering

The majority of human CYP450 enzymes are membrane-bound microsomal CYP450s that attach to the endoplasmic reticulum via an N-terminal anchor. Although CYP450 enzymes have been traditionally considered as monomers, evidence suggests that they can interact with each other within the membrane through homo- and hetero-oligomerization.<sup>197-199</sup> Previous studies of CYP3A4 intermolecular cross-linking suggests that P450-P450 interactions affect the catalytic properties of the cytochrome P450 system, which cannot be simply derived from the properties of the individual P450 components.<sup>199-201</sup> Detergents have been used to monomerize CYP3A4 in previous studies<sup>199, 202, 203</sup>, and trimer was predicted as the most probable size of the CYP3A4 oligomers in both solution and membrane.<sup>199</sup> We show that CYP3A4 forms tetramers in solution without detergent, both in the apo and Ketoconazole-bound states. The tetramers have the same arrangement as the proposed biological assembly of one of the apo structures (PDB ID 4i3q), which differs from the crystallographic tetramer of the published Ketoconazole-bound structure (PDB ID 2v0m).

The scattering profile data reveals a difference in the low  $q$ -range ( $\sim 0.1 \text{ \AA}^{-1}$ ) of the KETO scattering curve relative to the APO curve (Figure 26A and 27C), implying structural variations between the two tetramers. Observed differences in the 2D scattering curves represent specific coordinates in the 3D structure protein structure differ, as the scattering curve is a 2D representation of the structure. This data indicates that ketoconazole binding may modify the orientation of the monomeric subunits that constitute the tetramer. A possible explanation is that the binding site lid (Figure 6C) is closed in the apo tetramer monomeric subunits, whereas it is open in the ketoconazole-bound monomeric subunits. We hypothesize that better data quality will reveal more differences at higher  $q$ -ranges and lead to more accurate observations of structural differences between the two structures. The Kratky plot (Figure 26C) identifies the CYP3A4 tetramer as a

multidomain protein with folded regions and flexible segments. We speculate that the Kratky analysis may be identifying the different subunits of the tetramer and representing it as multiple domains of a monomeric structure. The pair distance distribution function (Figure 26D) was used to estimate the maximum diameter ( $D_{\max}$ ) of the protein, which was estimated to be 123 Å for unbound CYP3A4 (APO), and 134 Å for Ketoconazole bound CYP3A4 (KETO).



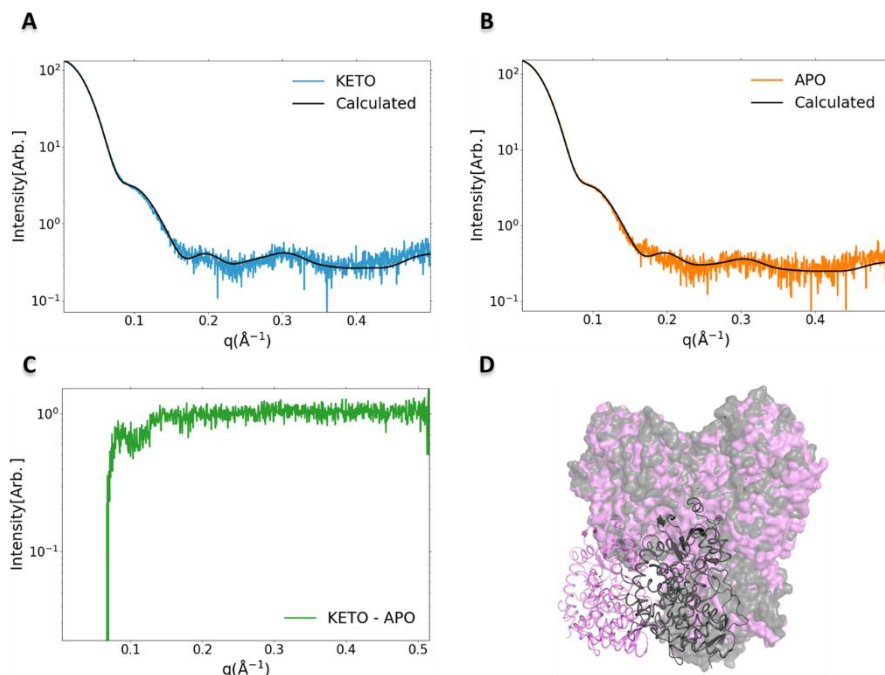
**Figure 26: SAXS analysis of CYP4A.** Ketoconazole bound protein is labeled “KETO”, and apo protein is labeled “APO”. (A) Scattering profile(s) on a log-lin scale. (B) Guinier fit(s) (top) and fit residuals (bottom). (C) Normalized Kratky plot. Dashed lines show where a globular system would peak. (D)  $P(r)$  function(s), normalized by  $I(0)$ . *Figure from paper IV.*

We calculated the 2D scattering curve from the crystallographic Ketoconazole bound tetramer (PDB ID: 2v0m asymmetric unit remodeled in the orientation of PDB ID: 4i3q biological assembly 2) and fitted it to the experimental data using CRY SOL<sup>204</sup> (Figure 27, A and B). The  $\chi^2$  value of the comparison between the experimental and calculated scattering curve was 0.83 for APO and 1.34 for KETO. These values indicate a good fit for both, with a slightly better fit for

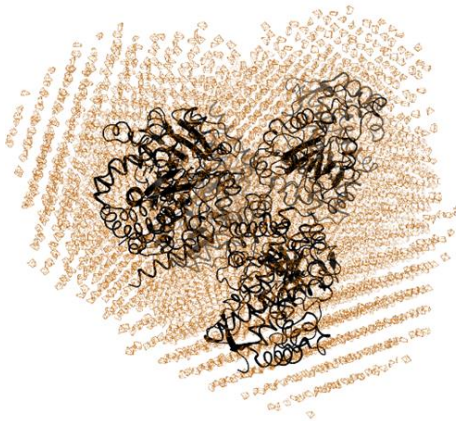
APO. We hypothesize that further reorientation of the crystal structure may improve the fit for KETO. APO scattering curve was subtracted from KETO scattering curve (Figure 27C) to better highlight the previously mentioned structural difference observable at q-range ( $\sim 0.1 \text{ \AA}^{-1}$ ). The *ab initio* model was calculated from the scattering curve and aligned using the crystallographic Ketoconazole bound tetramer mentioned above (Figure 28). The two different crystallographic tetramer assemblies that were used for the CRYSOLO fit are superimposed and shown along with their monomers in Figure 27D (2v0m in pink, 4i3q in black). SAXS analysis statistics, software used for calculations, fitting and modelling are summarized in Table 1.

### 4.2.3 Summary

We used UV-vis absorbance spectroscopy to confirm the binding of Ketoconazole and Lovastatin to CYP3A4 in solution, and we observed the two types of spectral changes in the UV-visible heme Soret spectrum, known as Type I and Type II binding. We also investigated the ligand interaction and oligomeric state of CYP3A4 by X-ray scattering, and our SAXS data showed that CYP3A4 formed tetramers in solution without detergent, both in the apo and Ketoconazole-bound states. The tetramers had the same arrangement as the suggested biological assembly of a previously published apo structure (PDB ID 4i3q), which was different from the crystallographic tetramer of the published Ketoconazole-bound structure (PDB ID 2v0m). We fitted the apo and Ketoconazole-bound SAXS data with the tetramer model composed of monomers of the Ketoconazole-bound structure (2v0m). We found distinct structural differences in the low q-range ( $\sim 0.1 \text{ \AA}^{-1}$ ) of the Ketoconazole-bound scattering curve compared to the apo curve by subtracting the apo scattering curve from the Ketoconazole-bound curve. We hypothesize that better data quality will reveal more differences at higher q-ranges and lead to more accurate observations of structural differences between the two structures.



**Figure 27: SAXS analysis of CYP4A.** Ketoconazole bound protein is labeled “KETO”, and apo protein is labeled “APO”. (A) Comparison of experimental KETO scattering data (blue) and calculated curve from PDB ID: 2v0m remodeled in the orientation of PDB ID: 4i3q biological assembly 2 (black) (B) Comparison of experimental APO scattering data (orange) and the same calculated scattering curve as “A” (black) (C) APO scattering curve subtracted from KETO scattering curve (D). The crystallographic tetramer of the ketoconazole structure (PDB ID 2V0M) is shown in pink overlaid on the suggested biological tetramer from an apo structure (PDB ID 4I3Q) shown in grey. Three monomers superpose well but the fourth monomer is positioned differently in the two tetramers. *Figure from paper IV.*



**Figure 28: Results from DAMMIF *ab initio* modelling based on data from the apo protein.** The model is shown in black and the *ab initio* envelope in orange.

**Table 1. SAXS results of CYP3A4.**

<i>Data collection parameters</i>	<b>KETO</b>	<b>APO</b>
<i>Beamline</i>	BM29 <sup>205, 206</sup> , ESRF, Grenoble, France	
<i>Wavelength (Å)</i>	0.99	
<i>q range (nm<sup>-1</sup>)</i>	0.025-5	
<i>Temperature (K)</i>	283.15	
<b>Structural parameters</b>		
<b>Guinier analysis</b>		
<i>I(0) (cm<sup>-1</sup>)</i>	152 ± 0.11	137 ± 0.12
<i>Rg (Å)</i>	43.49 ± 0.05	42.62 ± 0.04
<i>qRg min</i>	0.401	0.219
<i>qRg max</i>	1.468	1.438
<i>q-range (Å<sup>-1</sup>)</i>	0.0092 to 0.0338	0.0051 to 0.0338
<i>r<sup>2</sup></i>	0.99	1
<b>P(r) analysis</b>		
<i>I(0) (cm<sup>-1</sup>)</i>	136 ± 0.11	151 ± 0.11
<i>Rg (Å)</i>	42.98 ± 0.05	42.09 ± 0.004
<i>q-range (Å<sup>-1</sup>)</i>	0.0092 to 0.5156	0.0051 to 0.5156
<i>D<sub>max</sub>(Å)</i>	134	123
<i>χ<sup>2</sup></i>	0.611	0.622
<i>Porod volume estimate (Å<sup>3</sup>)</i>	311316	322573
<b>Molecular weight (MW) determination</b>		
<i>MW (kDa) – Bayesian inference primus<sup>207</sup></i>	208 (176.6 – 221.1) 55.2% probability	185.8 (162.7 – 221.1) 44.8% probability
<i>MW (kDa) Shape &amp; Size<sup>208</sup></i>	216.8	205.9
<b>Modelling parameters</b>		
<i>Shape reconstruction</i>	DAMMIF <sup>209</sup>	
<i>Symmetry</i>	P1	
<i>No. of models averaged/total</i>	20	
<i>DAMAVER NSD (var)<sup>210</sup></i>	1.45 ± 0.09	1.38 ± 0.07
<i>χ<sup>2</sup></i>	0.63	0.63
<b>Atomistic modelling</b>		
<i>Crystal structures</i>	PDB ID: 2v0m and 4i3q	
<i>χ<sup>2</sup></i>	1.34	0.83
<b>Software employed</b>		
<i>Data evaluation</i>	Primus qt, GNOM <sup>211</sup> BioXTAS RAW <sup>212</sup>	
<i>Computation of model intensities</i>	CRY SOL <sup>204</sup>	
<i>3D graphics representations</i>	Pymol <sup>213</sup>	
<i>Ab initio model alignment</i>	CIFSUP	

# Concluding Summary

The work in this thesis provides a detailed description of different methods for obtaining high-quality microcrystals suitable for serial crystallography and contributed some structural insight of three different proteins at ambient temperature. In **paper I**, we solved the room-temperature structure of human CYP3A4 at 2.95 Å and compared it to a high-resolution cryo-temperature structure to observe temperature-induced structural changes in the active site. We did not observe any significant differences, we did however, notice an increased order of loops in the room-temperature structure compared to the cryo-temperature structure. CYP3A4 is a suitable target for time-resolved experiments at XFELs. We aim to investigate how temperature influences the open-state conformation of CYP3A4, in contrast to the closed-state reported in the paper. We also plan to use reaction triggering to capture the intermediate states and the opening of the active site lid at different time intervals. Moreover, we want to examine the effect of protein ligands with different sizes on the active site flexibility at room temperature. We expect to obtain higher-resolution structures of CYP3A4 by using XFELs. We have achieved significant improvements in CYP3A4 microcrystallization, but we still need to optimize the batch method for better diffraction quality and test the hybrid method proposed in paper II. In **paper II**, we used serial crystallography to determine the structures of sEH in complex with seven different ligands and solved the ligand-bound structures at resolutions ranging from 2.1 to 2.5 Å. Comparing the room-temperature structural data with previously published cryo-temperature structures presented us with an example of a temperature-dependent difference in ligand-binding mode, and we observed that flexible loops were better resolved at ambient temperature. In **paper III**, we used time-resolved serial femtosecond crystallography to investigate the structural changes of the active site after flash-photolysis of carbon monoxide (CO) from the reduced heme  $a_3$  of  $ba_3$ -type CcO. This led to the observation that unlike the  $aa_3$ -type enzyme, CO is stabilized on  $\text{Cu}_B$  by interactions with a transiently ordered water molecule and

provided a structural rationale for the long lifetime of the Cu<sub>B</sub> – CO complex in *ba*<sub>3</sub>-type CcO, and consequently, the very high oxygen affinity of the enzyme. In **paper IV**, we used UV-vis absorbance spectroscopy to confirm the binding of Ketoconazole and Lovastatin to CYP3A4 in solution. We also investigated the ligand interaction and oligomeric state of the protein by X-ray scattering. We determined CYP3A4 formed tetramers in solution and observed a distinct structural difference at low q-range caused by ketoconazole binding. We predict that better data quality will reveal more differences at higher q-ranges and lead to more accurate observations of structural differences between the two structures. This thesis demonstrates some of the advantages of using serial crystallography at room temperature for the structural characterization of proteins. Serial crystallography offers several advantages over conventional crystallography, but it also poses challenges such as low throughput and high labor demand compared to conventional crystallography. However, I believe that these challenges can be overcome with further development and automation of serial crystallography methods. Serial crystallography is still in its infancy, similar to the situation of conventional crystallography decades ago, but it has the potential to provide valuable insights into protein dynamics at near-physiological conditions.

# Acknowledgements

Lundberg has been a lovely place to work. Thank you everyone for making me feel welcome. To my fellow lunatics who skipped directly to this page, I see you. **To everyone at Lundberg** whose name isn't included in this section, **thank you!** I would like to thank you all by name but the printing company charges by the page, and **Gisela** is making me pay out of pocket.

**Gisela**, thank you for having a warm presence and for creating a comforting environment around you. I have always felt comfortable talking to you about anything. Maybe too comfortable? You are the epitome of intelligence without being egotistical. I appreciate your sense of humor and kindness. You have been an excellent collaborator, and I am proud to have been part of your wonderful group. I would like to borrow your time-turner someday though, because you must own one to be able to attend the absurd number of meetings that you do, and you manage to make enough time for everyone. **Richard**, the dance floor king! Thanks for getting all that funding so we can have nice things. It is really nice of you to check up on people and not only your own students. Thanks to you and Helena for driving me home from the airport and I appreciate you serenading the hallway with your singing. Although I was in the general sense of "huh?" in many of our conversations, I have enjoyed them all the same. I miss the Elven long hair. **Anne F** - You are an absolute hero of mine, and I will never stop spreading the gospel of what an incredible educator you are. Thank you for making science fun and for conning me to stick with it. I still haven't played Frostpunk yet, but I will be unemployed soon with a lot of free time... I probably still won't though, it looks depressing. Thank you for being my friend.

**Ooi!** If all I got out of the last few years at Lundberg was gaining you as a friend, I would consider it time well spent. Maria and I look forward to babysitting and spoiling your future children. We thank you for lifting the curse! Remember, ~~if you murder him~~, we can start a charity to support his family. **Jonathan**, do you remember when you made me buy multiple controllers so we could play games together?

Yeah... Me too. Still waiting... You were the first friend I made in the lab, and you made me feel welcome and part of the group. I spent a lot of late evenings performing my experiments because I had spent the day chatting with you, only to realize I still have experiments to complete. It has been a lot of fun hanging out with you at and outside of work, and I hope that continues. **Dunge**, Gossip Girl here! Your one and only source into the scandalous lives of Lundberg's elites. There is no one else I would choose to lose my mind with staring at dots on a screen for 24 hours in a row with. I still haven't figured out if your love of the US is trolling or not, but I appreciate an excellent troll! XOXO. **Johan**, my Oompa Loompa, I think of you with the same fondness and mild annoyance as I do my younger siblings. Thanks for teaching me everything I know about romance. I hope you have a lot of fun with the rest of the time you have left, and I have no doubt you will be successful (ok maybe a little doubt... it's CYP after all). **Doris** – I always knew I was in for a fun and memorable day whenever you walked into my office. Thanks for all the conversations we have had over the years. I am glad HR was not listening. **Gabrielle**, my fellow horse girl! You are delightful, unintentionally hilarious, and I find your enthusiasm admirable. I think very fondly of every moment I have spent around you. Stay gullible! **Jessiiiiica!** You are an absolute fashion icon! I'm happy you have got that lovely house. Now that I am unemployed, I will need a basement to live in. I wonder how many of my experiments failed because I got distracted chatting with you! You were our unofficial 3rd office mate and I think Andrea is still a little upset that you ditched him for Flo. You were thoroughly missed when you were away on maternity leave. **Andrea**, my beautiful trash panda! No one sprints across a hallway in absolute panic like you. I am so happy I got to share offices with you over the years! You are very knowledgeable, kind, and you take your time to help others even when you are feeling stressed about your own work. We had a lot of fun trash talking our projects together and threatening to quit at least once a day. **Monika**, you have been very kind and generous with your time through the years, and I really appreciate you. You must have been a fast draw cowboy in your past life based on how quickly you pick up your phone to reply to my texts. I will forever be grateful to you. **Arpitha**, I am

always happy to see you, and cheers for always being yourself. You are one of a kind, and it just so happened to be my favorite kind. I still laugh out loud when I think about you walking around the lab in oversized ski gear. I hope you can still get crystals without me standing close by. **Emil**, I had a lot of fun exploring Osaka with you, we made excellent vacation friends. I have enjoyed hanging out with you outside of work over the years. **Emma**, it was a pleasure having you in the group. We had a lot of laughs, and it is unfortunate that covid interrupted our time together. Please present your thesis, the hard part was already done. **Gilbert and Jens**, the Tom and Jerry of Lundberg lab. When is the next game night? **Florian**, I think of you whenever see a tomato, and I eat a lot of tomatoes. **Damasus** - I would love to thank you but would just be me thanking myself. Ahuru m gi n'anya nwanne. **Julia**, I'm sorry for laughing at your very memorable walk at ESRF when you were in a lot of pain post-surgery. I hope you are now fully recovered and feeling much better. I admire how driven you are, and I wish for all your dreams to come through. You are welcome for the cardio gained chasing me around for Tuesday seminars. **Laras**, I always enjoyed talking to you, but I do not see you much anymore... I have missed you! **Tinna**, Thanks for keeping the lab running and keeping me safe by always reminding me to wear a lab coat. The lab completely fell apart without you, to such an extent we had to move to a new lab. **Leona**, Hi Leona :-) To whomever Leona keeps stealing Leona's bike/bike parts? Stop it! She is nice and needs it to get home from work. **Adams**, thanks for all you do. I like your sense of humor. **Kristina**, thank you for having me as a collaborator on one of your papers, it was a very relaxing experience. **Taru**, thanks for being kind and sharing blue dyes with me. **Lucija**, thank you for picking Andrea up for smoke breaks. It was nice to get some alone time. **Irena**, either you spent a lot of time in the äkta room, or our schedules were just perfectly aligned! It was always a pleasant experience running into you. **Emilie**, I love your outfits! To the remaining basement dwellers **Filippo, Charles, Božidar**, you all can finally leave the basement. It was fun playing board games with you guys. Do not miss out on the cardio from going up and down those stairs. **Hanna**, we have not interacted much, but I am a fan of anyone who enjoys making fun of

Ylber. To those we have loved and lost along the way, **Cecilia, Rebecka, Daniel, SwaG!, Darius, Petra, Rob, Amke, Elin, Leo Viktor, Maja, Davide, Rasmus, etc.**, we miss you and we mourn you. I will see you all in Valhalla shortly. I think **Davide** is back though? If that is true, welcome back **Davide**. Goodbye **Davide**. **Katharina**, you are welcome for the renovation. I have always been curious about one of the ingredients in the post-reno summer rolls, and I still haven't figure out what it was. I wished we had talked more over the years, but I am shy. You make an excellent HP trivia partner. **Catarina** and **Johanna**, thanks for looking out for us PhD students. **Valida, Lars** and **Brunno**, what you do is important, and we appreciate you! Thanks for keeping things running. To the **Janitorial staff**, thanks for keeping the building clean. We are all a bunch of lunatics, and it probably isn't easy cleaning up after us, so **thank you!**

To **my family**, you are all alright. To **my future family**, you are also wonderful and have raised a lovely daughter. To **Lea**, how much longer are you going to stay in the states? I miss you, and it brings me a lot of comfort to know you will be in my life forever. It is nice to have friends I do not have to make plans with. To all the **friends** I have gained and lost along the way, I cherish you.

**Maria**, hej älskling, kan du gissa vad jag tänker på just nu?

# Bibliography

1. Diederichs, K. and M. Wang, *Serial Synchrotron X-Ray Crystallography (SSX)*. Methods Mol Biol, 2017. **1607**: p. 239-272.
2. *Introduction to X-ray Scattering*, in *X-ray Scattering*. Teledyne Princeton Instruments.
3. Zhao, F.-Z., et al., *A guide to sample delivery systems for serial crystallography*. The FEBS Journal, 2019. **286**(22): p. 4402-4417.
4. Sarti, P., et al., *Cytochrome c oxidase and nitric oxide in action: Molecular mechanisms and pathophysiological implications*. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2012. **1817**(4): p. 610-619.
5. Urban, P., et al., *Ligand Access Channels in Cytochrome P450 Enzymes: A Review*. International Journal of Molecular Sciences, 2018. **19**(6): p. 1617.
6. Greule, A., et al., *Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism*||The authors would like to acknowledge the outstanding and inspirational contributions of Paul Ortiz de Montellano to the field of cytochrome P450 research and of Prof. Sir Alan R. Battersby (1925–2018) to the molecular understanding of biosynthesis over the course of their careers. Natural Product Reports, 2018. **35**(8): p. 757-791.
7. Hartley, H., *Origin of the Word 'Protein'*. Nature, 1951. **168**(4267): p. 244-244.
8. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*, in *Interdisciplinary Applied Mathematics*. 2010, Springer Nature. p. 77-104.
9. Poos Mary I, C.R., Carlson-Newberry Sydne J, *The role of protein and amino acids in sustaining and enhancing performance*. Committee on Military Nutrition Research: Activity Report, ed. I.o.M.U.C.o.M.N. Research. 1999, Washington, D.C.: National Academy Press. xv, 429 p.
10. Alberts B, J.A., Lewis J, et al., *Molecular Biology of the Cell. 4th edition*. The Shape and Structure of Proteins. 2002, New York: Garland Science.

11. De Chadarevian, S., *Sequences, Conformation, Information: Biochemists and Molecular Biologists in the 1950s*. Journal of the History of Biology, 1996. **29**(3): p. 361-386.
12. Branden, C.I. and J. Tooze, *Introduction to protein structure*. 2012: Garland Science.
13. Cozzone, A.J., *Proteins: Fundamental chemical properties*. Encyclopedia of Life Sciences. No month listed—2002. John Wiley & Sons Ltd, 2002: p. 1-10.
14. Kendrew, J.C., et al., *A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis*. Nature, 1958. **181**(4610): p. 662-666.
15. De Chadarevian, S., *John Kendrew and myoglobin: Protein structure determination in the 1950s*. Protein Science, 2018. **27**(6): p. 1136-1143.
16. Linderstrøm-Lang, K.U., *Lane Medical Lectures: Proteins and Enzymes*. 1952: Stanford University Press.
17. Wetlaufer, D.B., *Terminology of Protein Structure*. Nature, 1961. **190**(4781): p. 1113-1113.
18. Haynie, D.T. and B. Xue, *Superdomains in the protein structure hierarchy: The case of PTP-C2*. Protein Science, 2015. **24**(5): p. 874-882.
19. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(4096): p. 223-30.
20. Sun, P.D., C.E. Foster, and J.C. Boyington, *Overview of Protein Structural and Functional Folds*. Current Protocols in Protein Science, 2004. **35**(1).
21. Rehman I, F.M., Botelho S, *Biochemistry, Secondary Protein Structure*. 2022, National Library of Medicine: StatPearls.
22. Dyson, H.J., P.E. Wright, and H.A. Scheraga, *The role of hydrophobic interactions in initiation and propagation of protein folding*. Proceedings of the National Academy of Sciences, 2006. **103**(35): p. 13057-13061.
23. Malleshappa Gowder, S., et al., *Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins*. The Scientific World Journal, 2014. **2014**: p. 1-7.
24. Englander, S.W. and L. Mayne, *The nature of protein folding pathways*. Proceedings of the National Academy of Sciences, 2014. **111**(45): p. 15873-15880.

25. Thomas, J.M., *The birth of X-ray crystallography*. Nature, 2012. **491**(7423): p. 186-187.
26. Fersht, A.R., *From the first protein structures to our current knowledge of protein folding: delights and scepticisms*. Nature Reviews Molecular Cell Biology, 2008. **9**(8): p. 650-654.
27. Maveyraud, L. and L. Mourey, *Protein X-ray Crystallography and Drug Discovery*. Molecules, 2020. **25**(5): p. 1030.
28. Carvalho, A.L., J. Trincão, and M.J. Romão, *X-Ray Crystallography in Drug Discovery*, in *Ligand-Macromolecular Interactions in Drug Discovery*. 2010, Humana Press. p. 31-56.
29. Davis, A.M., S.J. Teague, and G.J. Kleywegt, *Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design*. Angewandte Chemie International Edition, 2003. **42**(24): p. 2718-2736.
30. Kubelka, J., et al., *Sub-microsecond Protein Folding*. Journal of Molecular Biology, 2006. **359**(3): p. 546-553.
31. Darnell, S., *Why Structure Prediction Matters*, in *Structural Biology*. 2020: DNASTAR.
32. Staker, B.L., G.W. Buchko, and P.J. Myler, *Recent contributions of structure-based drug design to the development of antibacterial compounds*. Curr Opin Microbiol, 2015. **27**: p. 133-8.
33. Siltberg-Liberles, J., J.A. Grahnen, and D.A. Liberles, *The evolution of protein structures and structural ensembles under functional constraint*. Genes (Basel), 2011. **2**(4): p. 748-62.
34. Thompson, M.C., T.O. Yeates, and J.A. Rodriguez, *Advances in methods for atomic resolution macromolecular structure determination*. F1000Res, 2020. **9**.
35. Nam, K.H., *Serial X-ray Crystallography*. Crystals, 2022. **12**(1): p. 99.
36. Zatsepin, N.A., et al., *The complementarity of serial femtosecond crystallography and MicroED for structure determination from microcrystals*. Current Opinion in Structural Biology, 2019. **58**: p. 286-293.
37. Chapman, H.N., et al., *Femtosecond X-ray protein nanocrystallography*. Nature, 2011. **470**(7332): p. 73-7.

38. Martin-Garcia, J.M., et al., *Serial femtosecond crystallography: A revolution in structural biology*. Archives of Biochemistry and Biophysics, 2016. **602**: p. 32-47.
39. Fromme, P. and J.C.H. Spence, *Femtosecond nanocrystallography using X-ray lasers for membrane protein structure determination*. Current Opinion in Structural Biology, 2011. **21**(4): p. 509-516.
40. Smith, N. and M.A. Wilson, *Understanding Cysteine Chemistry Using Conventional and Serial X-ray Protein Crystallography*. Crystals, 2022. **12**(11): p. 1671.
41. Roedig, P., et al., *High-speed fixed-target serial virus crystallography*. Nature Methods, 2017. **14**(8): p. 805-810.
42. Cohen, A.E., et al., *Goniometer-based femtosecond crystallography with X-ray free electron lasers*. Proceedings of the National Academy of Sciences, 2014. **111**(48): p. 17122-17127.
43. Gati, C., et al., *Serial crystallography on in vivo grown microcrystals using synchrotron radiation*. IUCrJ, 2014. **1**(Pt 2): p. 87-94.
44. Sherrell, D.A., et al., *Fixed-target serial crystallography at the Structural Biology Center*. Journal of Synchrotron Radiation, 2022. **29**(5): p. 1141-1151.
45. Andersson, R., et al., *Serial femtosecond crystallography structure of cytochrome c oxidase at room temperature*. Scientific Reports, 2017. **7**(1): p. 4518.
46. Weierstall, U., et al., *Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography*. Nat Commun, 2014. **5**: p. 3309.
47. Sugahara, M., et al., *Grease matrix as a versatile carrier of proteins for serial crystallography*. Nat Methods, 2015. **12**(1): p. 61-3.
48. Conrad, C.E., et al., *A novel inert crystal delivery medium for serial femtosecond crystallography*. IUCrJ, 2015. **2**(Pt 4): p. 421-30.
49. Vakili, M., et al., *Microfluidic polyimide gas dynamic virtual nozzles for serial crystallography*. Rev Sci Instrum, 2020. **91**(8): p. 085108.
50. Kloudová, B., et al., *Gas Dynamic Virtual Nozzle Sprayer for an Introduction of Liquid Samples in Atmospheric Pressure*

- Ionization Mass Spectrometry*. Analytical Chemistry, 2023. **95**(8): p. 4196-4203.
51. Wiedorn, M.O., et al., *Rapid sample delivery for megahertz serial crystallography at X-ray FELs*. IUCrJ, 2018. **5**(Pt 5): p. 574-584.
  52. DePonte, D., et al., *Gas dynamic virtual nozzle for generation of microscopic droplet streams*. Journal of Physics D: Applied Physics, 2008. **41**(19): p. 195505.
  53. Butryn, A., et al., *An on-demand, drop-on-drop method for studying enzyme catalysis by serial crystallography*. Nature Communications, 2021. **12**(1): p. 4461.
  54. Sevrioukova, I.F. and T.L. Poulos, *Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems*. Dalton Trans, 2013. **42**(9): p. 3116-26.
  55. Brodie, B.B., et al., *Detoxication of drugs and other foreign compounds by liver microsomes*. Science, 1955. **121**(3147): p. 603-4.
  56. Klingenberg, M., *Pigments of rat liver microsomes*. Arch Biochem Biophys, 1958. **75**(2): p. 376-86.
  57. Omura, T. and R. Sato, *A new cytochrome in liver microsomes*. J Biol Chem, 1962. **237**: p. 1375-6.
  58. Omura, T. and R. Sato, *The Carbon Monoxide-Binding Pigment Of Liver Microsomes. I. Evidence For Its Hemoprotein Nature*. J Biol Chem, 1964. **239**: p. 2370-8.
  59. McDonnell, A.M. and C.H. Dang, *Basic review of the cytochrome p450 system*. J Adv Pract Oncol, 2013. **4**(4): p. 263-8.
  60. Nelson, D.R., *The cytochrome p450 homepage*. Hum Genomics, 2009. **4**(1): p. 59-65.
  61. Bailey, D.G. and G.K. Dresser, *Interactions between grapefruit juice and cardiovascular drugs*. Am J Cardiovasc Drugs, 2004. **4**(5): p. 281-97.
  62. Bailey, D.G., et al., *Grapefruit juice-drug interactions*. Br J Clin Pharmacol, 1998. **46**(2): p. 101-10.
  63. Kato, M., *Intestinal first-pass metabolism of CYP3A4 substrates*. Drug Metab Pharmacokinet, 2008. **23**(2): p. 87-94.
  64. Barnaba, C., et al., *The catalytic function of cytochrome P450 is entwined with its membrane-bound nature*. F1000Res, 2017. **6**: p. 662.

65. Zhou, S.F., *Drugs behave as substrates, inhibitors and inducers of human cytochrome P450 3A4*. *Curr Drug Metab*, 2008. **9**(4): p. 310-22.
66. Schaffenburg, W., B.N. Lockshin, and C.M.C. Deklotz, *Polymorphisms*. *Comprehensive Dermatologic Drug Therapy*, 2021.
67. Bruno, R.D. and V.C. Njar, *Targeting cytochrome P450 enzymes: a new approach in anti-cancer drug development*. *Bioorg Med Chem*, 2007. **15**(15): p. 5047-60.
68. Isin, E.M. and F.P. Guengerich, *Substrate binding to cytochromes P450*. *Anal Bioanal Chem*, 2008. **392**(6): p. 1019-30.
69. Yun, C.H., et al., *Functional expression of human cytochrome P450 enzymes in Escherichia coli*. *Curr Drug Metab*, 2006. **7**(4): p. 411-29.
70. Griffin, B.W. and J.A. Peterson, *Camphor binding by Pseudomonas putida cytochrome P-450. Kinetics and thermodynamics of the reaction*. *Biochemistry*, 1972. **11**(25): p. 4740-6.
71. Shimada, T., et al., *Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians*. *J Pharmacol Exp Ther*, 1994. **270**(1): p. 414-23.
72. Isin, E.M. and F.P. Guengerich, *Kinetics and thermodynamics of ligand binding by cytochrome P450 3A4*. *J Biol Chem*, 2006. **281**(14): p. 9127-36.
73. Roberts, A.G. and W.M. Atkins, *Energetics of heterotropic cooperativity between alpha-naphthoflavone and testosterone binding to CYP3A4*. *Arch Biochem Biophys*, 2007. **463**(1): p. 89-101.
74. Denisov, I.G., et al., *Cooperativity in cytochrome P450 3A4: linkages in substrate binding, spin state, uncoupling, and product formation*. *J Biol Chem*, 2007. **282**(10): p. 7066-76.
75. Yoon, M.Y., A.P. Campbell, and W.M. Atkins, *"Allosterism" in the elementary steps of the cytochrome P450 reaction cycle*. *Drug Metab Rev*, 2004. **36**(2): p. 219-30.
76. Sohl, C.D., et al., *Cooperativity in oxidation reactions catalyzed by cytochrome P450 1A2: highly cooperative pyrene*

- hydroxylation and multiphasic kinetics of ligand binding*. J Biol Chem, 2008. **283**(11): p. 7293-308.
77. Fernando, H., J.R. Halpert, and D.R. Davydov, *Resolution of multiple substrate binding sites in cytochrome P450 3A4: the stoichiometry of the enzyme-substrate complexes probed by FRET and Job's titration*. Biochemistry, 2006. **45**(13): p. 4199-209.
  78. Ekroos, M. and T. Sjögren, *Structural basis for ligand promiscuity in cytochrome P450 3A4*. Proceedings of the National Academy of Sciences, 2006. **103**(37): p. 13682-13687.
  79. Hendrychová, T., et al., *Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations*. Biochim Biophys Acta, 2011. **1814**(1): p. 58-68.
  80. Cojocar, V., P.J. Winn, and R.C. Wade, *The ins and outs of cytochrome P450s*. Biochim Biophys Acta, 2007. **1770**(3): p. 390-401.
  81. Otyepka, M., et al., *What common structural features and variations of mammalian P450s are known to date?* Biochim Biophys Acta, 2007. **1770**(3): p. 376-89.
  82. Skopalík, J., P. Anzenbacher, and M. Otyepka, *Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences*. J Phys Chem B, 2008. **112**(27): p. 8165-73.
  83. Anzenbacher, P., et al., *Active Sites of Cytochromes P450: What are They Like?* Acta Chimica Slovenica, 2008. **55**: p. 63-66.
  84. Olsen, L., et al., *Prediction of Activation Energies for Hydrogen Abstraction by Cytochrome P450*. Journal of Medicinal Chemistry, 2006. **49**(22): p. 6489-6499.
  85. Locuson, C.W. and T.S. Tracy, *Identification of binding sites of non-I-helix water molecules in mammalian cytochromes p450*. Drug Metab Dispos, 2006. **34**(12): p. 1954-7.
  86. Wright, W.C., J. Chenge, and T. Chen, *Structural perspectives of the CYP3A family and their small molecule modulators in drug metabolism*. Liver Research, 2019. **3**(3): p. 132-142.
  87. Anzenbacher, P. and J. Hudeček, *Differences in flexibility of active sites of cytochromes P450 probed by resonance Raman*

- and UV-Vis absorption spectroscopy. *Journal of Inorganic Biochemistry*, 2001. **87**(4): p. 209-213.
88. Ohkura, K., et al., *Flexible structure of cytochrome P450: promiscuity of ligand binding in the CYP3A4 heme pocket*. *Anticancer research*, 2009. **29**(3): p. 935-942.
89. Teixeira, V.H., V. Ribeiro, and P.J. Martel, *Analysis of binding modes of ligands to multiple conformations of CYP3A4*. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 2010. **1804**(10): p. 2036-2045.
90. Sevrioukova, I.F. and T.L. Poulos, *Structural and mechanistic insights into the interaction of cytochrome P4503A4 with bromoergocryptine, a type I ligand*. *Journal of Biological Chemistry*, 2012. **287**(5): p. 3510-3517.
91. Graham, S.E. and J.A. Peterson, *How similar are P450s and what can their differences teach us?* *Arch Biochem Biophys*, 1999. **369**(1): p. 24-9.
92. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam*. *J Mol Biol*, 1987. **195**(3): p. 687-700.
93. Ducharme, J., *Investigation of cytochrome P450 3A4 allosteric properties*, in *Department of Chemistry*. 2021, McGill University.
94. Werck-Reichhart, D. and R. Feyereisen, *Cytochromes P450: a success story*. *Genome Biol*, 2000. **1**(6): p. Reviews3003.
95. Gotoh, O., *Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences*. *J Biol Chem*, 1992. **267**(1): p. 83-90.
96. Lüdemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways*. *J Mol Biol*, 2000. **303**(5): p. 813-30.
97. Winn, P.J., et al., *Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine*. *Proc Natl Acad Sci U S A*, 2002. **99**(8): p. 5361-6.

98. Johnson, E.F. and C.D. Stout, *Structural diversity of eukaryotic membrane cytochrome p450s*. J Biol Chem, 2013. **288**(24): p. 17082-90.
99. Newman, J.W., C. Morisseau, and B.D. Hammock, *Epoxide hydrolases: their roles and interactions with lipid metabolism*. Progress in Lipid Research, 2005. **44**(1): p. 1-51.
100. Shan, J. and K. Hashimoto, *Soluble Epoxide Hydrolase as a Therapeutic Target for Neuropsychiatric Disorders*. Int J Mol Sci, 2022. **23**(9).
101. Morisseau, C. and B.D. Hammock, *Impact of soluble epoxide hydrolase and epoxyeicosanoids on human health*. Annu Rev Pharmacol Toxicol, 2013. **53**: p. 37-58.
102. Harris, T.R. and B.D. Hammock, *Soluble epoxide hydrolase: gene structure, expression and deletion*. Gene, 2013. **526**(2): p. 61-74.
103. Gill, S.S. and B.D. Hammock, *Distribution and properties of a mammalian soluble epoxide hydrase*. Biochemical Pharmacology, 1980. **29**(3): p. 389-395.
104. Hammock, B.D., et al., *Soluble mammalian epoxide hydratase: action on juvenile hormone and other terpenoid epoxides*. Comp Biochem Physiol B, 1976. **53**(2): p. 263-5.
105. Imig, J.D. and B.D. Hammock, *Soluble epoxide hydrolase as a therapeutic target for cardiovascular diseases*. Nature Reviews Drug Discovery, 2009. **8**(10): p. 794-805.
106. Chiamvimonvat, N., et al., *The Soluble Epoxide Hydrolase as a Pharmaceutical Target for Hypertension*. Journal of Cardiovascular Pharmacology, 2007. **50**(3): p. 225-237.
107. Borsini, A., *The role of soluble epoxide hydrolase and its inhibitors in depression*. Brain Behav Immun Health, 2021. **16**: p. 100325.
108. Broadley, M., et al., *Chapter 7 - Function of Nutrients: Micronutrients*, in *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*, P. Marschner, Editor. 2012, Academic Press: San Diego. p. 191-248.
109. Li, Y., et al., *Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex*. J Bioenerg Biomembr, 2006. **38**(5-6): p. 283-91.

110. Tsukihara, T., et al., *Structures of Metal Sites of Oxidized Bovine Heart Cytochrome c Oxidase at 2.8 Å*. *Science*, 1995. **269**(5227): p. 1069-1074.
111. Sharma, V., et al., *Insights into functions of the H channel of cytochrome c oxidase from atomistic molecular dynamics simulations*. *Proceedings of the National Academy of Sciences*, 2017. **114**(48): p. E10339-E10348.
112. Watson, S.A. and G.P. McStay, *Functions of Cytochrome c oxidase Assembly Factors*. *Int J Mol Sci*, 2020. **21**(19).
113. Paulus, A., et al., *The cytochrome ba3 oxidase from Thermus thermophilus does not generate a tryptophan radical during turnover: Implications for the mechanism of proton pumping*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2015. **1847**(10): p. 1093-1100.
114. Schimo, S., et al., *Cytochrome c Oxidase Biogenesis and Metallochaperone Interactions: Steps in the Assembly Pathway of a Bacterial Complex*. *PLoS One*, 2017. **12**(1): p. e0170037.
115. Keightley, J.A., et al., *Molecular genetic and protein chemical characterization of the cytochrome ba3 from Thermus thermophilus HB8*. *J Biol Chem*, 1995. **270**(35): p. 20345-58.
116. Soulimane, T., et al., *Three-dimensional crystals of cytochrome-c oxidase from Thermus thermophilus diffracting to 3.8 Å resolution*. *FEBS Lett*, 1995. **368**(1): p. 132-4.
117. Siletsky, S.A., et al., *Time-resolved single-turnover of ba3 oxidase from Thermus thermophilus*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2007. **1767**(12): p. 1383-1392.
118. Ohtani, N., M. Tomita, and M. Itaya, *An extreme thermophile, Thermus thermophilus, is a polyploid bacterium*. *J Bacteriol*, 2010. **192**(20): p. 5499-505.
119. Smirnova, I., et al., *Functional role of Thr-312 and Thr-315 in the proton-transfer pathway in ba3 Cytochrome c oxidase from Thermus thermophilus*. *Biochemistry*, 2010. **49**(33): p. 7033-9.
120. Von Ballmoos, C., et al., *Proton transfer in ba3 cytochrome c oxidase from Thermus thermophilus*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2012. **1817**(4): p. 650-657.
121. von Ballmoos, C., et al., *Timing of Electron and Proton Transfer in the ba3 Cytochrome c Oxidase from Thermus thermophilus*. *Biochemistry*, 2012. **51**(22): p. 4507-4517.

122. Kannt, A., et al., *Electrical current generation and proton pumping catalyzed by the ba3-type cytochrome c oxidase from Thermus thermophilus*. FEBS letters, 1998. **434**(1-2): p. 17-22.
123. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges*. Frontiers in Microbiology, 2014. **5**.
124. Xia, X.-X., Z.-G. Qian, and S.Y. Lee, *Comparative proteomic and genetic analyses reveal unidentified mutations in Escherichia coli XLI-Blue and DH5 $\alpha$* . FEMS Microbiology Letters, 2011. **314**(2): p. 119-124.
125. Williams, P.A., et al., *Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone*. Science, 2004. **305**(5684): p. 683-6.
126. Chang, M.C., et al., *Engineering Escherichia coli for production of functionalized terpenoids using plant P450s*. Nat Chem Biol, 2007. **3**(5): p. 274-7.
127. Brustad, E.M., et al., *Structure-guided directed evolution of highly selective p450-based magnetic resonance imaging sensors for dopamine and serotonin*. J Mol Biol, 2012. **422**(2): p. 245-62.
128. Briand, L., et al., *A self-inducible heterologous protein expression system in Escherichia coli*. Scientific Reports, 2016. **6**(1): p. 33037.
129. Martínez, L.M., *Protein Purification Methods*, in *How Protein Purification works*. 2019: Sepmag.
130. Chakraborty, S., K. Trihemasava, and G. Xu, *Modifying Baculovirus Expression Vectors to Produce Secreted Plant Proteins in Insect Cells*. J Vis Exp, 2018(138).
131. *Traditional methods of cell lysis for Protein Extraction*. Thermo Fisher Scientific 2006.
132. Thomas, S.N., et al., *Liquid chromatography–tandem mass spectrometry for clinical diagnostics*. Nature Reviews Methods Primers, 2022. **2**(1): p. 96.
133. Wong, J.W., R.L. Albright, and N.-H.L. Wang, *Immobilized Metal Ion Affinity Chromatography (IMAC) Chemistry and Bioseparation Applications*. Separation and Purification Methods, 1991. **20**(1): p. 49-106.
134. *Immobilized Metal Chelate Affinity Chromatography (IMAC)*. Sigma Aldrich, 2023.

135. Harris, L.J. and T.W. Birch, *Zwitterions: Proof of the zwitterion constitution of the amino-acid molecule. II. Amino-acids, polypeptides, etc., and proteins as zwitterions, with instances of non-zwitterion ampholytes*. *Biochem J*, 1930. **24**(4): p. 1080-97.
136. Cummins, P.M., K.D. Rochfort, and B.F. O'Connor, *Ion-Exchange Chromatography: Basic Principles and Application*. *Methods Mol Biol*, 2017. **1485**: p. 209-223.
137. Burgess, R.R., *A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting*. *Protein Expr Purif*, 2018. **150**: p. 81-85.
138. Fekete, S., et al., *Theory and practice of size exclusion chromatography for the analysis of protein aggregates*. *J Pharm Biomed Anal*, 2014. **101**: p. 161-73.
139. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization*. *Acta Crystallogr F Struct Biol Commun*, 2014. **70**(Pt 1): p. 2-20.
140. Hartel, R.W. and A.V. Shastry, *Sugar crystallization in food products*. *Crit Rev Food Sci Nutr*, 1991. **30**(1): p. 49-112.
141. Cole, J.J., et al., *Carbon dioxide supersaturation in the surface waters of lakes*. *Science*, 1994. **265**(5178): p. 1568-70.
142. Hwang, H., et al., *Hydration breaking and chemical ordering in a levitated NaCl solution droplet beyond the metastable zone width limit: evidence for the early stage of two-step nucleation*. *Chem Sci*, 2020. **12**(1): p. 179-187.
143. *Hanging Drop Vapor Diffusion Crystallization*, in *Crystal Growth 101*. 2020, Hampton Research.
144. Dessau, M.A. and Y. Modis, *Protein crystallization for X-ray crystallography*. *J Vis Exp*, 2011(47).
145. Li, D., et al., *Crystallizing Membrane Proteins in the Lipidic Mesophase. Experience with Human Prostaglandin E2 Synthase I and an Evolving Strategy*. *Crystal Growth & Design*, 2014. **14**(4): p. 2034-2047.
146. Moraes, I. and M. Archer, *Methods for the Successful Crystallization of Membrane Proteins*, in *Structural Proteomics: High-Throughput Methods*, R.J. Owens, Editor. 2015, Springer New York: New York, NY. p. 211-230.
147. Ishchenko, A., E.E. Abola, and V. Cherezov, *Crystallization of Membrane Proteins: An Overview*, in *Protein Crystallography*:

- Methods and Protocols*, A. Wlodawer, Z. Dauter, and M. Jaskolski, Editors. 2017, Springer New York: New York, NY. p. 117-141.
148. Kubicek, J., et al., *Controlled in meso phase crystallization--a method for the structural investigation of membrane proteins*. PLoS One, 2012. **7**(4): p. e35458.
  149. Zabara, A., et al., *The nanoscience behind the art of in-meso crystallization of membrane proteins*. Nanoscale, 2017. **9**(2): p. 754-763.
  150. van 't Hag, L., et al., *In Meso Crystallization: Compatibility of Different Lipid Bicontinuous Cubic Mesophases with the Cubic Crystallization Screen in Aqueous Solution*. Crystal Growth & Design, 2014. **14**(4): p. 1771-1781.
  151. Conn, C.E., *Lipid Mesophases for Crystallizing Membrane Proteins*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1269-1273.
  152. Devey, R. *Crystallography - An Overview*. Life Science 2021.
  153. Huxford, T., *X-Ray Crystallography*, in *Brenner's Encyclopedia of Genetics (Second Edition)*, S. Maloy and K. Hughes, Editors. 2013, Academic Press: San Diego. p. 366-368.
  154. Nam, K.H., *Molecular Dynamics—From Small Molecules to Macromolecules*. International Journal of Molecular Sciences, 2021. **22**(7): p. 3761.
  155. Pflugrath, J.W., *Practical macromolecular cryocrystallography*. Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 6): p. 622-42.
  156. Durdagi, S., et al., *Near-physiological-temperature serial crystallography reveals conformations of SARS-CoV-2 main protease active site for improved drug repurposing*. Structure, 2021. **29**(12): p. 1382-1396. e6.
  157. Watkin, D., *Structure refinement: some background theory and practical strategies*. Journal of Applied Crystallography, 2008. **41**(3): p. 491-522.
  158. Melnikov, I., et al., *The complex analysis of X-ray mesh scans for macromolecular crystallography*. Acta Crystallographica Section D, 2018. **74**(4): p. 355-365.

159. Huang, C.Y., et al., *In meso in situ serial X-ray crystallography of soluble and membrane proteins*. Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 6): p. 1238-56.
160. Diederichs, K. and M. Wang, *Serial Synchrotron X-Ray Crystallography (SSX)*, in *Protein Crystallography: Methods and Protocols*, A. Wlodawer, Z. Dauter, and M. Jaskolski, Editors. 2017, Springer New York: New York, NY. p. 239-272.
161. Dubach, V.R.A. and A. Guskov, *The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy*. Crystals, 2020. **10**(7): p. 580.
162. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallographica Section D, 2013. **69**(7): p. 1204-1214.
163. Powell, H.R., *A beginner's guide to X-ray data processing*. The Biochemist, 2021. **43**(3): p. 46-50.
164. White, T.A., et al., *CrystFEL: a software suite for snapshot serial crystallography*. Journal of Applied Crystallography, 2012. **45**(2): p. 335-341.
165. *X-Ray electron density maps*, in *RSCB PDB General help*. 2022: RCSB Protein data bank.
166. Evans, P. and A. McCoy, *An introduction to molecular replacement*. Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 1-10.
167. Shabalin, I.G., P.J. Porebski, and W. Minor, *Refining the macromolecular model – achieving the best agreement with the data from X-ray diffraction experiment*. Crystallography Reviews, 2018. **24**(4): p. 236-262.
168. Vitkup, D., et al., *Why protein R-factors are so large: a self-consistent analysis*. Proteins, 2002. **46**(4): p. 345-54.
169. Engh, R.A., *X-Ray Crystallography, Basic Principles*, in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 2026-2029.
170. Agirre, J., et al., *The CCP4 suite: integrative software for macromolecular crystallography*. Acta Crystallogr D Struct Biol, 2023. **79**(Pt 6): p. 449-461.
171. Potterton, L., et al., *CCP4i2: the new graphical user interface to the CCP4 program suite*. Acta Crystallogr D Struct Biol, 2018. **74**(Pt 2): p. 68-84.

172. Engler, N., et al., *Protein Dynamics in an Intermediate State of Myoglobin: Optical Absorption, Resonance Raman Spectroscopy, and X-Ray Structure Analysis*. Biophysical Journal, 2000. **78**(4): p. 2081-2092.
173. Chu, B. and B.S. Hsiao, *Small-Angle X-ray Scattering of Polymers*. Chemical Reviews, 2001. **101**(6): p. 1727-1762.
174. Wall, M.E., S.E. Ealick, and S.M. Gruner, *Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease*. Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6180-4.
175. Meisburger, S.P., D.A. Case, and N. Ando, *Diffuse X-ray scattering from correlated motions in a protein crystal*. Nature Communications, 2020. **11**(1): p. 1271.
176. Takahara, A., et al., *Application of Synchrotron Radiation X-ray Scattering and Spectroscopy to Soft Matter*. Polymers (Basel), 2020. **12**(7).
177. Hémonnot, C.Y.J. and S. Köster, *Imaging of Biological Materials and Cells by X-ray Scattering and Diffraction*. ACS Nano, 2017. **11**(9): p. 8542-8559.
178. Choi, K.H. and M. Morais, *Use of small-angle X-ray scattering to investigate the structure and function of dengue virus NS3 and NS5*. Methods Mol Biol, 2014. **1138**: p. 241-52.
179. Lombardo, D., P. Calandra, and M.A. Kiselev, *Structural Characterization of Biomaterials by Means of Small Angle X-rays and Neutron Scattering (SAXS and SANS), and Light Scattering Experiments*. Molecules, 2020. **25**(23): p. 5624.
180. Bunaciu, A.A., E.G. Udriștioiu, and H.Y. Aboul-Enein, *X-ray diffraction: instrumentation and applications*. Crit Rev Anal Chem, 2015. **45**(4): p. 289-99.
181. Bopp, J., et al., *X-ray Phase Contrast: Research on a Future Imaging Modality: An Introductory Guide*. 2018. p. 191-205.
182. Changizi, V., A.A. Kheradmand, and M.A. Oghabian, *Application of small-angle X-ray scattering for differentiation among breast tumors*. J Med Phys, 2008. **33**(1): p. 19-23.
183. Da Vela, S. and D.I. Svergun, *Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution*. Current Research in Structural Biology, 2020. **2**: p. 164-170.

184. Pauw, B.R., *Everything SAXS: small-angle scattering pattern collection and correction*. Journal of Physics: Condensed Matter, 2013. **25**(38): p. 383201.
185. Tants, J.N. and A. Schlundt, *Advances, Applications, and Perspectives in Small-Angle X-ray Scattering of RNA*. Chembiochem, 2023: p. e202300110.
186. Dods, R., et al., *From Macrocystals to Microcrystals: A Strategy for Membrane Protein Serial Crystallography*. Structure, 2017. **25**(9): p. 1461-1468.e2.
187. Gomez, G.A., et al., *Human soluble epoxide hydrolase: structural basis of inhibition by 4-(3-cyclohexylureido)-carboxylic acids*. Protein Sci, 2006. **15**(1): p. 58-64.
188. Öster, L., et al., *Successful generation of structural information for fragment-based drug discovery*. Drug Discov Today, 2015. **20**(9): p. 1104-11.
189. Shen, H.C. and B.D. Hammock, *Discovery of Inhibitors of Soluble Epoxide Hydrolase: A Target with Multiple Potential Therapeutic Indications*. Journal of Medicinal Chemistry, 2012. **55**(5): p. 1789-1808.
190. Vázquez, J., et al., *Screening and Biological Evaluation of Soluble Epoxide Hydrolase Inhibitors: Assessing the Role of Hydrophobicity in the Pharmacophore-Guided Search of Novel Hits*. J Chem Inf Model, 2023. **63**(10): p. 3209-3225.
191. Shimada, A., et al., *A nanosecond time-resolved XFEL analysis of structural changes associated with CO release from cytochrome c oxidase*. Sci Adv, 2017. **3**(7): p. e1603042.
192. Liu, B., et al., *Combined Microspectrophotometric and Crystallographic Examination of Chemically Reduced and X-ray Radiation-Reduced Forms of Cytochrome ba3 Oxidase from Thermus thermophilus: Structure of the Reduced Form of the Enzyme*. Biochemistry, 2009. **48**(5): p. 820-826.
193. Leibman, K.C., A.G. Hildebrandt, and R.W. Estabrook, *Spectrophotometric studies of interactions between various substrates in their binding to microsomal cytochrome P-450*. Biochem Biophys Res Commun, 1969. **36**(5): p. 789-94.
194. Das, A., et al., *Screening of type I and II drug binding to human cytochrome P450-3A4 in nanodiscs by localized surface plasmon resonance spectroscopy*. Anal Chem, 2009. **81**(10): p. 3754-9.

195. Dawson, J.H., L.A. Andersson, and M. Sono, *Spectroscopic investigations of ferric cytochrome P-450-CAM ligand complexes. Identification of the ligand trans to cysteinate in the native enzyme.* Journal of Biological Chemistry, 1982. **257**(7): p. 3606-3617.
196. Jones, J.P., et al., *The effects of nitrogen-heme-iron coordination on substrate affinities for cytochrome P450 2E1.* Chem Biol Interact, 2011. **193**(1): p. 50-6.
197. Arendse, L.B. and J.M. Blackburn, *Effects of polymorphic variation on the thermostability of heterogenous populations of CYP3A4 and CYP2C9 enzymes in solution.* Scientific Reports, 2018. **8**(1).
198. Reed, J.R. and W.L. Backes, *Formation of P450· P450 complexes and their effect on P450 function.* Pharmacology & therapeutics, 2012. **133**(3): p. 299-310.
199. Davydov, D.R., et al., *Interactions among cytochromes P450 in microsomal membranes: oligomerization of cytochromes P450 3A4, 3A5, and 2E1 and its functional consequences.* Journal of Biological Chemistry, 2015. **290**(6): p. 3850-3864.
200. Davydov, D.R., *Microsomal monooxygenase as a multienzyme system: the role of P450-P450 interactions.* Expert Opinion on Drug Metabolism & Toxicology, 2011. **7**(5): p. 543-558.
201. Davydov, D.R., et al., *Pivotal role of P450-P450 interactions in CYP3A4 allostery: the case of  $\alpha$ -naphthoflavone.* Biochem J, 2013. **453**(2): p. 219-30.
202. Grinkova, Y.V., I.G. Denisov, and S.G. Sligar, *Functional reconstitution of monomeric CYP3A4 with multiple cytochrome P450 reductase molecules in Nanodiscs.* Biochem Biophys Res Commun, 2010. **398**(2): p. 194-8.
203. Denisov, I.G., et al., *Allosteric Interactions in Human Cytochrome P450 CYP3A4: The Role of Phenylalanine 213.* Biochemistry, 2019. **58**(10): p. 1411-1422.
204. Svergun, D., C. Barberato, and M.H.J. Koch, *CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates.* Journal of Applied Crystallography, 1995. **28**(6): p. 768-773.
205. Pernot, P., et al., *New beamline dedicated to solution scattering from biological macromolecules at the ESRF.* Journal of Physics: Conference Series, 2010. **247**(1): p. 012009.

206. Pernot, P., et al., *Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution*. Journal of Synchrotron Radiation, 2013. **20**(4): p. 660-664.
207. Hajizadeh, N.R., et al., *Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data*. Sci Rep, 2018. **8**(1): p. 7204.
208. Franke, D., C.M. Jeffries, and D.I. Svergun, *Machine Learning Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions*. Biophys J, 2018. **114**(11): p. 2485-2492.
209. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering*. J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
210. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering*. Journal of Applied Crystallography, 2003. **36**(3 Part 1): p. 860-864.
211. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis*. Journal of Applied Crystallography, 2003. **36**(5): p. 1277-1282.
212. Hopkins, J.B., R.E. Gillilan, and S. Skou, *BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis*. Journal of Applied Crystallography, 2017. **50**(5): p. 1545-1553.
213. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.8*. 2015.